Makoto Takizawa
Leonard Barolli
Tomoya Enokido (Eds.)

LNCS 5186

Network-Based Information Systems

2nd International Conference, NBiS 2008
Turin, Italy, September 2008
Proceedings

Lecture Notes in Computer Science 5186

Commenced Publication in 1973
Founding and Former Series Editors:
Gerhard Goos, Juris Hartmanis, and Jan van Leeuwen

Editorial Board

David Hutchison
Lancaster University, UK

Takeo Kanade
Carnegie Mellon University, Pittsburgh, PA, USA

Josef Kittler
University of Surrey, Guildford, UK

Jon M. Kleinberg
Cornell University, Ithaca, NY, USA

Alfred Kobsa
University of California, Irvine, CA, USA

Friedemann Mattern
ETH Zurich, Switzerland

John C. Mitchell
Stanford University, CA, USA

Moni Naor
Weizmann Institute of Science, Rehovot, Israel

Oscar Nierstrasz
University of Bern, Switzerland

C. Pandu Rangan
Indian Institute of Technology, Madras, India

Bernhard Steffen
University of Dortmund, Germany

Madhu Sudan
Massachusetts Institute of Technology, MA, USA

Demetri Terzopoulos
University of California, Los Angeles, CA, USA

Doug Tygar
University of California, Berkeley, CA, USA

Gerhard Weikum
Max-Planck Institute of Computer Science, Saarbruecken, Germany

Makoto Takizawa Leonard Barolli
Tomoya Enokido (Eds.)

Network-Based Information Systems

2nd International Conference, NBiS 2008
Turin, Italy, September 1-5, 2008
Proceedings

Springer

Volume Editors

Makoto Takizawa
Seikei University, Department of Computer and Information Science
3-3-1 Kichijoji-Kitamachi, Musashinao, Tokyo 180-8633, Japan
E-mail: makoto.takizawa@st.seikei.ac.jp

Leonard Barolli
Fukuoka Institute of Technology (FIT), Faculty of Information Engineering
Department of Information and Communication Engineering
3-30-1 Wajiro-Higashi, Higashi-ku, Fukuoka 811-0214, Japan
E-mail: barolli@fit.ac.jp

Tomoya Enokido
Rissho University, Faculty of Business Administration
4-2-16, Osaki, Shinagawa, Tokyo 141-8602, Japan
E-mail: eno@ris.ac.jp

Library of Congress Control Number: Applied for

CR Subject Classification (1998): C.2, D.4.4, H.3, D.1.3

LNCS Sublibrary: SL 3 – Information Systems and Application, incl. Internet/Web and HCI

ISSN 0302-9743
ISBN-10 3-540-85692-7 Springer Berlin Heidelberg New York
ISBN-13 978-3-540-85692-4 Springer Berlin Heidelberg New York

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, re-use of illustrations, recitation, broadcasting, reproduction on microfilms or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer. Violations are liable to prosecution under the German Copyright Law.

Springer is a part of Springer Science+Business Media

springer.com

© Springer-Verlag Berlin Heidelberg 2008
Printed in Germany

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper SPIN: 12457984 06/3180 5 4 3 2 1 0

Preface

Welcome to the proceeding of the Second International Conference on Network-Based Information Systems (NBiS 2008), held in conjunction with the 19th International Conference on Database and Expert Systems Applications, DEXA-2008, in Turin, Italy, September 1–2, 2008.

The main objective of NBiS 2008 was to bring together scientists, engineers, and researchers from the fields of network systems and information systems with the aim of encouraging the exchange of ideas, opinions, and experience between these two communities.

NBiS started as a workshop, and it was held as such for 9 years together with the DEXA International Conference, making it the oldest of the DEXA workshops. The workshop was very successful in quantity and quality. We received every year a large number of paper submissions, but as a workshop we could only accept a limited number of papers. This is the second year that NBiS was run as an international conference together with DEXA.

This year, we received 81 research papers from all over the world. The submitted papers were carefully reviewed by at least three reviewers. Based on the review results, the Program Committee members selected 32 high-quality papers to be presented during the NBiS 2008 international conference.

Many volunteers kindly helped us to prepare and organize NBiS 2008. First of all, we would like to thank all the authors for submitting their papers, the Program Committee members, and the reviewers, who carried out the most difficult work in evaluating the submitted papers. We would like to thank the Keynote Speaker of NBiS 2008, Arjan Durresi, Indiana University-Purdue University Indianapolis, USA, for accepting our invitation.

We would also like to thank the DEXA Association for giving us the chance to organize the conference. We are particularly grateful to Gabriela Wagner for her kind support and help, and also for dealing with the conference registration. We would like to express our appreciation to the staff of the Politecnico di Torino for their work as local organizers. Finally, we would like to thank all the participants of the conference. We look forward to meeting you again at future NBiS conferences.

June 2008

Makoto Takizawa
Leonard Barolli
Tomoya Enokido

Organization

Conference Chairpersons

Makoto Takizawa	Seikei University, Japan
Leonard Barolli	Fukuoka Institute of Technology, Japan

Program Committee Chair

Tomoya Enokido	Risho University, Japan

Web Management Chair

Alixier Aikebaier	Tokyo Denki University, Japan

Program Committee

Markus Aleksy	AAB Corporate Research Center, Germany
Irfan Awan	University of Bradford, UK
Bhed Bahadur Bista	Iwate Prefectural University, Japan
Goutam Chakraborty	Iwate Prefectural University, Japan
Kuo-Ming Chao	Coventry University, UK
Misbah Deen	Keele University, UK
Alex Delis	Polytechnic University, USA
Andrei Doncesku	University Paul Sabatier, France
Arjan Durresi	Indiana University-Purdue University Indianapolis, USA
Thomas Grill	University of Linz, Austria
Lin Guan	Loughborough University, UK
Takahiro Hara	Osaka University, Japan
Yoshiaki Hori	Kyushu University, Japan
Hui-huang Hsu	Tamkang University, Taiwan
Jinn-Ke Jan	National Chung Hsing University, Taiwan
Ismail Khalil Ibrahim	Johannes Kepler University Linz, Austria
Hiroaki Kikuchi	Tokai University, Japan
Akio Koyama	Yamagata University, Japan
Vincent Lee	Monash University, Australia
Yinsheng Li	Fudan University, China
Wen-Yang Lin	National University of Kaohsiung, Taiwan
Giuseppe De Marco	Toyota Institute of Technology, Japan
Wenny Rahayu	La Trobe University, Australia

Michel Raynal	IRISA-INRIA, France
Fumiaki Sato	Toho University, Japan
Nobuyoshi Sato	Iwate Prefectural University, Japan
Elhadi Shakshuki	Acadia Univiversity, Canada
Yoshitaka Shibata	Iwate Prefectural University, Japan
Timothy K. Shih	Tamkang University, Taiwan
David Taniar	Monash University, Australia
Nguyen Manh Tho	Vienna University of Technology, Austria
Minoru Uehara	Toyo University, Japan
Fatos Xhafa	Polytechnic University of Catalonia, Spain
Chu-Sing Yang	National Su Yut-Sen University, Taiwan
Muhammed Younas	Oxford Brookes University, UK

Referees

Markus Aleksy	Lin Guan	Fumiaki Sato
Irfan Awan	Takahiro Hara	Nobuyoshi Sato
Bhed Bahadur Bista	Hui-huang Hsu	Elhadi Shakshuki
Leonard Barolli	Timothy K. Shih	Yoshitaka Shibata
Valbona Barolli	Hiroaki Kikuchi	Kuei-Ping Shih
Santi Caballe	Akio Koyama	Takuo Suganuma
Goutam Chakraborty	Tsuneo Kagawa	Makoto Takizawa
Kuo-Ming Chao	Kuan-Ching Li	David Taniar
Giuseppe De Marco	Wen-Yang Lin	Dian Tjondronegoro
Mieso Denko	Takuo Nakashima	Minoru Uehara
Andrei Doncesku	Hiroaki Nishino	Fatos Xhafa
Arjan Durresi	Vamsi Paruchuri	Muhammed Younas
Tomoya Enokido	Wenny Rahayu	Qing-An Zeng

Table of Contents

NBiS-2008 Keynote Talk

Wireless Networks

Heterogeneous Networks

Ad-Hoc Networks

P2P, Grid and Internet Computing

Ad-Hoc and Sensor Networks

Intelligent Algorithms and Systems

Secure Systems and Applications

Network Tools and Architectures

Patterns in Internet Architecture

Arjan Durresi

Department of Computer and Information Science
Indiana University Purdue University Indianapolis
durresi@cs.iupui.edu

Abstract. The networking research community is working to design the Next Generation Internet, which will meet the needs of the twenty-first century. The Next Generation Internet has to be commerce friendly and secure. It should allow receivers to set policies for how and where they receive their information. The Next Generation Internet should be designed for mobile objects. Naming, addressing architecture, and routing have to be such that these objects can move and decide how and where they want to receive their Internet traffic with full rights of privacy of their location, if desired. Most of such capabilities turn out to be mostly a consequence of the network architecture. Therefore, it is adamant a better understanding of issues related to Internet architectures. In this talk, I will identify several architectural patterns that have emerged from the Internet evolution, from its beginning until now. Furthermore, we will discuss about lessons that could be learned from these patterns in network architecture.

M. Takizawa, L. Barolli, and T. Enokido (Eds.): NBiS 2008, LNCS 5186, p. 1, 2008.
© Springer-Verlag Berlin Heidelberg 2008

GAMesh: Automatic Placement of Wireless Mesh Nodes Via Genetic Algorithms

Giuseppe De Marco

Toyota Technological Institute, Tenpaku-Hisakata 2-12-1, Nagoya 468-8511, Japan
demarco@toyota-ti.ac.jp

Abstract. Wireless mesh networks are a profitable application of ad-hoc networks for their low installation cost and high performance properties. Here, we are concerned with the problem of how the mesh nodes should be placed in a given service area to satisfy some network constraints. We investigate this problem by means of genetic algorithms, which are used in order to find optimal topologies of a mesh network. The resulting algorithm has been called as GAMesh, and it is a numerical framework where the design constraints of the network, such as the connectivity and the traffic constraints, can be easily written by means of a scalar fitness function. We cast the problem in an innovative graph drawing technique, which furnishes good solutions in a reasonable computational time. Differently from other solutions, GAMesh explores all the points in the discretized plane. We show GAMesh in the context of infrastructure based wireless mesh networks, typically used by wireless internet service providers, which must minimize cost and maximize the performance.

1 Introduction

Definitions

1. Mesh Node (MN): A wireless router which can route the traffic of a set of mesh clients.
2. Mesh Portal (MP): An MN with wired connection to the Internet.
3. Mesh Client (MC): User device connecting to the mesh network through an MN.
4. Access Network(AN): The single-hop network composed of some MCs connected to one MN.
5. Backhaul Network(BN): The multi-hop network composed of the totality of MNs.

Usually, wireless mesh (WM) networks are intended for extending broadband Internet connection to locations where it is weak or absent [1] [9]. Nevertheless, mesh networks are economical solutions for developping countries and small wireless Internet providers which cannot afford the expensive deployment of cable-based connections. Two types of mesh networks have been proposed so far, namely Municipal WM (MWM) and Infrastructure-based WM (IWM). In

M. Takizawa, L. Barolli, and T. Enokido (Eds.): NBiS 2008, LNCS 5186, pp. 2–11, 2008.
© Springer-Verlag Berlin Heidelberg 2008

the first type, a number of MNs are deployed in a given area on the top of buildings or on street lamps. Residential users as well as nomadic users connect to the network by means of off-the-shelf wireless NICs. In the second types of mesh nets, the user installs on the top of her/his building an antenna pointing toward the "best" MN. The latter type is clearly an infrastructure model, where the operator control precisely the location of the accesses. MNs are usually multi-radio nodes in order to improve the capacity of the network. The user traffic is routed toward the "best" MP by selecting radio channels orthogonal with those used for the access.

An important design parameter of such network is the density of wireless nodes to deploy, i.e. how many MNs one should deploy in order to form a connected network given a target user coverage. Its value depends roughly on the transmission range of each MN and MC, and theoretical results from ad-hoc network community can help estimate the total number of MNs. [3,6,7,8]. However, the estimate does not provide the designer with the physical location of MNs, especially in the case where constraints such as interference and traffic demands must be satisfied. To the best knowledge of the author, only one paper deals with the problem of the placement of MNs in a given area. In [2], the authors consider a set of Candidate Sites (CS) for the location of MNs and a set of Test Points (TP), which practically represent the positions of potential users. An ILP formulation of the problem provides the minimum number of CSs needed to cover all the TPs. The CS set is assumed to be given, e.g. it can be derived by means of a measurement campaign on feasible installation sites. This step can be quite expensive, especially for large networks. In this paper, we consider the dual of the work in [2]: All the sites on a given area are CSs. Clearly, this assumption requires a discretization of the area, and we will assume a grid of cells, where every cell is occupied or not by an MN [1]. Here, we cast the planning problem as a graph drawing problem. The WM net is represented as a graph in the discretized Euclidean plane, where every vertex represents an MN. The problem is how and where to place vertexes of the graph in the cells of the grid. We call our algorithm Genetic Algorithm based Mesh planner (GAMesh). In particular, the solution of the problem is a candidate graph, which is an *individual* in the genetic parlance. This individual is the best result of a searching process among all possible candidates. It is worth noting that the present work deals only with IWM networks. The more general case will be analyzed in further studies.

The structure of the paper is as follows. In Section 2, we explain how the individuals (graphs) are coded in genes, which are then used for the evolution process of the GA. In particular, we explain the three fundamental operators of GA, selection, crossover and mutation. In the same section, we discuss how to choose the objective function in order to fairly represent the constraints of the WM net. In Section 4, we demonstrate the ability of GAMesh by means of simulations. Conclusions and discussions are given in Section 5.

[1] This formulation is also desirable, because it can be extended to include details of radio propagation models.

2 GAMesh

The goal of GAMesh is to find optimum layouts of the WM network, which satisfy some design constraints. In GA parlance, these layouts are solutions or rather best individuals for the optimization problem, which are the result of three fundemental operations: selection,crossover and mutation. These individuals are the best "fit" w.r. to some objective function to be minimized/maximized. In our problem, we consider a maximization problem. Since we are looking for best placements of inter-connected MNs, our individuals are graphs.

The particular characteristic of our problem is that if we used standard genetic coding such as strings of integer or boolean values, an interesting problem could arise during mating and crossover. It can be states as follows: Good parents can generate always bad children, or rather unfit individuals. This is contrary to the general schema mechanism of GA, which states that if a parent is a good individual with respect to its fitness value, its goodness should be transferred to the prole. Therfore, we use a genetic coding similar to TimGA [5]. The main difference with our work is that TimGA uses an aesthetic criteria to optimize the layout of the graph. In other words, in TimGA, and other classical graph drawing techniques, links among nodes are established a priori, while in GAMesh they depend on the layout itself.

The WM network is supposed to be deployed in a squared area A. The area is divided in $M = \lceil \frac{A}{\ell^2} \rceil$ cells, where ℓ is the length of a cell side. The BN is represented by a graph $G = (V, E)$, with V the set vertexes and E the link set. A chromosome is represented by a triple $s = (\mathbf{A}, \mathbf{x}, \mathbf{y})$, where $\mathbf{A}$ is the $n \times n$ adjacent matrix of a graph and n is the number of MNs[2]. The vectors $\mathbf{x}, \mathbf{y} \in \mathbb{N}_m^n$ are the coordinate of the mesh points, i.e. $\mathbf{x} = (x_1, \ldots, x_n)$, $1 \leq x_i \leq m$, $\forall i, 1 \leq i \leq n$.

Further definitions are in order. The AN is represented by a graph $G_a = (V_a, E_a)$, where $V_a = V \cup V_t$, $n_a = |V_a|$ and V_t is the set of nodes representing TPs or MCs. The set E_a is build by associating an MC one and only one of the nearest MN of the BN. In general, an MC can have links to multiple MNs, but for now we consider only single associations. Therefore, the graph G_a is obtained by 1) eliminating links between MCs, because we are supposing that mesh clients cannot be in ad-hoc mode one another, and 2) by eliminating multiple links from an MC to many MNs, i.e. we suppose that the MCs are uniquely associated with one and only one MN. The adjacency matrix associated with G_a is the $(n_a + n) \times (n_a + n)$ matrix, $\mathbf{A}_{cc} = (s)_{ij}$. The degree of an MN $v \in V$ is $D_i = \sum_{t=1}^{n} a_{it}$, where a_{ij} are the elements of $\mathbf{A}$. The degree of the same node in G_a is labeled as d_i.

The edges of graphs G and G_a are selected according to the radio characteristics of the associated channels. For example, two nodes $v, u \in G$ will be connected by an edge if their euclidean distance r is s.t. the path loss is less than a given threshold L_{th}.

[2] The adjacency matrix $\mathbf{A} = (a)_{ij}$ of G is the $n \times n$ matrix where $a_{ij} = 1$ if there is a link between (mesh) node i and j. However, since A is sparse, the simulation code use an efficient representation of $\mathbf{A}$.

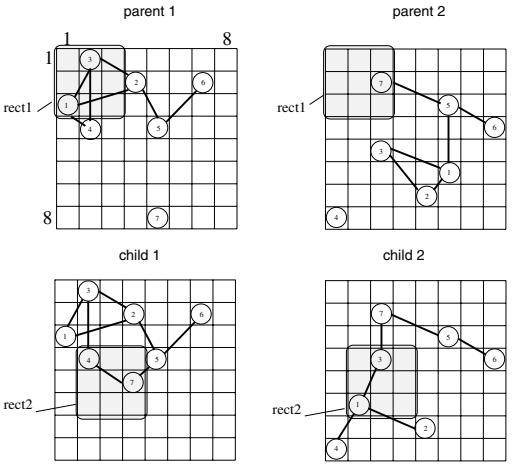


Fig. 1. Example of crossover operations of GAMesh

2.1 Crossover

The crossover operations cannot be spun in the classical way, i.e. by exchanging parts of the strings which code the chromosome, because here we have not strings at all, but triples. Exchanging substring of vectors $\mathbf{x}, \mathbf{y}$ will lead to a misleading behavior of the GA. A simple example, not shown here for space constraints, could readily convince of this fact. Therefore, the crossover takes place by selecting a square of R cells inside two parents and two children. If the squares are empty, the respective parents go unchanged throughout the crossover. The node positions inside these squares are exchanged between parents and children. An example is in order. In Fig. 1, we have two candidate solutions, i.e. parent1 and parent2. Two random squares, rect1 and rect2, are drawn within the parents and the children, respectively. The nodes which fall inside the parent squares are transferred to the children. Nodes of rect1 are transferred into rect2 of child2, and viceversa. For example, in the child1, the node 1 of parent2 has been transferred inside rect2. Let us note that child1 and child2 have an improved connectivity with respect to that of the respective parents.

2.2 Mutation

We designed four types of mutation operations, described in the following list.

1. SingleMutate: Picks up a single node and move it in a random position in the plane.
2. RectMutate: Picks up two squares in the plane and move the content of one square to the other one.
3. SmallMutate: Picks up a single node and move it by a fixed step (=3).
4. SmallRectMutate: Picks a square of nodes and move it by a fixed step (=3).

These operations are selected by drawing a discrete random variable X with distribution p_i, $i \in [1,4]$. When $X = i$, the mutation operator is accordingly selected. For example, if $X = 1$ the SingleMutate operator is selected. In our experiments, we found empirically that a quite acceptable distribution should give more chances to the second and the fourth operator. In particular, we set $p_1 = 0.1, p_2 = 0.3, p_3 = 0.1, p_4 = 0.5$.

2.3 Objective Function

The objective function is evaluated for every chromosome, and it is a function of the independent variables. With a bit of terminology abuse, it is also called the fitness. The only independent variables are the positions of the nodes. Other graph properties, such as the size of the giant component[3], $|C_0|$, or the node degree, are functions of the independent variables. Constraints on the graph properties can be embodied in the objective function. A first criteria to build the fitness funciont is the evaluation of the the giant component of a chormosome, because we desire connected BNs. The level of acceptance of individuals with respect to their giant component size is set by a threshold value, T_h. If we set T_h too low, we risk to favor a lot of bad individuals, and on the other hand, if we set T_h too high, we risk to favor few super individuals, thus reducing the exploration power of the GA. In our experiment, we set $T_h = n$, i.e. the number of MNs. Another factor which affects the goodness of an individual is its ability to cover a number of TPs as high as possible. We measure this number by simply counting how many MCs are in radio communication with a particular MN, or, equivalently, by counting the number of isolated TPs in G_a. Accordingly, our objective function is a function of three basic graph properties:

$$f(s) = \begin{cases} \frac{|C_0|}{n}, & \text{if } |C_0| < T_h \\ -\frac{\max(D)}{n-1} + 2(2-p_0)^2, & \text{if } |C_0| \geq T_h \end{cases} \tag{1}$$

where p_0 is the percentage of isolated TPs in G_a. The fitness function is less than 1 if the graph is not connected, i.e. if the giant component is less than the threshold. The expression of $f(s)$ after the threshold has been chosen to be a decreasing function of p_0. In this way, the more nodes are isolated, the lesser the fitness. The meaning of the maximum of the degree of G in (1) will be clearer in the following section, by considering interference arguments. Accordingly[4], the range of the fitness is in the interval $[1, 8]$.

The algorithm is terminated when no improvement is noted from one generation to another. A controllable way to check this improvement is to accumulate a certain number of samples of the best individual into a window of values as the generation process proceeds. If the relative difference of the fitness of the individuals within this window is less than a given value, the algorithm has reached or is near the optimum and it terminates. In our experiments we set the window to 40.

[3] The giant component is the largest connected component of a graph G, where a component is a maximal connected subgraph $C \subseteq G$.

[4] In particular, $\max(|C_0|) = n$, $\max(D) = n - 1$ and then the assertion.

3 Constraints

In order to take into account others constraints of the network, we can rewrite the fitness function by introducing a penalty function[5]. The equation in (1) becomes:

$$f(s) = -\frac{\max(D)}{n-1} + 2(2-p_0)^2 + \sum_{i=1}^{O} \epsilon_i \Phi_i \,, \quad |C_0| \geq T_h \tag{2}$$

where Φ_i is the function representing the i-th constraint, O the number of constraints and ϵ_i can be $\mp$, according to the monotonicity of Φ_i. In the following, we define the total bit rate of the access network of an MN i as $f_i = \lambda_0 d_i$, where λ_0 is the bit rate guaranteed to the MC and d_i is the degree of node i with respect to G_a. A first constraint is on the traffic transported between two MNs, i and j. By assuming a pefect link scheduling algorithm which activates one backhaul link at time, we must satisfy the following constraint:

$$f_i + f_j \leq \rho_{ij} a_{ij}, \quad \forall i, j \in V \tag{3}$$

where ρ_{ij} is the wireless capacity of the link connecting node i and j. By summing up for all indexes, and assuming that $\rho_{ij} = \rho$, we have:

$$\sum_{(i,j)\in E} (f_i + f_j) \leq \rho n \overline{D} \,, \tag{4}$$

where $\overline{D}$ is the mean degree of G. Then, the constraint on the traffic can be written as:

$$\Phi_1 = \left(1 - \frac{1}{\rho n \overline{D}} \sum_{(i,j)\in E} (f_i + f_j) \right) \,, \tag{5}$$

and $\epsilon_1 = 1$. Another constraint is related to the wireless capacity, θ_v of the access interface of every MN node, $v \in V$. The user traffic toward a particular MN cannot exceed the capacity of its access channel. If we suppose that $\theta_i = \theta$,$\forall i$, we can write the second contraint as:

$$d_i \lambda_0 \leq \theta, \quad \forall i \in V \,, \tag{6}$$

by summing up for all i, we have:

$$\Phi_2 = \left(1 - \frac{\lambda_0 \overline{d}}{\theta} \right) \tag{7}$$

and $\epsilon_2 = 1$.

[5] Penalty functions are not well accepted in the GA community because, besides their lack of generality, they tend to generate a non neglegible number of unfeasible solutions. However, for weak constrained problems as in this case, penalty function is quite a controllable technique to cast constrained problems in unconstrained ones.

3.1 Other Constraints: Interference

In general, radio interference limits the capacity of the radio links. The interference depends on the MAC protocol and the net architecture. For CSMA/CA protocols like IEEE 802.11, it is well known that the throughput which a given node can achieve is limited by the number of transmission which are 1-hop and 2-hop away. However, if we consider directional antennae, the interference is minimized. Here, for simplicity, we suppose that the IWM network uses directional antennae, but the general case can be easily formulated. Although, this solution increases the cost of the node, the use of directional antennae in WM nets is not an unrealistic solution, but rather a good compromise between cost and performance, as witnessed also by recent studies [4] [10]. Clearly, the lower is the number of antennae, the lower is the cost of the MN. The price of this solution is that the network is less redundant. Similarly, for multi-radio nodes, we would use a number of radio interfaces as low as possible. In conclusion, these costs can be minimized by simply minimizing the maximum of D, by implicitly assuming that the physical meaning of D depends on the type of WM.

4 Simulations

GAMesh has been designed in MATLAB, because of its intensive use of matrix operations. The spatial distribution of the MC or TPs in the plane is assumed to be uniform in a given part of the whole service area. For instance, the position of the i-th MC is assumed to be drawn from a continuous bi-dimensional uniform r.v. $U = (u_1, u_2) = (\mathcal{U}(l_1), \mathcal{U}(l_2))$, where $\mathcal{U}(x)$ is a uniform r.v. in the range $[0, x]$. For the path loss thresholds, we set $L_{\text{Th}} = 55$dB and $L_{\text{Th}} = 48$dB, for the BN and the AN, respectively.

In this section, we give some preliminary results of GAMesh in order to show its usage. The parameter used in this test are shown in Table 1 and 2. We experienced that the value of 50 for the initial population is enough to avoid the premature convergence of GAMesh. In order to judge the goodness of GAMesh, we use the following metrics:

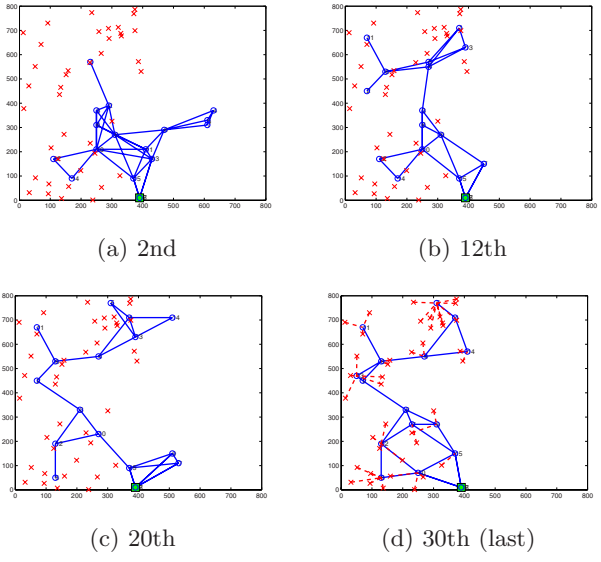
- Max of the fitness, $\max(f)$ and the percentage of converged individuals.
- Minimum of the number of isolated nodes, $\min(n_0)$.

Table 1. GAMesh parameters

Initial pop.	Mutation rate	Crossover square(m)	Grid step(m)
50	0.02	8	20

Table 2. Simulation parameters. Bit rates are in Mbps

Service Area(m^2)	Position of MCs	n	MPs	λ_0	θ	ρ
$L^2 = 800^2$	$l_1 = \frac{L}{2}, l_2 = L$	15	1	0.5	11	54



(a) 2nd (b) 12th

(c) 20th (d) 30th (last)

Fig. 2. Sample best topologies found by GAMesh during the evolution process

and the following metrics related to the success rates w.r. to the constraints:

- Median of $(1 - T_v)$, where T_v is the number of backhaul traffic constraint violations.
- Median of $(1 - T_{av})$, where T_{av} is the number of access traffic constraint violations.

Clearly and ideally, the optima should be reached whenever the following equalities hold: $\min(n_0) = 0$, and the medians of success rates equal to 1.

As benchmark example, we put all the TPs on the left side of the plane, in order to test whether GAMesh can correctly "bias" the search space toward the correct place in the plane. As it can be seen in Fig. 2, GAMesh correctly places the MNs close to TPs. For sake of clarity, we shown only four sample of optimal topologies, i.e. best individuals found at different steps of the evolution process. It is worth noting that since the fitness depends on the degree of G, the resulting topology will "resemble" a tree, i.e. a graph with a degree as low as possible. In fact, GAMesh revealed a self-repulsion property: The closer nodes are, the higher the number of links, and then the degree, which in turn decreases the fitness. This can be see by comparing the sample topologies of Fig. 2. The same happens for other runs of GAMesh, which are not reported due to space limits. In some runs, it can be observed that a number of MNs lower than the user-defined one could be used. The automatic setting of the minimum number of MNs is left for further improvements of GAMesh. In Fig. 3-a, we have the evolution of the fitness values for a single run of GAMesh. After roughly 150 iterations, GAMesh touches the maximum of the fitness, i.e. we are near the optimum solution. It can be seen that the number of converged individuals, i.e. the fraction of the

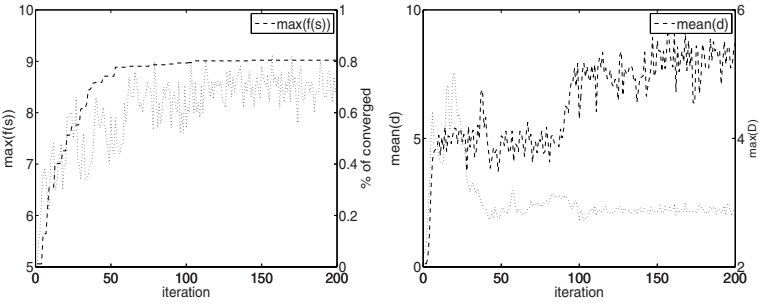


Fig. 3. Metrics

Table 3. GAMesh iterations (ite) for different values of the grid size

grid size	$\overline{ite}$	$\sqrt{\mathrm{Var}(ite)}$
50	158.60	42.3089
80	180	80.1
100	179	60.0462

population whose fitness value equals the maximum, is not a good indicator of the termination condition, as witnessed by the high variability of the related function in Fig. 3-a.

The degree D is as low as possible, and its max value is about 5, as shown in Fig. 3-b. The degree stricly depends on the number of nodes and the level of redundancy desired by the designer. For example, in Fig 2-c, some links could be removed without violating the constraints, or, similarly, a lower number of nodes could be used. Eventually, for the topologies of Fig. 3, the constraints are always satisfied, i.e. medians of T_v and T_{av} are always 1. In particular, the median of n_0 decreases towards zero in roughly 60 iterations. These simulations confirm our expectations. GAMesh can successfully find optimum constrained layouts of WM nets in a reasonable computational time, which depends on the grid size, n and n_a. GAMesh slighty depends on the grid size, because the grid size impact the range of $\mathbf{x}$ and $\mathbf{y}$, only and not the size of the matrices. To test this feature, we performed a Monte Carlo simulation for a particular value of the grid size. In Table 3, we see that the number of iterations, ite, does not change substantailly, given that the sampling window of the fitness values is 40, as explained in section 2.3.

5 Conclusions

In this paper, we introduced a novel approach to the design of the topology of WM networks. The low cost of such networks should encompass also the design cost. In particular, we concerned with the physical placement of MNs

within a given service area. The core of the algorithm is based on the theory of graph drawing. A novel genetic algorithm has been designed for this purpose. Simulation results of the present work shown that GAMesh can find candidate solutions in a reasonable computational time which translates qualitatively in about 100 iterations for $10 - 20$ nodes. On commodity desktop computers, this means $10 - 20$ seconds. Additional constraints can be inserted in the fitness function in order to include other details of the wireless network, such as the interference and the wireless gains As further investigations, we envision a multi-objective optimization version of GAMesh in order to take into account also the total cost of installation of MNs, and a deeper analysis of the performance of GAMesh in more complicated scenarios as well.

References

1. Akyildiz, I., Wang, X., Wang, W.: Wireless mesh networks: a survey. Computer Networks 47(4) (March 2005)
2. Amaldi, E., Capone, A., Cesana, M., Malucelli, F.: Optimization models for the radio planning of wireless mesh networks. In: Akyildiz, I.F., Sivakumar, R., Ekici, E., Oliveira, J.C.d., McNair, J. (eds.) NETWORKING 2007. LNCS, vol. 4479, pp. 287–298. Springer, Heidelberg (2007)
3. Bettstetter, C., Hartmann, C.: Connectivity of wireless multihop networks in a shadow fading environment, September 2005, vol. 11(5) (2005)
4. Das, S.M., Pucha, H., Koutsonikolas, D., Hu, Y.C., Peroulis, D.: Dmesh: Incorporating practical directional antennas in multichannel wireless mesh networks. IEEE Journal on Selected Areas in Communications 24(11), 2028–2039 (2006)
5. Eloranta, T.: Timga: A genetic algorithm for drawing undirected graphs. Divulgaciones Matematicas 9(2), 155–171 (2001), http://www.emis.de//journals/DM/v92/art5.pdf
6. Gupta, P., Kumar, P.R.: Critical power for asymptotic connectivity in wireless networks. Stochastic Analysis, Control, Optimization and Applications: A Volume in Honor of W.H. Fleming, 547–566 (1998)
7. Hicham Khalife, N.M.
8. Marco, G.D., Postiglione, F., Longo, M.: Connectivity of ad hoc networks with link asymmetry induced by shadowing. IEEE Communications Letters 11(6), 495–497 (2007)
9. Nandiraju, N., Nandiraju, D., Santhanama, L., He, B., Wang, J., Agrawal, D.: Wireless mesh networks: Current challenges and future direction of web-in-the-sky. IEEE Wireless Communications, 79–89 (2007)
10. Tang, J., Xue, G., Chandler, C., Zhang, W.: Interference-aware routing in multihop wireless networks using directional antennas. In: Proceedings of IEEE INFOCOM 2005, March 2005, vol. 1(13-17), pp. 751–760 (2005)

Context-Aware Loading for Mobile Applications

Markus Aleksy[1], Thomas Butter[2], and Martin Schader[2]

[1] ABB Corporate Research Center Germany
Industrial Software and Applications
[2] University of Mannheim
Department of Information Systems

Abstract. The dynamic adaptability to the current context of the user is one of the most important challenges to a context-aware mobile application since context decides on the available services. In this paper, we present an approach, which supports dynamic loading of the different elements of a mobile application. This includes context sensors, adapted input and output components, or the business logic required for the use of a context-aware service.

1 Introduction

A main characteristic of mobile context-aware applications is their dynamics, which requires that the restricted resources of the mobile device on which such an application is executed are used efficiently. The reason is that the current context of the user, such as the present location, decides on the available services. The framework such an application is based on must therefore both have a flexible functionality regarding context discovery and processing and have the ability for the efficient discovery, management, and use of location-based services. Scheer et al. [22] distinguish four kinds of context information:

- Environmental Context
  This type of context information includes the environmental conditions of the current location, such as the weather or the ruling lighting conditions.
- Activity Context
  This category describes the present activity of the user. Professional as well as private activities, such as journeys or shopping acts may be noted here.
- Temporal Context
  The recording of the current temporal conditions represents the third class of context information. Besides the current time, also other temporal factors, such as the season, can be found in this category.
- Personal Context
  This category includes the individual preferences and characteristics of the user. The range shown here is very wide and must take into account a variety of factors. Aspects, which can be handled relatively simply such as smoker/non-smoker just like more complex points of view, such as the state of health or the interests or hobbies of the user may be captured.

M. Takizawa, L. Barolli, and T. Enokido (Eds.): NBiS 2008, LNCS 5186, pp. 12–20, 2008.
© Springer-Verlag Berlin Heidelberg 2008

Another distinction is proposed by Korpipää et al. [17]. It is based on the Blackboard approach [6] and orientates itself at the categorization of the context information of a mobile application according to the abilities of the mobile terminal, or, more exactly, on its Symbian platform. The authors mention the categories location, time, environment, user, as well as device.

Some of this context information can be captured by the mobile device itself if the device exhibits corresponding technical prerequisites. The current location, the line of sight, the datebook, etc. are to be noted here. Since the mobile terminals currently offered on the market only have a relatively low number of context sensors at their disposal, some part of the context information must be collected by the context sensors on the server side. In this category the current weather conditions or the season are to be found.

The restricted resources of the mobile devices presently offered on the market represent a huge problem, which must be handled when realizing the necessary flexibility of a mobile application. When compared to conventional computer systems, many of today's mobile devices, e.g., PDAs or mobile telephones show restrictions such as low computer power and storage capacity, limited input/output functionality, low bandwidth, high latent periods when using wireless communication, or restricted power supply. Despite the ongoing further development and improvement of mobile systems, there will remain a discrepancy between them and their static counterparts [21].

It is therefore necessary that both requirements, i.e., achieving flexibility of the mobile application as well as handling the resources of the mobile device carefully are taken into account.

The introduced approach has resulted from work in the context of the Generic Environment for Mobile business (GEM) framework and uses some of its components. A short summary of the GEM client framework is therefore carried out in the following section.

2 GEM Client Framework

An architecture for the development of context-aware mobile applications must fulfill a variety of different tasks at the same time. Next to the recording and management of different context information and the efficient management of resources of the mobile device the search of the users for services must be designed with maximum flexibility. Under consideration of the conditions and restrictions of the mobile device, the found results must be presented suitably to users. Furthermore, communication means should be designed flexible to make the information exchange with existing information systems easier.

To do justice to these claims, the GEM client framework we developed is composed of the following components (see figure 1):

- Context Manager
  This component is responsible for the recording and management of the current values of the individual context sensors. In addition, it provides an interface, which can be used by the Service Discovery Architecture as well as by other context-aware applications to reach the present context information.

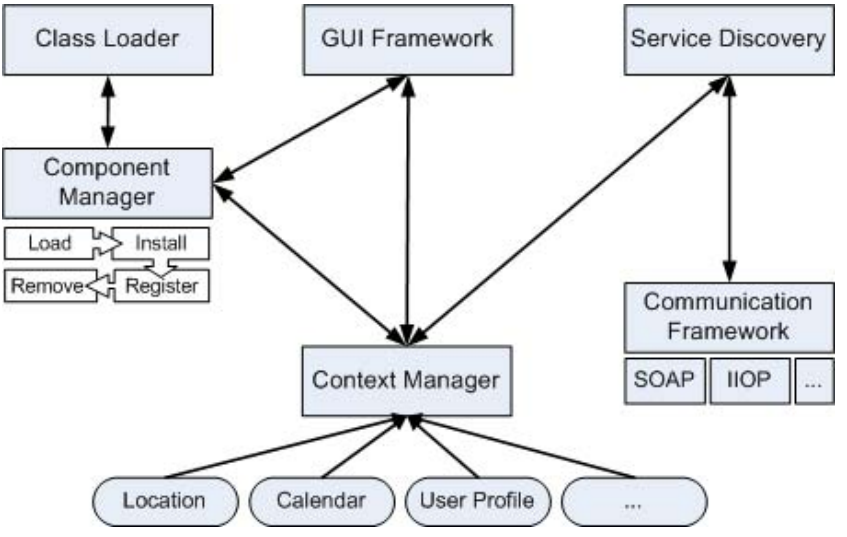


Fig. 1. Overview of the GEM Client Framework

– Component Manager
  The Component Manager administrates the life cycle of any application component, which is used on the mobile device. It is supported by a Class Loader component. This component allows for dynamic loading of application components, which provide a specific business logic or offer a user interface specifically adapted to the needs of the just requested service.
– Service Discovery
  This part of the whole architecture is responsible for the on-finding of services. It complements user requests by the current context information. If the result of the enquiry should require the installation of a new component, then it goes back to the functionality offered by the Component Manager. An exact representation of the Service Discovery Architecture can be found in [1].
– Context-aware Graphical User Interface (GUI) Framework
  This element of our architecture is helpful to bridge heterogeneity of the various devices. The simultaneous realization of both device-independent and at the same time user-friendly GUIs is supported by this framework. A detailed description is given in [4].
– Generic Communication Framework
  To integrate mobile context-aware applications into existing business applications smoothly, the offered communication infrastructure must be very flexible. Besides the default support of protocols frequently used in the area of Enterprise Computing (such as SOAP [9] or the Internet Inter-ORB Protocol (IIOP) [18]) it should also offer the genericity, which allows the employment of further communication protocols.

3 Context-Aware Loading for Mobile Applications

The Context Manager component is responsible for the recording and passing of context information. Should the current context of the user change, e.g., through the alteration of his location, then this information is reported to the Context Manager by the corresponding context sensor. The manager then informs all context-aware applications registered with it about the occurred changes. If this event should be judged by the application as relevant, then this can result in different actions. These actions are discussed more exactly in the following sections.

3.1 Reloading Context Sensors

As mentioned already, both the number and the type of the context sensors residing on a mobile device can be very different. To avoid the restrictions of storage capacity of the mobile device, only certain context sensors should be installed per default. Localization, user profile, time planning, etc. are examples. When required, additional context sensors can be loaded and installed via an existing network connection.

As not all context sensors do always have to be available, an alteration of the current context can imply that reloading additional context sensors is necessary. For example, per default, one sensor based on the Global Positioning System (GPS) [14] can be used for the recording of the current location of the mobile device or the user. Although GPS has gained wide acceptance out-of-doors, there are different approaches, which try to cover the areas in which GPS does not work due to missing intervisibility. Therefore, different projects try to use other existing communication infrastructures, such as GSM radio masts, WiFi or Bluetooth access points, etc. (cf. [24], [2], [20], [5], [25], [15] and others).

If the user of the mobile device should not be in the position to be identified any longer via GPS since he has just entered a building, for example a corresponding indoor localizing sensor can be reloaded and installed.

3.2 Reloading Input/Output Components

The alteration of the current context of a mobile application can lead to a change in the preferential manner of input or output. Should the current context information of the user show that he is just walking, the output could be adapted correspondingly by the GUI using larger fonts and icons. To reach a high flexibility of the visual representation, our framework bases on the use of the XML User Interface Language (XUL) [26] and Cascading Style Sheets (CSS) [3]. This way, the design of the user interface is founded on three different elements: the structure of the GUI, the corresponding actions/behavior, as well as on the actual representation itself. Furthermore you can download the GUI either in the form of descriptions or as compiled code. While the first variant is showing a greater flexibility regarding possible changes, the second alternative can be executed considerably faster [4].

The output on the mobile device can just the same be adjusted, e.g., if it is noticed that the mobile user is just situated in a moving vehicle. Here, the device could switch over from the default output on the display to an optional acoustic output. Since these input or output types are context-dependent, it makes sense to download them onto the mobile device first when actually required.

3.3 Reloading Service Components

Furthermore, context alterations can require that a new search for suitable context-specific services must take place. This search is supported by the Service Discovery Architecture (cf. section 2). Should it turn out that a found service needs a specific GUI or requires the installation of a dedicated business logic, the corresponding components will be reloaded and installed.

3.4 Useful Design Patterns

During realization of the context-dependent reloading of the various elements of a mobile application different design patterns may be utilized. Such patterns are simple, concise and already proven solutions for programming tasks frequently appearing in practice.

Design patterns, such as Lazy Acquisition [16] or Partial Acquisition [13] aim at allocating the resources needed by an application to the latest possible time or gradually. The Lazy Acquisition design pattern is based on the components roles: User, who wants to use a certain resource, Resource, offering a certain functionality, Virtual Proxy, having the same interface as the resource it represents, thus being able to function as a local substitute, and Resource Environment, which is responsible for management of the respective resources. With the exception of the Virtual Proxy, the Partial Acquisition design pattern is based on identical roles.

The Lazy Load design pattern described by Fowler [7] is based on the idea that an object does not immediately load all data but only those, which it immediately needs. However, it "knows" how it can reach the remaining data so that these can be reloaded when required.

3.5 Deallocation of Reloaded Elements

In the case of mobile clients, the use of the Evictor pattern [11], [12] seems to us the solution most suitable for the management of reloaded elements. This approach tries to cope with the objective of using services efficiently by Monitoring. Every time an access to a reloaded element is carried out, a marker is put. Elements that were not accessed for quite a long time (least recently used – LRU) or that are only rarely used (least frequently used - LFU) are candidates for deallocation. The release of these elements can be happen either periodically or upon request.

4 Design Considerations for Context-Aware Loading

The ability to reload GUI elements, scripts, classes, or components constitute the core functionality of the Class Loader component. It should, therefore, not be equated directly with the Java class `ClassLoader`. In the following sections, the essential features of our design and the message flows between the individual components are described.

4.1 Design Issues

In principle, there exist different approaches how one could realize context-aware class loading. The first approach relies on the strict separation between the individual components. I.e., there is a Context Manager component, a Component Manager component, as well as a Class Loader component. This structure would meet the requirements of the Separation of Concerns aspect [23]. It is a key concept, which is of special importance for the development of software systems and aims at the clear separation of the different areas of responsibility. Different aspects of a software system are considered independently of each other and integrated at a later time.

In our case we have decided, however, to combine the Component Manager and the Class Loader components to conserve the restricted resources of a mobile device and to reduce the needed internal communication executed on the mobile device itself.

4.2 Message Flow

The UML sequence diagram in figure 2 illustrates the message flow in our case. A context-aware application can register with the `ContextManager` component by means of the method `addContextlistener()`. As soon as one of the context sensors has discovered that a change of context has occurred it notifies the

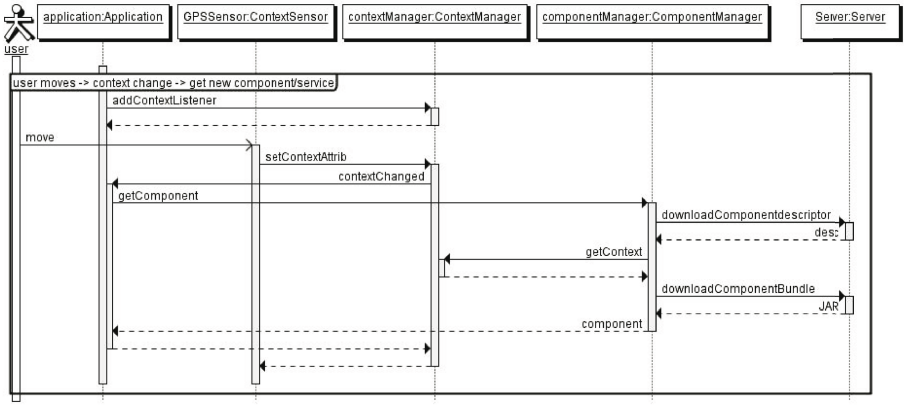


Fig. 2. A simplified example scenario presenting the context-aware class loading functionality

`ContextManager` about this through a call of the method `setContextAttr()`. The latter informs all components and applications registered at it with the help of method `contextChanged()`. The application can then use the functionality offered by the `ComponentManager` component to load a component dynamically. A call of the method `getComponent()` is the reason that, in the first step, `ComponentDescriptor` is downloaded. This object contains information about possibly suitable classes and about their respective restrictions. With an invocation of `getContext()`, the restrictions of the different classes are evaluated and the class with the best fit is selected. Then, the proper component is loaded, installed, and registered so that it can be used by the application.

4.3 Related Work

Mobility of code is used in many areas. Java applets and mobile agents belong to the best known ones. A survey on the different approaches to code mobility can be found in [8]. In the area of mobile devices, the OSGi platform [19] represents one of the most well-known frameworks, which have a class loader component. Hall [10] describes a Policy-Driven class loader, which separate the Policy Decisions from the Class Loading infrastructure. However, no publication in which such a generic possibility to realize context-aware loading is described is known to us.

5 Conclusions

The introduced approach makes it possible to realize context-aware applications even on mobile devices with severely restricted resources. Storage requirements can be kept low through our framework by having the necessary application components not being installed first of all on the mobile device but by downloading, installing, using, and releasing them upon demand, i.e., dependent on context. A main feature of the approach introduced by us therefore is the ability to load different elements of a mobile application at the proper point in time. These elements might be either context sensors or input or output components, as well as business logic components. The current user context alone decides whether and when any and which of these elements are transferred to the mobile device, respectively. In this way, it is possible to realize an application adapting itself to the changing and current needs of the user.

Acknowledgements

This work was funded in part by Deutsche Forschungsgemeinschaft.

References

1. Aleksy, M., Atkinson, C., Bostan, P., Butter, T., Schader, M.: Interaction Styles for Service Discovery in Mobile Business Applications. In: Proceedings of the 17th International Conference on Database and Expert Systems Applications (DEXA 2006) / 9th Workshop Network-Based Information Systems (NBiS 2006), Krakow, Poland, September 4-8, 2006. IEEE Computer Society, Los Alamitos (2006)

2. Bahl, P., Padmanabhan, V.N.: RADAR: An In-Building RF-Based User Location and Tracking System. In: Proceedings of the 19th International Conference on Computer Communications (Infocom), Tel Aviv, Israel, March 2000, vol. 2, pp. 775–784. IEEE, Los Alamitos (2000)
3. Bos, B., Lie, H.W., Lilley, C., Jacobs, I.: Cascading Style Sheets, Level 2, CSS2 Specification. W3C Recommendation, World Wide Web Consortium (W3C) (1998)
4. Butter, T., Aleksy, M., Bostan, P., Schader, M.: Context-aware User Interface Framework for Mobile Applications. In: Proceedings of the 27th IEEE International Conference on Distributed Computing Systems Workshops (ICDCSW-2007), Toronto, Kanada, June 25-29, 2007. IEEE Computer Society, Los Alamitos (2007)
5. Castro, P., Chiu, P., Kremenek, T., Muntz, R.R.: A Probabilistic Room Location Service for Wireless Networked Environments. In: Abowd, G.D., Brumitt, B., Shafer, S. (eds.) UbiComp 2001. LNCS, vol. 2201, pp. 18–34. Springer, Heidelberg (2001)
6. Engelmore, R., Morgan, T.: Blackboard Systems. Addison-Wesley, Reading (1998)
7. Fowler, M.: Patterns of Enterprise Application Architecture. Addison-Wesley Professional, Reading (2002)
8. Fuggetta, A., Picco, G.P., Vigna, G.: Understanding Code Mobility. IEEE Transactions on Software Engineering 24(5), 342–361 (1998)
9. Gudgin, M., Hadley, M., Mendelsohn, N., Moreau, J.J., Nielsen, H.F., Karmarkar, A., Lafon, Y.: W3C SOAP Version 1.2 Part 1: Messaging Framework (Second Edition), W3C Recommendation, 27 (April 2007), http://www.w3.org/TR/soap12-part1/
10. Hall, R.S.: A Policy-Driven Class Loader to Support Deployment in Extensible Frameworks. In: Emmerich, W., Wolf, A.L. (eds.) CD 2004. LNCS, vol. 3083, pp. 81–96. Springer, Heidelberg (2004)
11. Henning, M., Vinoski, S.: Advanced CORBA Programming with C++. Addison-Wesley, Reading (1999)
12. Jain, P.E.: Proceedings of 8th Patterns Languages of Programs Conference (PLoP 2001), Allerton Park, Monticello, Illinois, USA, September 11-15, 2001 (2001)
13. Jain, P., Kircher, M.: Partial Acquisition. In: Proceedings of 9th Conference on Pattern Language of Programs (PLoP 2002), Allerton Park, Monticello, Illinois, USA, September 8-12, 2002 (2002)
14. Kaplan, E.: Understanding GPS. Artech House Publishers (1996)
15. King, T., Kopf, S., Haenselmann, T., Lubberger, C., Effelsberg, W.: COMPASS: A Probabilistic Indoor Positioning System Based on 802.11 and Digital Compasses. In: Proceedings of the First ACM International Workshop on Wireless Network Testbeds, Experimental evaluation and CHaracterization (WiNTECH 2006), Los Angeles, CA, USA, September 2006. ACM, New York (2006)
16. Kircher, M.: Lazy Acquisition. In: Proceedings of 6th European Conference on Pattern Languages of Programs (EuroPLoP 2001), Irsee, Germany, July 4-8, 2001 (2001)
17. Korpipää, P., Mäntyjärvi, J., Kela, J., Keränen, H., Malm, E.J.: Managing context information in mobile devices. IEEE Pervasive Computing 2(3), 42–51 (2003)
18. Object Management Group The Common Object Request Broker: Architecture and Specification. Version 3.0.3. OMG Technical Document Number formal/04-03-01 (2004), ftp://ftp.omg.org/pub/
19. OSGi Alliance. About the OSGi Service Platform – Technical Whitepaper, Revision 4.1 (June 7, 2007), http://www.osgi.org/documents/collateral/OSGiTechnicalWhitePaper.pdf

20. Priyantha, N.B., Chakraborty, A., Balakrishnan, H.: The Cricket Location-Support System. In: Proceedings of the 6th Annual International Conference on Mobile Computing and Networking (MobiCom), pp. 32–43. ACM, New York (2000)
21. Satyanarayanan, M.: Mobile computing. IEEE Computer 26(9), 81–82 (1993)
22. Scheer, A.W., Feld, T., Göbl, M., Hoffmann, M.: Das mobile Unternehmen. In: Silberer, G., Wohlfahrt, J., Wilhelm, T. (eds.) Mobile Commerce-Grundlagen, Geschäftsmodelle, Erfolgsfaktoren, pp. 87–107. Gabler Verlag, Wiesbaden (2002)
23. Völter, M.: Server-Side Components–A Pattern Language. In: Proceedings of 6th European Conference on Pattern Languages of Programs (EuroPLoP 2001), Irsee, Germany, July 4-8 (2001)
24. Want, R., Hopper, A., Falcao, V., Gibbons, J.: The Active Badge Location System. ACM Transactions on Information Systems 10(1), 91–102 (1992)
25. Youssef, M., Agrawala, A.: The Horus WLAN Location Determination System. In: Proceedings of the 3rd International Conference on Mobile Systems, Applications, and Services (Mobisys), pp. 205–218 (2005)
26. XUL Tutorial (2006), http://www.xulplanet.com/tutorials/xultu/

Performance Analysis of Angle Routing in MANETs

Othman A. Al-Amoudi, Mohamed S. El-Azhari, Mike Woodward, and Irfan Awan

Mobile Computing, Networks and Security Research Group
School of Informatics, University of Bradford,
Bradford, BD7 1DP, U.K.
{O.A.Al-amoudi,m.el-azhari,i.u.awan,m.e.woodward}@bradford.ac.uk

Abstract. Links between nodes in mobile ad hoc networks (MANETs) are vulnerable to breakage because of the highly dynamic nature of MANETs, this results in frequent changes and unpredictability in network topologies. So, discovering and maintaining routes between nodes is one of the biggest challenges in MANETs. This makes the routing area perhaps the most active research area within the MANET domain. The ultimate goal of the MANET community is to provide a set of standardized protocols that can be both robust and scalable. This paper proposes a routing protocol based on the angles (directions) of the adjacent mobile nodes. Each pair of nodes that form a hop should ideally be moving in the same or similar direction, so the connection between the source and the destination will consist of a series of nodes that are moving in a similar direction. We measure the performance of the proposed approach by evaluating it against the ad-hoc on demand distance vector (AODV) routing protocol. The simulation results reveal that the proposed approach demonstrates better performance than AODV for most scenarios.

1 Introduction

Ad hoc networks are basically peer-to-peer multi-hop mobile wireless networks where information packets are transmitted in a store-and-forward manner from a source to an arbitrary destination, via intermediate nodes. Mobile networks can be classified into infrastructure networks and mobile ad hoc networks [1] according to their dependence on fixed infrastructures.

A mobile ad-hoc network (MANET) is a kind of wireless ad-hoc network, and is a self-configuring network of mobile routers (and associated hosts) connected by wireless links – the union of which form an arbitrary topology. The highly dynamic nature of mobile ad hoc networks results in frequent changes and unpredictability in network topologies. These make the routing area perhaps the most active research area within the MANET domain. The dynamics of a MANET result in numerous link breakages, discovering and maintaining routes between nodes that maintain long connection time is one of the biggest challenges in MANETs.

To compare and analyze MANET routing protocols, appropriate classification methods are important. Classification methods help researchers and designers to understand distinct characteristics of a routing protocol and find its relationship with

M. Takizawa, L. Barolli, and T. Enokido (Eds.): NBiS 2008, LNCS 5186, pp. 21–29, 2008.
© Springer-Verlag Berlin Heidelberg 2008

others. One of the most popular methods to distinguish MANET routing protocols is based on how routing information is acquired and maintained by mobile nodes. Using this method, mobile ad hoc network routing protocols can be divided into proactive routing, reactive routing and hybrid routing. The Dynamic Source Routing (DSR) [2] and Ad hoc on- demand Distance Vector routing (AODV) [3] are examples of reactive routing protocols for mobile ad hoc networks. The Zone Routing Protocol (ZRP), which is a hybrid routing protocol has been proposed to combine the merits of both proactive and reactive routing protocols and overcome their shortcomings [4].

The ultimate goal of the MANET community is to provide a set of standardized protocols that can be both robust and scalable.

In this paper, we undertake a study of a protocol that uses a new concept in MANET routing. The objective is to find a "good" path between a source and a destination. The most widely used metric in MANET routing is hop count [3]. It is used in both the static and dynamic networks. The term hop count relates to the number of legs traversed by a packet between the source and destination. If there are multiple routing paths available, the path with the minimum hop count will be selected.

We evaluate our proposed approach against the hop count approach by implementing a modified version of the AODV protocol. The simulation results show that broken links can be significantly reduced through the proposed approach.

The rest of this paper is structured as follows. Section 2 includes the background and related work of dissemination in MANETs. Section 3 presents the proposed angle direction approach. The parameters used in the experiments and the performance results and analysis to evaluate the effectiveness and limitations of the proposed technique are presented in Section 4. Section 5 concludes the paper and outlines the future work.

2 Related Work

This section analyses the related work which directly or indirectly aims at using the angles to form the route which consist of a number of legs (nodes) between the source and the destination. The proposed protocol is based on the heading angle and a mobile node that propagates a message in the network in order to find the best route to the destination, where each node is assumed to be equipped with a digital compass. Also, each node classifies its neighbor into eight different zone ranges (d1… d8) according to their direction [12].

The Random Waypoint mobility model has been used in many studies (e.g.,[7]) the results show that the Random Waypoint mobility model is a good approximation for simulating the motion of vehicles on a road, but there are situations in which a different model is better suited.

In [8] various MANET routing protocols, including DSR, AODV and DSDV have been evaluated and the results show that the protocol performance may vary drastically across mobility models and performance rankings of protocols may vary with the mobility models used. Random waypoint is a simple model that may be

applicable to some scenarios. However, we believe that it is not suitable to capture some important mobility characteristics of scenarios in which MANETs may be deployed.

3 Angle Direction Algorithms

There are varieties of ad hoc routing protocols. No matter how different they may be, in every routing protocol it is a key common task to find a "good" path between a source and a destination. Evaluation depends on a path metric such as hop count, expected delay, expected lifetime, etc. As a result, we have to find out a path that is optimal or at least nearly optimal with respect to the given or used path metric.

Our new suggested metric is the angles (directions) of the adjacent nodes. Each two nodes that forming a leg which is part of the distance between the source and the destination should be moving in the similar direction, so the connection between the source and the destination will consist of a series of nodes that have the approximate direction (Figure 2). The stability of the links has been considered in the route construction phase by the new metric.

Figures 1-2 explain the new suggested metric. This metric aims to have the connection between the nodes as long as it possible (Figure 2). In contrast, the hop count metric does not consider this (Figure 1).

The source node S, node 5, node 6, and node D represent a minimum hop route between the source and the destination D. For forming this route, the hop count metric is used for routing path construction. The node 5 and node 6 are coming closer but at the same time node 6 and node D are diverging. This implies leads to that the connection between node 6 and node D sooner will be disconnected (Figure 1).

As shown in Figure2 node S, node 2, node 3, node 4, and node D represent a route between the source and the destination. For forming this route, the new suggested metric is used for routing path construction. The directions of these nodes are the same or almost the same. This metric gives an expectation that the connection will remain as long as possible (Figure 2).

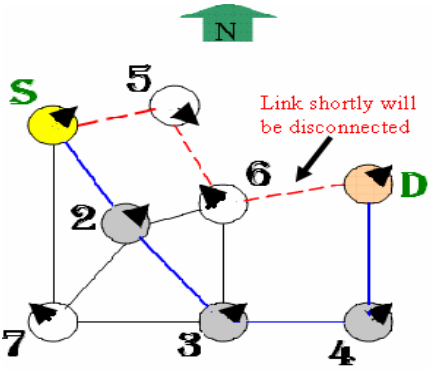


Fig. 1. Direction Angle and Hop Count

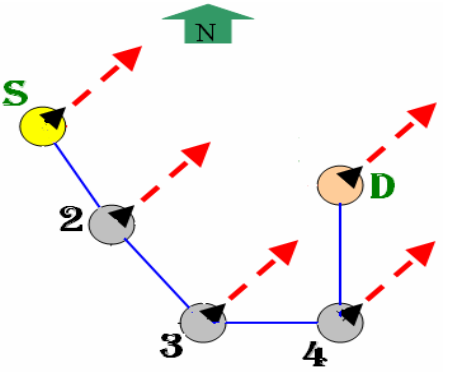
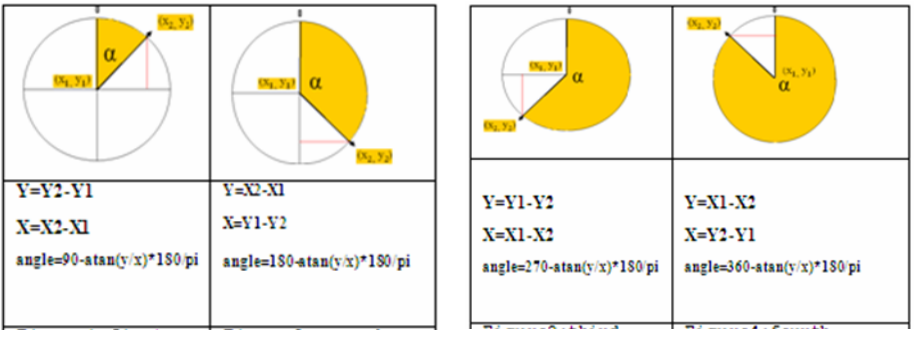


Fig. 2. Direction Angle



First and secod quadrants Third and fourth quadrants

Fig. 3.

The mobility procedure: On movement from position (x1, y1) to position (x2, y2); there are four quadrants as shown in Figure. 3.

Calculate the Node Angle:

Previous position x1, y1
Current position x2, y2
At the first quadrants
Angle=90-atan ((y2-y1)/(x2-x1)*180/3.14
At the second quadrants
Angle=180-atan ((x2-x1)/(y2-y1) *180/3.14
At the third quadrants
Angle=270-atan ((y1-y2)/(x1-x2)*180/3.14
At the fourth quadrants
Angle=360-atan ((x1-x2)/(y1-y2) *180/3.14

When a node receives a request packet:

Nod1: the sender node
Node2: the receiver node

Calculate the Angle difference between the two Nodes:

Def = |Node1Angle-Node2Angle|
If Def>180
Angle=360-Def
Else
Angle=def;
End if
Return Angle;

Procedure: Handle Request (angle process)

1: if packet received for the first time
2: if the route is new add it to Routing Table
3: else check if the angle of the last node is better (Angle in the table, Angle of the node, and Angle of the last node)

Procedure: Handle Reply
1: if there is a better route than those available update the Routing Table

4 Performance Analyses

In this section, we evaluate the performance of the proposed Angle direction algorithm. We compare the proposed algorithm with the basic AODV algorithm. We implement the proposed algorithm embedded in the AODV protocol. The metrics for comparison include the average number of collisions and number of broken links.

A. Simulation Setup

We have used the GloMoSim network simulator (version 2.03) [9] to conduct extensive experiments to evaluate the behaviour of the proposed algorithm. We study the performance comparison with the hop count approach, i.e AODV protocol [3, 10, 11] which is included in the GloMoSim package. The original AODV protocol uses hop count for discovering and maintaining routes between source and destination nodes. We have thus implemented AODV additionally using angle direction, called EAODV (AODV + enhanced by using angle). In our simulation, we use a 1000m × 1000m area with a random waypoint mobility model [3] of 100 mobile hosts. The network bandwidth is 2 Mbps and the medium access control (MAC) layer protocol is IEEE 802.11[8]. Other simulation parameters are shown in Table 1.

Table 1. Summary of the parameters used in the simulation experiments

Parameter	Value
Network range	1000m×1000m
Transmission range	250m
Number of mobile nodes	100
Number of connections	40
bandwidth	2Mbps
Traffic type	Constant bit rate (CBR)
Packet rate	packets per second
Packets during simulation time	50,100,150 packets
Packet size	512 bytes
Simulation time	900s
Speed	5,7.5,10 (m/s)

The main idea behind the proposed approach is to reduce the number of broken links in the route discovery phase, thus reducing the network traffic and decreasing the probability of channel contention and packet collision. Since our algorithm is based on an angle approach, it may not fit every scenario, however it significantly reduces the collisions.

B. Broken Links

Figure. 4 shows that our improved algorithm can significantly reduce the broken links for a network of 100 nodes and varying number of packets.

Fig.5 illustrates the improvement of reducing the number of broken links when the EAODV is used with 100 nodes moving at different speeds.

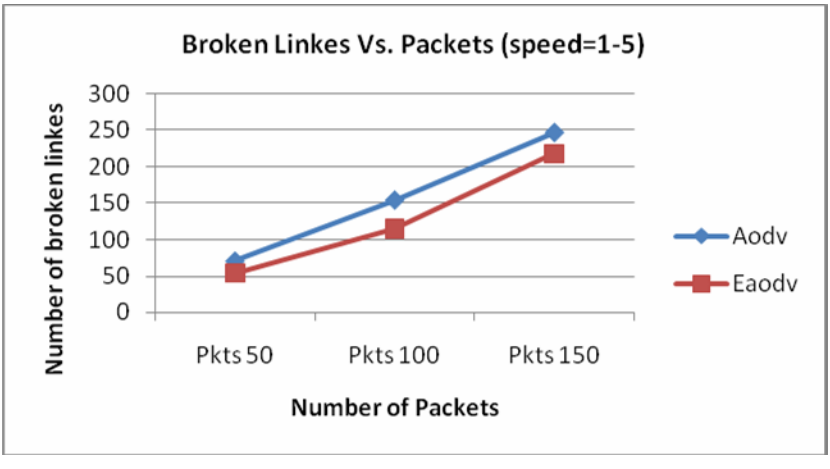


Fig. 4. Broken Links Vs. Packets

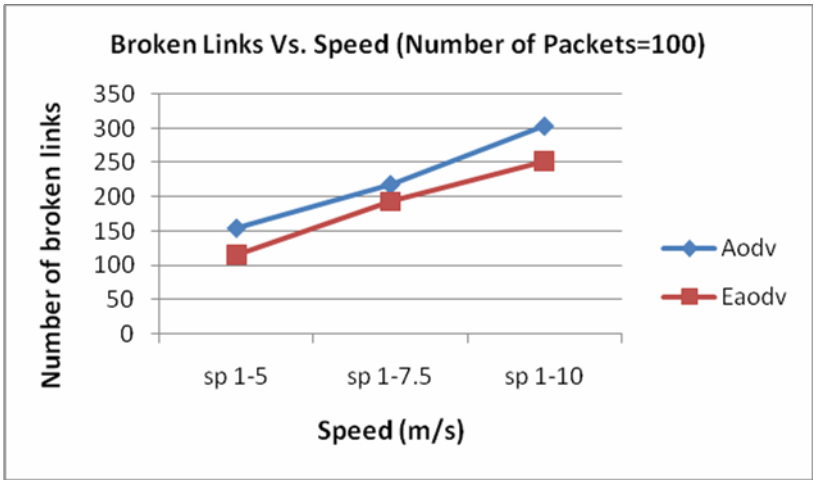


Fig. 5. Broken Links Vs. Speeds

C. Collisions

We also measure the number of collisions for these schemes at the physical layer. Since data packets and control packets share the same physical channel, the collision probability is high when there are a large number of control packets.

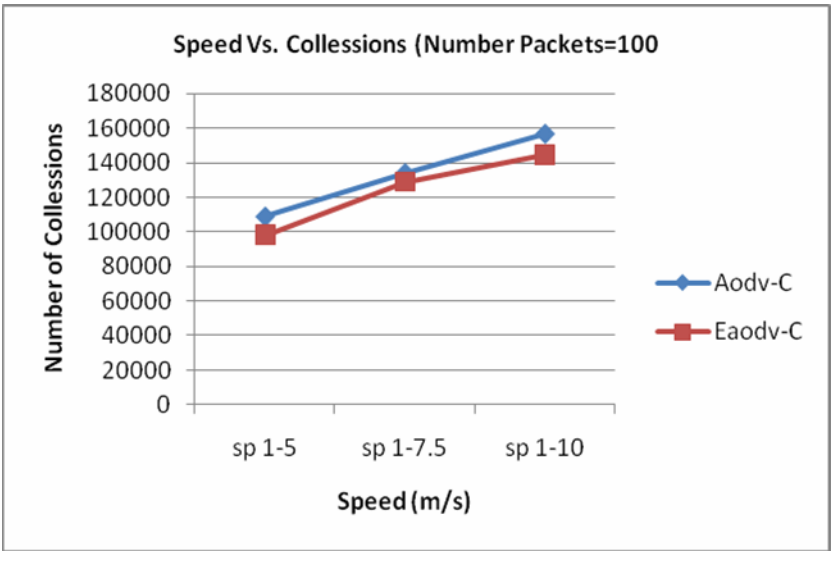


Fig. 6. Collisions Vs. Speeds

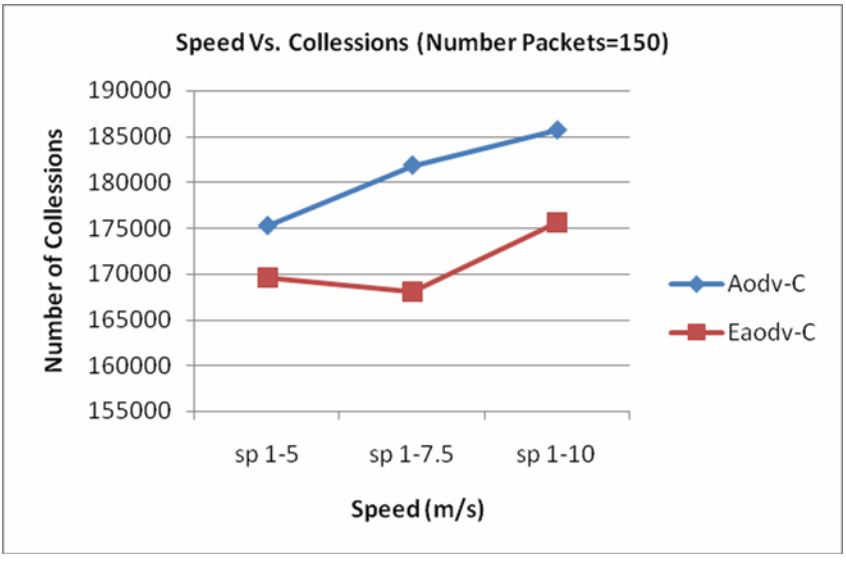


Fig. 7. Collisions Vs. Speeds

Fig. 6 shows the number of collisions for network with 100 nodes, 100 packets and different speeds. As shown in Fig. 6 our algorithm incurs fewer collisions than AODV.

Fig. 7 shows the number of collisions for a network with 100 nodes, 150 packets and different speeds. The improvement in reducing collisions is now more significant with a large number of packets.

5 Conclusions

This paper has presented an angle direction approach to routing for mobile ad-hoc networks. The proposed algorithm calculates the angle direction for every host node according to the node's coordination. Our simulation results prove this approach can generate fewer broken links than that of the AODV. It also demonstrates fewer collisions than the existing AODV approach.

For future work it would be interesting to explore the algorithm for different mobility in models. We also plan to evaluate the performance of angle direction routing compared with the Dynamic Source Routing (DSR) algorithm.

References

1. IETF Manet charter, http://www.ietf.org/html.charters/manet-charter.html
2. Johnson, D., Maltz, D.A.: Dynamic source routing in ad hoc wireless networks. In: Imielinski, T., Korth, H. (eds.) Mobile Computing. Kluwer Acad. Publ., Dordrecht (1996)
3. Perkins, C.E., Royer, E.M.: Ad-hoc on-demand distance vector routing. In: Proceedings of the 1999 Second IEEE Workshop on Mobile Computing Systems and Applications, February 1999, pp. 90–100. IEEE Computer Society, New York (1999)
4. Haas, Z.J.: The Zone Routing Protocol (ZRP) for ad hoc networks. Internet Draft (November 1997)
5. Saha, A.K., Johnson, D.B.: Modeling mobility for vehicular ad-hoc networks. In: Proceedings of the 1st ACM international workshop on Vehicular ad hoc networks, Philadelphia, PA, USA, October 1 (2004)
6. Bai, F., Sadagopan, N., Helmy, A.: The IMPORTANT Framework for Analyzing the Impact of Mobility on Performance of Routing for Ad Hoc Networks. AdHoc Networks Journal 1(4), 383–403 (November 2)
7. El-Nabi, T.H.A., Ahmed, A.: Modeling and simulation of a routing protocol for Ad hoc networks combining queuing network analysis and Ant colony algorithms. In: Institut für Informatik und Wirtschaftsinformatik Essen, April 2005. Universität Duisburg-Essen (Campus Essen) (2005)
8. Zhang, Q., Agrawal, D.: Dynamic probabilistic broadcasting in MANETs. Journal of ParallelDistributed Computing 65, 220–233 (2005)
9. Zeng, X., Bagrodia, R., Gerla, M.: GloMoSim: a library for parallel simulation of large-scale wireless networks. In: Proceedings of the 1998, 12th Workshop on Parallel and Distributed Simulations. PADS 1998, May 26–29, pp. 154–161. Banff, Alb., Canada (1998)
10. Sasson, Y., Cavin, D., Schiper, A.: Probabilistic broadcast for flooding in wireless mobile ad hoc networks. In: Proc. IEEE Wireless Communications & Networking Conference (WCNC 2003), March 2003, pp. 1124–1130 (2003)

11. Sasson, Y., Cavin, D., Schiper, A.: Probabilistic broadcast for flooding in wireless mobile ad hoc networks, EPFL Technical Report IC/2002/54, Swiss Federal Institute of Technology(EPFL) (2002)
12. Al-Akaidi, M., ALchaitia, M.: A New Heading Directional Angle Routing Protocol For Mobile Ad hoc Networks. In: The 5th IEE International Conference on 3G Mobile Communication Technologies (3G 2004) – The Premier Technical Conference on 3G and Beyond, October 18-20, 2004, Savoy Place, London (2004)

Dynamic Network Reconfiguration by Combination of Different Wireless LANs

Yoshitaka Shibata[1], Yosuke Sato[1], Naoki Ogasawara[1], Go Chiba[1], and Kazuo Takahata[2]

[1] Iwate Prefectural University, 152-52 Sugo, Takizawa, Iwate, Japan 020-0193
{shibata,yosuke,ogasawara,chiba}@sb.soft.iwate-pu.ac.jp
[2] Saitama Institute Technology, 1690 Fuzaiji, Fukaya, Saitama, Japan 369-0293
takahata@sit.ac.jp

Abstract. In this paper, we introduce a mobile network for disaster communication network by combination of different wireless LANs and mobile network. Currently available wireless LANs such as IEEE802.11b,g,j,n, IEEE802.16 (WiMAX), cellular network are combined with a mobile router and loaded on a car to build a mobile network node. Using multiple mobile network nodes, a large disaster communication network is organized.

Keywords: disaster network, wireless LAN, mobile network.

1 Introduction

As recent advent of wireless communication network technology, various types of wireless networks have been emerged and used in various application fields, such as hot spot at public areas, Intelligent Trafiic System, disaster communication network, etc. In particular, wireless and mobile networks performed very important role for disaster use.

So far we have developed disaster information networks which are very effective and robust using commercial available wireless network such as IEEE802.b.g.j and could verify their usefulness[1][2][3]. Those networks are consisted of multiple wireless LANs with the same standard as network configuration. However, since connectivity among the whole network depends on distance between the wireless network nodes, power density, transmission frequency and their cover area, network transmission speed, communication may or may not be disconnected depending on the communication environment. Also mobile environment was not considered in the previous networks.

In this paper, by combining different standard Wireless LANs and dynamically changing the communication link and route according to the application, media to be transmitted, communication distance, network connectivity could be preserved.

In the followings, system configuration and architecture is explained in section 2. Path selection method between the nodes is explained in section 3. Prototyped system and it performance evaluation is described in section 4. Finally we conclusions and future works e are summarized.

M. Takizawa, L. Barolli, and T. Enokido (Eds.): NBiS 2008, LNCS 5186, pp. 30–37, 2008.
© Springer-Verlag Berlin Heidelberg 2008

2 System Configuration

Our proposed system is shown in Fig. 1 and configured by mobile nodes loaded on a vehicle and fixed nodes. The mobile node is consisted of different standard wireless LANs and their antennas, a Router PC, a note PC and a video camera. Each node is linked one another and organizes an adhoc network and finally connected to the fixed node which is a gateway to Internet or other wide area network. Each link from one node to another has multiple channels by different wireless LANs. One of the suitable link among them are selected depending on the network environment such as the distance between node, power density of LANs, and organizes an adhoc network. The router PC finds the route to the fixed node and dynamically determines the other route when the communication to the neighbor node is disconnected. Thus, by automatically selecting the best the link and dynamically change the path to the fixed node depending on the network environment, various application services such as file transfer, Web, VoIP and video communication services can be supported.

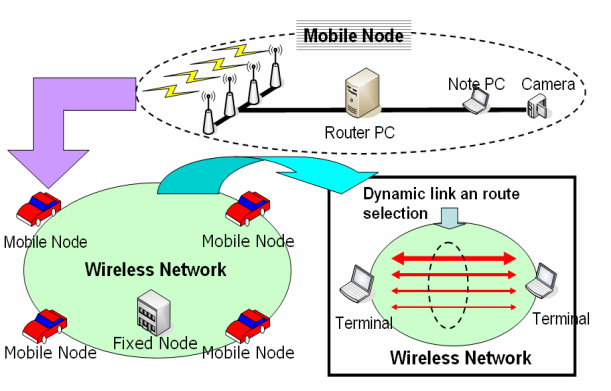


Fig. 1. System Configuration

2.1 System Architecture

The system architecture of our system in mobile node is shown in Fig. 2 and is consisted of four modules including sender/receiver module, connection management module, route decision module and monitor module on top of wireless LANs. The monitor module monitors the change of throughput and registers the current link of LANs, path and traffic load on the node to the Route list. The Route decision module determines the route of LAN based on Route list and Policy list. The connection management module observes various parameters including throughput, delay, power density, packet loss rate by periodically sending sensing packet to the neighbor nodes. The sender/receiver module sends and receives packet stream information and control information. Policy list manages poly list for router PC set by user through browser GUI. Three different data including control data, connection acknowledgement data and media data are transmitted between the neighbor mobile nodes. The control data is transmitted between the connection management modules or route lists and used to

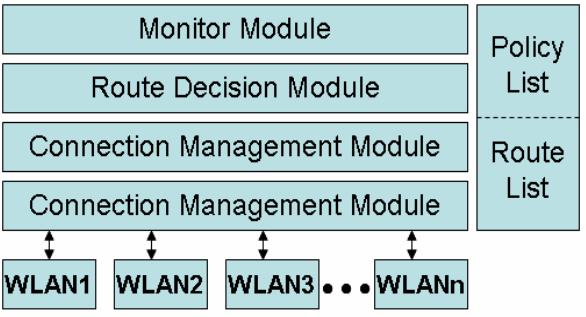


Fig. 2. System Architecture

determine the suitable route and link. The connection acknowledgement module is used to know and confirm the connection status to the neighbor mobile nodes by sending sensing packet. The media data is used to transmit the various users' media data including Web., VoIP, video streaming and etc..

3 Path Selection Method

There are two types of path selection method including link section method between neighbor nodes and route selection method for route between the source node and destination node.

3.1 Link Selection

Fig. 3 indicates the case of link selection method. The proper link is changed among the several possible links depending on the connection status. When current connection status is worse by decreasing the power density, throughput, packet succeeding rate, then another link of LANs are selected and changed.

The link selection is realized on Router PC on mobile node. The router PC periodically observes throughput, delay time, power density, throughput, packet loss rate and understands whether those links are possibly available or not on each link. The policy to use the link is also set through the GUI in advance. The link selection is

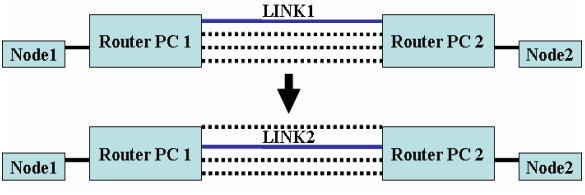


Fig. 3. Link Selection Method

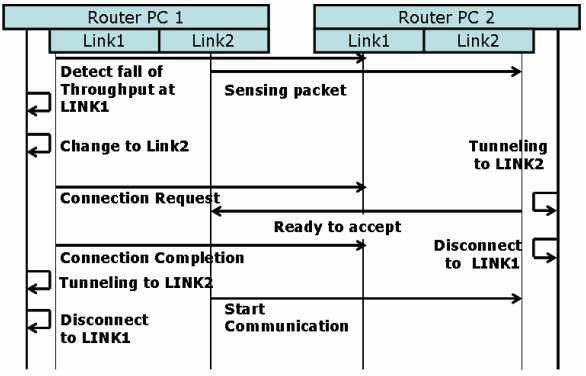


Fig. 4. Data Flow of Link Selection

carried out based on the following procedure as shown in Fig. 4. The link selection method is carried out in the following algorithm.

1. Each node grasps the end-to-end route status by periodically flooding sensing packets
2. Router PC1 detects that the state parameter, such as throughput at the Link 1 between Router PC1 and PC2 decreased.
3. Router PC1 sends link Reconnect Request message to Router PC2 based on Policy list.
4. Router PC2 makes a tunnel for Link 2 and sends a Ready to Accept message to Route PC1.
5. Router PC1 makes a tunnel for Link 2 and sends a Ready to Reconnect message to Router PC2.
6. Communication between Router PC1 and PC2 starts and both tunnels are released.
7. The current link state is recorded.

3.2 Route Selection

Fig. 5 shows an example Router PC selection method. The status information of currently connecting Router PCs including the Router PC number, IP address and link state is always recognized on each Router PC. When some Router PC, Router PC 2 in Fig. 5 have to be exchanged from PC3 to other due to degradation of throughput, the PC Router selection method is carried out in the followings steps similar to the link selection algorithm:

1. Each node grasps the end-to-end route status by periodically flooding sensing packets
2. Router PC2 finds that the throughput between Router PC2 and PC3 decreased
3. Router PC2 determines Router PC4 as a next route
4. Router PC2 send a connection request to Route PC4 to change a link to another channel

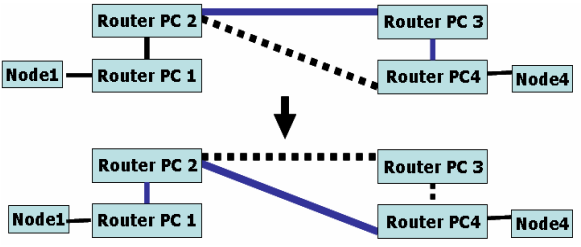


Fig. 5. Router PC selection Method

5. Router PC4 creates a tunnel for path and send reconnect ready to Router PC2
6. Router PC2 creates a tunnel to router PC4 and eliminates the path to route PC3
7. The current path is registered to route list in both

3.3 Policy List

When the current link state between the nodes gets worse, a new link connection is established. There are several parameters to be observed during communication including standard, power density, frequency of LANs, throughput, delay, packet loss, distance between the neighbor nodes. Distance between the nodes can be calculated using GPS at each node. In our system, we introduce priority method how the current connection is changed to a new link and predefined as Policy list as follows:

1. User selected default values.
2. Throughput between nodes
3. Power density of LAN
4. Packet Loss
5. Standard LAN

By predefining the threshold of changing link, the communication can continue without disconnection between nodes.

3.4 Route List

In the Route list, two routing tables are created. One table includes default route information with normal connection nodes and their IP addresses and their priority numbers. The other table includes alternative routing information with possible connection nodes and their IP addresses.

4 Prototype and Evaluation

In order to verify the effects of the suggested system, we constructed a prototype system on the vehicle as shown in Fig. 6. As communication network, four different wireless LANs including IEEE802.11a (4.9 GHz, 10mW, 54Mbps), IEEE802.11b,g (2.4GHz, 10mW, 11Mbps & 54Mbps), D-Star(1.2GHz, 1W, 128Kbps) and CDMA 1X

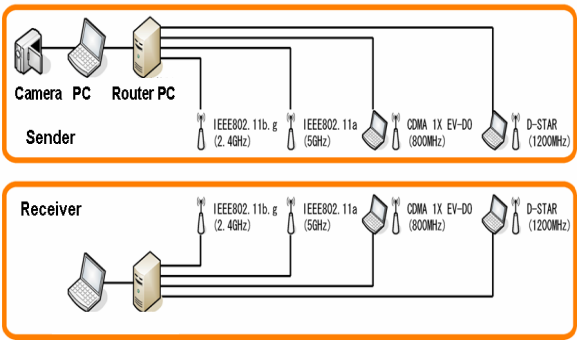


Fig. 6. Prototype System

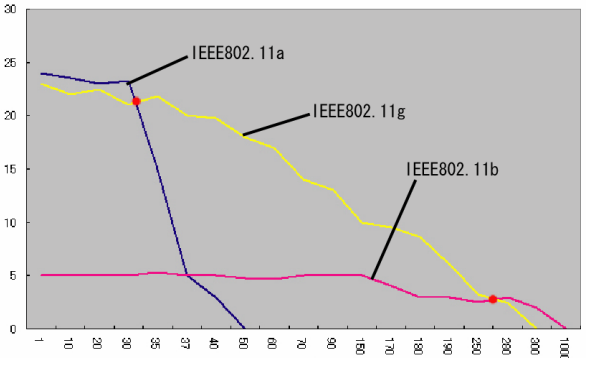


Fig. 7. Performance Result: Throughput(1)

EV-DO(0.8GHz, 400mW, 1.2Mbps) are installed. We developed and implemented the function modules using C++ on mobile router PC (Mobile Pentium 3) running on 4.3 BSD Unix as shown in Fig. 6.

First we observed throughput and packet loss of each LAN link by changing the distance between two nodes. The result is shown in Fig. 7. ~ Fig. 9. Based on this figure, the thresh hold for Policy list to change the communication link is determined. Next we implemented video stream application module between the nodes. As video streaming, Midfield was applied by controlling its video quality from 56Kbps to 20Mbps using its transcoding function.

In this prototype, when the distance between the mobile nodes are short (1~ 250m), the video stream with 5 Mbps was transmitted using default LAN by IEEE802.11g. As the distance increases, the throughput decreased under 5 Mbps when the distance was more than 250m. Then Router PC changed the wireless LAN from IEEE802.11g to IEEE802.11b with 1.5Mbps. When the distance is about 1,000m, the Rout PC changed to CDMA2000 1X EV-DO with about 200 Kbps. Thus, depending on the distance, the throughput reduces dynamically and the LAN changed to transmit the video steam by changing video quality.

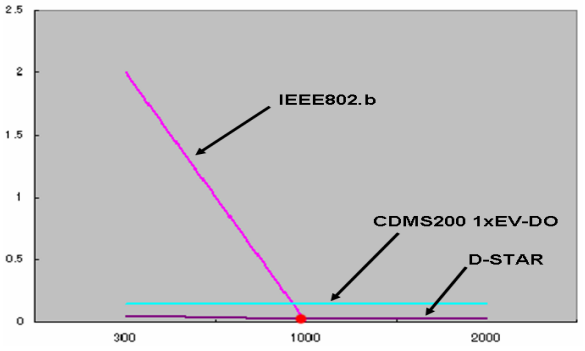


Fig. 8. Performance Result: Throughput(2)

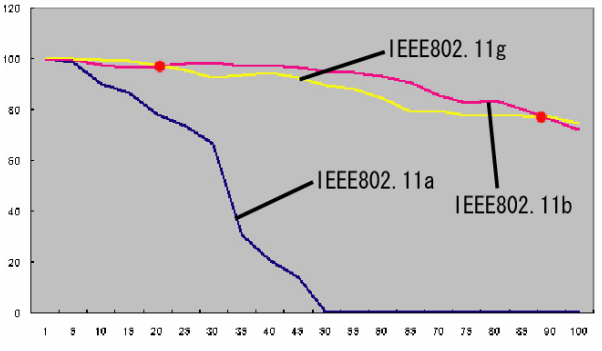


Fig. 9. Packet Succeeding Rate

5 Conclusions

In this paper, we introduced a mobile network for disaster communication network by combination of different wireless LAN and mobile network using currently available different wireless LANs to organize a large disaster communication network when public communication network was damaged. A communication path between nodes has multiple links and the suitable link among them is selected based on policy role such as the distance, power and transmission channel frequency. By multi-hopping those nodes, user can communicate with other user and send/receive disaster information even though some of information infrastructure are damaged. In this paper, we designed and implemented a wireless mobile network with currently available different wireless networks and constructed a prototyped system to evaluate the functional and performance. Currently we are implementing dynamic routing method among adhoc network to provide end-to-end path. In the future work, the parameter values and threshold for routing method and policy rule should be optimized.

References

[1] Shibata, Y., Nakamura, D., Uchida, N., Takahata, K.: Residents Oriented Disaster Information Network. In: IEEE Proc on SAINT 2003, January 2003, pp. 317–322 (2003)

[2] Sakamoto, D., Hashimoto, K., Takahata, K., Shibata, Y., et al.: Performance Evaluation of Evacuation information Network System based on Wireless Wide Area Network. DPS, 100–112 (in Japanese) (2000)

[3] Nakamura, D., Uchida, N., Asahi, H., Takahata, K., Hashimoto, K., Shibata, Y.: Wide Area Disaster Information Network and Its Resource Management System. In: AINA 2003 (March 2003)

[4] Hashimoto, K., Shibata, Y.: Design of A Middleware System for Flexible Intercommunication Environment. In: IEEE Proc. on Advanced Information Networking and Applications, pp. 59–64 (March 2003)

[5] Kyoung Paik, E., Cho, H.-s., et al.: Load Sharing and Session Preservation with Multiple Mobile Routers for Large Scale Network Mobility. In: AINA 2004, pp. 393–398 (2004)

[6] Isomura, M., Imai, N., et al.: Performance Evaluation of Inverse Multiplexing of Heterogeneous Communication Media for Mobile Router. In: FIT 2004, pp. 239–240 (2004)

[7] Oda, S., Ogashiwa, N., et al.: A New Multi-Homing Architecture Based on Overlay Network. IEICE Transaction J87-B(10), 1564–1573 (2004)

[8] Sharma, S., Baek, I., et al.: OmniCon: A Mobile IP-based Vertical Hando® System for Wireless LAN and GPRS Links. In: ICPPW 2004, pp. 330–337 (2004)

[9] Stem, M., Katz, R.: Vertical Hando®s in Wireless Overlay Networks. In: MONET, pp. 335–350 (1997); Special Issue on Mobile Networking in the Internet

Making an Agreement in an Order-Heterogeneous Group

Ailixier Aikebaier[1], Tomoya Enokido[2], and Makoto Takizawa[3]

[1] Tokyo Denki University, Japan
alisher.akber@computer.org
[2] Rissho University, Japan
eno@ris.ac.jp
[3] Seikei University, Japan
makoto.takizawa@st.seikei.ac.jp

Abstract. In traditional agreement protocols of multiple peer processes (peers), every peer just aims at agreeing on one value out of values shown by the peers. In meetings of human societies, agreement procedures are so flexible that persons can change their opinions and can use various types of agreement conditions. We already discuss E- and P-precedent relations $v_1 \rightarrow_i^E v_2$ and $v_1 \rightarrow_i^P v_2$ on values v_1 and v_2 of a peer p_i, which show that p_i can take v_2 after taking v_1 and prefers v_1 to v_2, respectively. If a peer autonomously takes values only based on its precedent relations, the peers might not make an agreement even if the values satisfy the agreement condition. We discuss what previous values the peer can take again an order-heterogenous system where some pair of peers have different precedent relations. In this paper, we discuss a *cut*, i.e. a satisfiable set of previous values in a history of values which the peers have so far taken, in addition for each peer to taking a new value at each round.

1 Introduction

Most peer-to-peer (P2P) systems [1] are managed by coordinators like indexes and superpeers [16, 17]. In this paper, we consider a fully distributed P2P system where there is no centralized coordinator and each peer directly communicates with other peers. In P2P applications like social network service (SNS) [12, 15], multiple peer processes (peers) $p_1, \ldots, p_n$ are cooperating to make a decision. Traditional agreement protocols [5, 8, 10, 11] just aim at agreeing on one value out of values shown by the peers. In human societies, agreement procedures are so flexible that members can change their opinions and can use various types of agreement conditions like majority-condition. The authors [3, 4, 13] discuss the *existentially (E-) precedent relation* $v_1 \rightarrow_i^E v_2$ and the *preferentially (P-) precedent relation* $v_1 \rightarrow_i^P v_2$ to show that a peer p_i can take a value v_2 after taking a value v_1 and prefers v_1 to v_2, respectively.

Based on the precedent relations, we discuss protocols for multiple peers to make an agreement which satisfies types of agreement conditions [3, 4]. Since each

M. Takizawa, L. Barolli, and T. Enokido (Eds.): NBiS 2008, LNCS 5186, pp. 38–48, 2008.
© Springer-Verlag Berlin Heidelberg 2008

peer autonomously selects values which can be taken in the E- and P- precedent relations, the peers may fail to make an agreement depending on the order in which a peer takes values even if there exist a tuple of values which each peer can take and satisfy the agreement condition. Each peer p_i keeps in record a local history of values which p_i has so far sent and received. A checkpoint mechanism to rollback to the previous rounds is discussed in a paper [2]. In order to make an agreement, we discuss another approach to finding a satisfiable tuple of previous values in addition for each peer to finding a new value. We define a *cut* which is a tuple of previous values, each of which is taken by a peer at a previous round. If a *cut* which satisfies the agreement condition is found, every peer can make an agreement by backing to the cut. Even if there is a satisfiable cut in the history, some peer may not be able to back to the cut because the peer cannot withdraw some values which the peer has taken. For example, after a person says "no", he cannot withdraw it in some meeting. We define an *uncompensatable* value which each peer cannot withdraw after showing to other peers. We define a *recoverable* cut in a history to which every peer can back and which satisfies the agreement condition on the basis of the uncompensatable values. Then, we discuss how every peer to back to the cut in an order-heterogeneous system where some pair of peers have different precedent relations.

In section 2, we discuss types of precedent relations on values. In section 3, we define a history of values. In section 4, we discuss to which previous round each peer can back in the history. In section 5, we discuss a coordination protocol.

2 Precedent Relations

2.1 Coordination Procedure

In fully distributed P2P applications, multiple peers $p_1, \ldots, p_n$ have to make an agreement, for example, e.g. to fix a date for a meeting of members in a society. A *domain* D_i of a peer p_i is set of possible values which p_i can take. We assume every peer is reliable and every pair of peers can reliably communicate with one another in the underlying network.

Each peer p_i initially takes a value v_i^0 in the domain D_i. p_i sends v_i^0 and receives values $v_1^0, \ldots, v_n^0$ from the peers $p_1, \ldots, p_n$, respectively. The agreement condition AC_i like *all*, *majority*, *weighted majority*, *consonance*, and others [3, 4] is checked for the tuple $\langle v_1^0, \ldots, v_n^0 \rangle$. If AC_i is not satisfied, p_i takes another value $v_i^1 = LD_i(v_1^0, \ldots, v_n^0)$ in D_i, where LD_i is a local decision function. This is the first round. p_i sends v_i^1 to the other peers and receives values from the other peers. Thus, at each round t, each peer p_i collects a tuple $\langle v_1^{t-1}, \ldots, v_n^{t-1} \rangle$ of values received from the peers. If AC_i is satisfied, p_i obtains one agreement value $v = GD_i(v_1^{t-1}, \ldots, v_n^{t-1})$ by performing a global decision function GD_i. Unless AC_i is satisfied, p_i takes a value $v_i^t = LD_i(v_1^{t-1}, \ldots, v_n^{t-1})$. Then, p_i notifies the other peers of v_i^t. Here, the values $v_i^0, v_i^1, \ldots, v_i^{t-1}$ are *previous* ones at round t and v_i^t is a *current* value of p_i. Every peer p_i is assumed to have the same ones $AC_i = AC$, $LD_i = LD$, and $GD_i = GD$.

2.2 Precedent Relations

We define the *existentially (E-) precedent* relation $\rightarrow_i^E$ and *preferentially (P-) precedent* relation $\rightarrow_i^P$ on a domain D_i to show with which value a peer p_i can take after a value at each round.

Definition. For every pair of values v_1 and v_2 in a domain D_i of a peer p_i,

1. v_1 *E-precedes* v_2 in p_i ($v_1 \rightarrow_i^E v_2$) iff p_i is allowed to take v_1 after v_2.
2. v_1 *P-precedes* v_2 in p_i ($v_1 \rightarrow_i^P v_2$) iff p_i prefers v_1 to v_2.
3. $v_1 \rightarrow_i^E v_2$ and $v_1 \rightarrow_i^P v_2$ if $v_1 \rightarrow_i^E v_3 \rightarrow_i^E v_2$ and $v_1 \rightarrow_i^P v_3 \rightarrow_i^P v_2$ for some value v_3, respectively.

In a distributed auction system [6], each person cannot show a cheaper value v_2 than a previous value v_1, i.e. $v_1 \rightarrow_i^E v_2$ where $v_2 > v_1$. Suppose a peer p_i can take a pair of values v_1 and v_2 after taking a value v in the E-dominant relation $\rightarrow_i^E$, i.e. $v \rightarrow_i^E v_1$ and $v \rightarrow_i^E v_2$. Suppose neither $v_1 \rightarrow_i^E v_2$ nor $v_2 \rightarrow_i^E v_1$. If p_i prefers v_1 to v_2 ($v_2 \rightarrow_i^P v_1$), p_i would like to take v_1. The precedent relations $\rightarrow_i^E$ and $\rightarrow_i^P$ with the domain D_i are *a priori* specified when each peer p_i is initiated.

There are the following relations between a pair of values v_1 and v_2 in D_i:

1. v_1 is *E-equivalent* with v_2 in p_i ($v_1 \equiv_i^E v_2$) iff $v_1 \rightarrow_i^E v_2$ and $v_2 \rightarrow_i^E v_1$.
2. v_2 is more *E-significant* than v_1 in p_i ($v_1 \prec_i^E v_2$) iff $v_1 \rightarrow_i^E v_2$ but $v_2 \not\rightarrow_i^E v_1$.
3. v_1 *E-dominates* v_2 in p_i ($v_1 \preceq_i^E v_2$) iff $v_1 \prec_i^E v_2$ or $v_1 \equiv_i^E v_2$.
4. v_1 is *E-incomparable* with v_2 in p_i ($v_1 \mid_i^E v_2$) iff neither $v_1 \rightarrow_i^E v_2$ nor $v_2 \rightarrow_i^E v_1$.
5. v_1 is *P-equivalent* with v_2 in p_i ($v_1 \equiv_i^P v_2$) iff $v_1 \rightarrow_i^P v_2$ and $v_2 \rightarrow_i^P v_1$.
6. v_1 is more *P-significant* than v_2 in p_i ($v_2 \prec_i^P v_1$) iff $v_1 \rightarrow_i^P v_2$ but $v_2 \not\rightarrow_i^P v_1$.
7. v_1 *P-dominates* v_2 in p_i ($v_2 \preceq_i^P v_1$) iff $v_2 \prec_i^P v_1$ and $v_2 \equiv_i^P v_1$.
8. v_1 and v_2 are *P-incomparable* in p_i ($v_1 \mid_i^P v_2$) iff neither $v_1 \rightarrow_i^P v_2$ nor $v_2 \rightarrow_i^P v_1$.

A value v_1 is *maximal* and *minimal* iff there is no value v_2 such that $v_1 \rightarrow_i^E v_2$ and $v_2 \rightarrow_i^E v_1$ in D_i, respectively. Let $Corn_i(v)$ be a set of values which p_i can take after taking a value v in D_i, $Corn_i(v) = \{ y \mid v \rightarrow_i^E y \} \subseteq D_i$. If v is maximal, $Corn_i(v) = \phi$. $|Corn_i(v)| \geq 2$, v is *branchable*.

A *least upper bound* (*lub*) of values v_1 and v_2 ($v_1 \sqcup_i^E v_2$) is a value v_3 in the domain D_i such that $v_1 \rightarrow_i^E v_3$, $v_2 \rightarrow_i^E v_3$, and there is no value v_4 such that $v_1 \rightarrow_i^E v_4 \rightarrow_i^E v_3$ and $v_2 \rightarrow_i^E v_4 \rightarrow_i^E v_3$ in a peer p_i. A *greatest lower bound* (*glb*) of values v_1 and v_2 ($v_1 \sqcap_i^E v_2$) is similarly defined.

3 A History of a Peer

3.1 History

Each peer p_i takes a value while exchanging values with the other peers at each round. A *history* H_i^t of a peer p_i is a collection of local histories $\langle H_{i1}^t, \ldots, H_{in}^t \rangle$

at round t. A local history H_{ii}^t is a sequence $\langle v_i^0, v_i^1, \ldots, v_i^{t-1} \rangle$ of values which p_i has taken until round t from the initial round 0. H_{ij}^t is a sequence of values $\langle v_j^0, v_j^1, \ldots, v_j^{t-1} \rangle$ which p_i has received from p_j until round t ($j = 1, \ldots n, i \neq j$). Here, $H_{ij}^t = H_{jj}^t$. Initially, $H_{ij}^0 = \phi$. A notation $H_{ij}^t|_u$ shows a value v_j^u which p_i receives from p_j at round u ($u \leq t$). $H_{ii}^t|_u$ shows a value v_i^u which p_i takes at the round u.

Let H be a sequence $\langle x_1, \ldots, x_m \rangle$ ($m \geq 1$) of values. Here, a value x_l *precedes* x_h ($x_l \Rightarrow x_h$) if $l < h$ in H. "$H + x$" shows a sequence $\langle x_1, \ldots, x_m, x \rangle$ of values obtained by adding a value x to H. $\langle x_1, \ldots, x_l \rangle$ ($l \leq m$) and $\langle x_k, \ldots, x_m \rangle$ ($1 < k$) are a *prefix* and *postfix* of H, respectively. For example, $\langle a, b, c, d \rangle$ and $\langle c, d, e \rangle$ are a prefix and postfix of $H = \langle a, b, c, d, e \rangle$, respectively.

A value x may occur multiple times in a sequence H. Let $H_{ij}^t[x]$ show a subsequence $\langle x, \ldots, x \rangle$ of instances of a value x in H_{ij}^t. $|H_{ij}^t[x]|$ is the number of instances of a value x in a local history H_{ij}^t.

A peer p_i takes a current value v_i^t after receiving a tuple $\langle v_1^{t-1}, \ldots, v_n^{t-1} \rangle$ from the other peers p_1, ..., p_n, at round t. Then, p_i sends v_i^t and receives v_j^t from p_j. At $t + 1$, the local history H_{ij}^{t+1} of p_i is $H_{ij}^t + v_j^t = \langle v_j^0, v_j^1, \ldots, v_j^{t-1}, v_j^t \rangle$ ($j = 1, \ldots, n$).

Let D_i^* show a set of possible sequences of values obtained from the domain D_i. Here, every local history H_{ii}^t is an element of D_i^*. We introduce the following coordination methods on the history H_i^t:

1. *forward*: $D_i^* \rightarrow D_i^*$. For a sequence H_1 in D_i^*, H_1 is a prefix of $forward(H_1)$.
2. *compensate*: $D_i^* \rightarrow D_i^*$. For a sequence H_1 in D_i^*, $compensate(H_1)$ is a prefix of H_1.
3. *null*: $D_i^* \rightarrow D_i^*$. For a sequence H_1 in D_i^*, $null(H_1) = H_1$.

Each peer p_i takes one of the coordination methods at each round as shown in Figure 1. In *forward*, p_i selects a new value v_i^t as presented in the basic

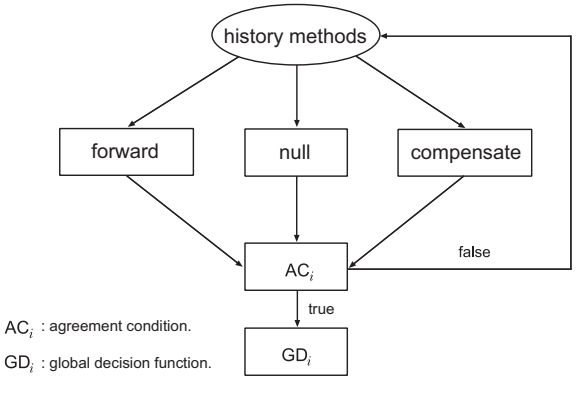


Fig. 1. Coordination procedure of a peer p_i

coordination procedure. In *null*, p_i does not select a new value. In *compensate*, p_i backs to the previous round.

3.2 Compensation

In some meeting of multiple persons, there may be some rule that each person cannot withdraw a special remark "no". A value x is *primarily uncompensatable* in p_i iff p_i cannot withdraw x after showing x to the other peers. Otherwise, x is *primarily compensatable* in p_i. In a local history $H^4_{ii} = \langle a, b, c, d, e\rangle$, suppose a value c is primarily uncompensatable and the other values are primarily compensatable. Here, d and e can be compensated but c cannot. Although a and b are primarily compensatable, neither a nor b can be withdrawn because c is primarily uncompensatable.

Definition. In a local history $H^t_{ii} = \langle v^0_i, \ldots, v^{u-1}_i, \ldots, v^{t-1}_i\rangle$ of a peer p_i, a value v^{u-1}_i is *uncompensatable* iff v^{u-1}_i is primarily uncompensatable or some value v_i preceded by v^{u-1}_i ($v^{u-1}_i \Rightarrow v_i$) is uncompensatable. u^{u-1}_i is *compensatable* iff v^{u-1}_i is not uncompensatable.

A sequence $\langle v^0_i, v^1_i, \ldots, v^{t-1}_i\rangle$ is *uncompensatable* iff v^{t-1}_i is primarily uncompensatable or $\langle v^0_i, \ldots, v^{t-2}_i\rangle$ is uncompensatable. A peer p_i cannot back to the previous round u at t ($u < t$) if v^s_i ($u < s < t$) is uncompensatable in H^t_{ii}. In the example, the local history $H^4_{ii} = \langle a, b, c, d, e\rangle$ is uncompensatable, because c is primarily uncompensatable while $\langle d, e\rangle$ is compensatable. In $H^t_{ii} = \langle v^0_i, v^1_i, \ldots, v^{t-1}_i\rangle$, an uncompensatable value v^u_i is a *most recently uncompensatable* (MRU) value iff $\langle v^{u+1}_i, \ldots, v^{t-1}_i\rangle$ is compensatable. A postfix $\langle v^{u+1}_i, \ldots, v^{t-1}_i\rangle$ a *maximally compensatable* in $H^t_{ii} = \langle v^0_i, v^1_i, \ldots, v^u_i, \ldots, v^{t-1}_i\rangle$ iff the postfix is compensatable and a prefix $\langle v^0_i, v^1_i, \ldots, v^u_i\rangle$ is uncompensatable. In $H^4_{ii} = \langle a, b, c, d, e\rangle$, c is the most recently uncompensatable value. $\langle d,\ e\rangle$ is maximally compensatable in H^4_{ii}.

At round t, a peer p_i takes a value v^t_i from the tuple $\langle v^{t-1}_1, \ldots, v^{t-1}_n\rangle$. Here, $v^{t-1}_i \rightarrow^E_i v^t_i$. Suppose $v^{t-1}_j \rightarrow^E_i v^t_i$, i.e. $v^t_i = v^{t-1}_i \sqcap^E_i v^{t-1}_j$ and p_j withdraws v^{t-1}_j. If p_j takes another value v ($\neq v^{t-1}_j$), p_i may take a different value from v^t_i. Hence, if p_j compensates v^{t-1}_j, p_i has to compensate v^t_i since p_i takes v^t_i based on v^{t-1}_j. For each value v^t_i, a *minimal domain* $MD_i(v^t_i)$ is a subset of values in the tuple $\langle v^{t-1}_1, \ldots, v^{t-1}_n\rangle$ such that $v^t_i = \sqcap^E_{i\ x \in MD(v^t_i)} x$ and $v^t_i \neq \sqcap^E_{i\ x \in MD(v^t_i)-y} x$ for every y in $MD_i(v^t_i)$. If any value in $MD_i(v^t_i)$ is omitted, the *lub* of $MD_i(v^t_i)$ is not v^t_i. v^t_i *depends on* v^{t-1}_j in p_i ($v^{t-1}_j \vdash_i v^t_i$) iff $v^{t-1}_j \in MD_i(v^t_i)$.

It is straightforward for the following theorem to hold from the definitions.

Theorem. A value v^t_i is required to be compensated in a peer p_i if at least one value in the minimal domain $MD_i(v^t_i)$ is compensated.

Theorem. If a value v^t_i is uncompensatable in a peer p_i, every value in the minimal domain $MD_i(v^t_i)$ is uncompensated.

3.3 Constraints on Values

Each value v in a domain D_i is characterized in terms of the maximum occurrence $MO_i(v)$, i.e. how many times a peer p_i can take. For example, each person can say "no" at most once in some meeting. If $MO_i(v) = 1$, p_i can take v at most once. If $|H_{ii}^t[v]| < MO_i(v)$, p_i can take v again at round t. At t, p_i can take a value v_i^t which satisfies the following conditions: 1) for every value x in the local history H_{ii}^t, $x \rightarrow_i^E v_i^t$ and 2) $|H_{ii}^t[v_i^t]| < MO_i(v_i^t)$. Let $P_i^t(v_i^{t-1})$ be a set of possible values which p_i can take at t; $P_i^t(v_i^{t-1}) = \{\ v \mid |H_{ii}^t[v]| < MO_i(v)$ and for every value x in H_{ii}^t, $v_i^{t-1} \rightarrow_i^E v\ \}$. For each value v in $P_i^t(v_i^{t-1})$, v is removed from $P_i^t(v_i^{t-1})$ if $|H_{ii}^t\ [v]| = MO_i(v)$. Then, p_i takes a value v_i^t in $P_i^t(v_i^{t-1})$ if $P_i^t(v_i^{t-1}) \neq \phi$ and sends v_i^t to the other processes. If $P_i^t(v_i^{t-1}) \neq \phi$, there is no value which p_i can take at t after taking v_i^{t-1} at t - 1. If $|P_i^t(v_i^{t-1})| \geq 2$, v_i^{t-1} is *branchable* at t. Even if v_i^{t-1} is branchable, $|Corn_i(v_i^{t-1})| \geq 2$, v_i^{t-1} may be taken in previous rounds of H_{ii}^t.

4 Backwarding Strategies

4.1 Cuts

Let δ_i be the current round of a peer p_i with a local history $H_{ii}^{\delta_i}$=$\langle v_i^0, v_i^1, \ldots, v_i^{\delta_i - 1}\rangle$. A history $H_i^{\delta_i}$ is a tuple $\langle H_{i1}^{\delta_1}, \ldots, H_{ii}^{\delta_i}, \ldots, H_{in}^{\delta_n}\rangle$ of local histories.

Definition. A *cut* of a history $H_i^{\delta_i}$ is a tuple of values $\langle v_1^{t_1}, \ldots, v_n^{t_n}\rangle$ where $t_j < \delta_j$ for each $j = 1, \ldots, n$.

Definition. A cut $ct = \langle v_1^{t_1}, \ldots, v_n^{t_n}\rangle$ of a history $H_i^{\delta_i}$ is *satisfiable* in a peer p_i iff ct satisfies the agreement condition AC_i.

A cut $\langle v_1^{t_1}, \ldots, v_n^{t_n}\rangle$ is *current* iff $t_j = \delta_j$ for every peer p_j. In the coordination protocol, a peer p_i takes a *forward* function, i.e. takes a new value. However, even if the current cut $\langle v_1^{\delta_1}, \ldots, v_n^{\delta_n}\rangle$ is not satisfiable, there might be a satisfiable cut $\langle v_1^{t_1}, \ldots, v_n^{t_n}\rangle$ in a history $H_i^{\delta_i}$. Here, if every peer p_i backs to the round $t_i + 1$ by compensating the local history $H_{ii}^{\delta_i}$ ($i = 1, \ldots, n$), every peer p_i can make an agreement on a value $v = GD_i(v_1^{t_1}, \ldots, v_n^{t_n})$.

Definition. Let $H_{ii}^{\delta_i}$ be a local history $\langle v_i^0, v_i^1, \ldots, v_i^{\delta_i - 1}\rangle$ of each peer p_i ($i = 1, \ldots, n$). A cut ct=$\langle v_1^{t_1}, \ldots, v_n^{t_n}\rangle$ is *obtainable* in p_i iff a postfix $\langle v_i^{t_i+1}, \ldots, v_i^{\delta_i - 1}\rangle$ of ct is compensatable in p_i. A cut $ct = \langle v_1^{t_1}, \ldots, v_n^{t_n}\rangle$ is *obtainable* iff ct is obtainable in every peer.

A cut $ct = \langle v_1^{t_1}, \ldots, v_n^{t_n}\rangle$ is obtainable in a history $H_i^{\delta_i}$ if the following conditions hold:

1. Let mru_j be the most recently uncompensatable value in a local history $H_{ij}^{\delta_j}$. A value $v_j^{t_j}$ in ct precedes mru_j in $H_{ij}^{\delta_j}$ ($mru_j \rightarrow v_j^{t_j}$).
2. Let v_k mean a value from a peer p_k in the minimal domain $MD_j(v_j^{t_j})$ ($v_k \vdash_j v_j^{t_j}$). From each value $v_j^{t_j}$ in ct, every value v_k in $MD_j(v_j^{t_j})$ precedes a value $v_j^{t_j}$ in $H_{ij}^{\delta_j}$ ($v_k \rightarrow v_j^{t_j}$) as shown in Figure 2.

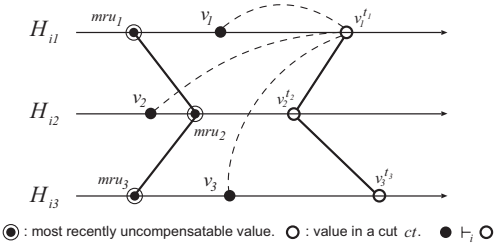


Fig. 2. Obtainable cut

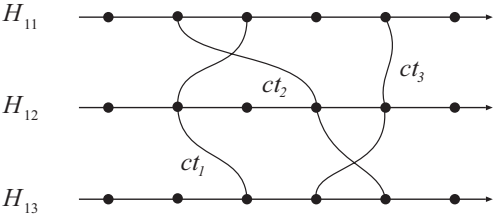


Fig. 3. Cuts

Even if a cut $ct = \langle v_1^{t_1}, \ldots, v_n^{t_n} \rangle$ satisfies the agreement condition AC_i, $v_i^{t_i}$ may not be obtainable in some peer p_i.

Definition. A cut $ct = \langle v_1^{t_1}, \ldots, v_n^{t_n} \rangle$ is *recoverable* iff ct is satisfiable and obtainable in every peer.

Theorem. If there is a recoverable cut $ct = \langle v_1^{t_1}, \ldots, v_n^{t_n} \rangle$ in a history $H_i^{\delta_i}$, every peer p_i can make an agreement on ct by backing to the round $t_i + 1$.

In a history $H_i^{\delta_i}$, there might be multiple recoverable cuts $ct_1, \ldots, ct_m$ ($m > 1$). Every peer p_i has to take the same cut ct_l out of the possible cuts $ct_1, \ldots, ct_n$. Let ct_1 and ct_2 be a pair of recoverable cuts $\langle v_{11}^{t_{11}}, \ldots, v_{1n}^{t_{1n}} \rangle$ and $\langle c_{21}^{t_{21}}, \ldots, c_{2n}^{t_{2n}} \rangle$ of a history $H_i^{\delta_i}$, respectively. First, ct_1 *precedes* ct_2 in $H_i^{\delta_i}$ ($ct_1 \rightarrow ct_2$) if $t_{1j} \leq t_{2j}$ for every j ($= 1, \ldots, n$). Otherwise, ct_1 and ct_2 *intersect*. Figure 3 shows three cuts ct_1, ct_2 and ct_3 in $H_i^{\delta_i}$. Here, ct_1 precedes ct_2 ($ct_1 \rightarrow ct_2$). ct_1 and ct_2 intersect while ct_2 and ct_3 intersect. Suppose there are a pair of recoverable cuts ct_1 and ct_2 in $H_i^{\delta_i}$. Here, each peer p_i has to make a decision on which ct_1 or ct_2 to be taken. In this paper, each peer p_i takes ct_1 if ct_2 precedes ct_1 in $H_i^{\delta_i}$ ($ct_2 \rightarrow ct_1$). A cut ct is *maximal* in $H_i^{\delta_i}$ iff there is no cut ct' in $H_i^{\delta_i}$ where ($ct \rightarrow ct'$). Next, suppose ct_1 and ct_2 intersect in $H_i^{\delta_i}$. Here, we introduce the weight $|ct|$ for a cut $ct = \langle v_1^{t_1}, \ldots, v_n^{t_n} \rangle$ in a history $\langle H_{i1}^{\delta_1}, \ldots, H_{in}^{\delta_n} \rangle$ as $|ct| = \sum_{j=1,\ldots,n} (\delta_j - t_j)$. ct_1 is *smaller* than ct_2 ($ct_1 < ct_2$) if $|ct_1| < |ct_2|$. ct_1 is taken if ct_1 and ct_2 intersect and $|ct_1| < |ct_2|$.

1. Let CT be a set of recoverable cuts in $H_i^{\delta_i}$. Let MCT be a set of maximal cuts in the set CT with respect to the precedent relation $\rightarrow$.
2. A smallest cut ct is selected in the set MCT.

4.2 Reselectable Values

Suppose a peer p_i takes a maximal value v_i^t in the domain D_i at round t. At $t+1$, p_i cannot take another value since v_i^t is maximal in D_i. Here, p_i has to back to the previous round u by compensation and take another value.

Let a history $H_i^{\delta_i}$ be $\langle v_i^0, v_i^1, \ldots, v_i^u, \ldots, v_i^{\delta_i - 1}\rangle$. A peer p_i takes a value v_i^{u+1} after taking v_i^u. If there is only one value v_i^{u+1} where $v_i^u \rightarrow_i^E v_i^{u+1}$, p_i cannot take another value different from v_i^{u+1} at round u. Hence, it is meaningless to compensate a subsequence $\langle v_i^{u+1}, \ldots, v_i^{\delta_i - 1}\rangle$ in $H_{ii}^{\delta_i}$, i.e. to back to the previous round $u+1$. On the other hand, suppose there are multiple values $v_1, \ldots, v_m$ $(m \geq 2)$ where $v_i^u \rightarrow_i^E v_l$ $(l = 1, \ldots, m)$. Suppose p_i takes a value v_l as v_i^{u+1} in the values $v_1, \ldots, v_m$. If the postfix $\langle v_i^{u+1}, \ldots, v_i^{\delta_i - 1}\rangle$ of $H_{ii}^{\delta_i}$ is compensated, p_i takes another value v_k $(\neq v_l)$ where $v_i^u \rightarrow_i^E v_k$ by backing to the round $u+1$.

Each time p_i backs to the round $u+1$, p_i has to take a value in $Corn_i(v_i^u)$ which has not been so far taken. For each branchable value v_i^u, $Used_i(v_i^u)$ indicates a set of values in $Corn_i(v_i^u)$ which p_i has taken until $u+1$. If v_i^u is taken at u, $Used_i(v_i^u) = \phi$. Then, a value v in $Corn_i(v_i^u)$ is taken. Here, $Used_i(v_i^u) = \{v\}$. Suppose p_i backs to $u+1$. Here, p_i takes a value v in $Corn_i(v_i^u)$ but not in $Used_i(v_i^u)$, i.e. $v \in P_i^{u+1}(v_i^u) = Corn_i(v_i^u)$ - $Used_i(v_i^u)$. Then, v is added to $Used_i(v_i^u)$. v_i^u is *branchable* in $H_i^{\delta_i}$ iff $P_i^{u+1}(v_i^u) \neq \phi$.

Definition. Let $H_{ii}^{\delta_i}$ be a local history $\langle v_i^0, \ldots, v_i^u, \ldots, v_i^{\delta_i - 1}\rangle$ of values which a peer p_i has taken until round δ_i. v_i^u is *reselectable* in H_{ii}^{δ} iff v_i^u is branchable in H_{ii}^{δ} and a subsequence $\langle v_i^{u+1}, \ldots, v_i^{\delta_i - 1}\rangle$ is compensatable.

By compensation, a peer p_i can back to a reselectable value v_i^u taken at u+1. Then, p_i takes a new value in $P_i^{u+1}(v_i^u) = Corn_i(v_i^u)$ - $Used_i(v_i^u)$.

5 A Coordination Protocol

5.1 Compensation to Cut

First, we discuss how each peer p_i compensates the history $H_i^{\delta_i}$ to a recoverable cut. Suppose a peer p_i obtains values $v_1^{\delta_1}, \ldots, v_n^{\delta_n}$ from the peers $p_1, \ldots, p_n$ at round δ_i. p_i searches $H_i^{\delta_i}$ for recoverable cuts. As discussed, we cannot compensate a subsequence $\langle x_1, \ldots, x_n\rangle$ of values if x_n is uncompensatable. Let MRU_j denote the most recently uncompensatable value v in a local history $H_{ij}^{\delta_i}$. Initially, $MRU_j = \phi$ for every p_j. At each round, p_i takes a new value v. $MRU_i = v$ if v is uncompensatable. There is no need that every value in $H_i^{\delta_i}$ is searched. In $H_{ij}^{\delta_i}$, only a subsequence $H_{ij}^{\delta} \mid_{MRU_j}^{\delta_i} = \langle v_j^{MRU_j}, v_j^{MRU_j+1}, \ldots, v_j^{\delta_j - 1}\rangle$ from a value denoted by MRU_j to the current value $v_j^{\delta_j}$ is searched to obtain a set of recoverable cuts. In the *all* agreement condition AC_i, a set SC_i of satisfiable cuts are obtained as follow:

$SC_i = \phi$;
for $t_i = MRU_i, \ldots, \delta_i - 1$ { $v = v_i^{t_i}$;
  for $t_1 = MRU_1, \ldots, \delta_1 - 1$;
    if $v = v_1^{t_1}$ for $t_2 = MRU_2, \ldots, \delta_2 - 1$;
      ...
        if $v = v_{n-1}^{t_{n-1}}$ for $t_n = MRU_n, \ldots, \delta_n - 1$;
          if $v = v_n^{t_n}$, $SC_i = SC_i \cup \{\langle v_{i1}^{t_1}, \ldots, v_{in}^{t_n}\rangle\}$;
}

First, p_i takes a maximal cut $ct_i = \langle v_1, \ldots, v_n\rangle$ in SC_i with respect to the precedent relation $\rightarrow$. p_i sends a cut request $ct_i = \langle v_1, \ldots, v_n\rangle$ to every peer. On receipt of the cut request $ct_i = \langle v_1, \ldots, v_n\rangle$ from p_i, a peer p_j checks if v_j is compersatable in $H_j^{\delta_j}$. If MRU_j precedes the value v_j ($MRU_j \rightarrow v_j$) in $H_{jj}^{\delta_i}$, v_j is compersatable. Otherwise, v_j is uncompersatable. If v_j is uncompersatable, p_j sends NAK to p_i. Otherwise, p_j sends ACK to p_i. p_i waits for responses from all the other peers. If every peer p_j sends ACK, p_i sends *Agree* to every peer. On receipt of *Agree*, each peer p_j obtains a value $v = GD_j(v_1, \ldots, v_n)$ and then terminates. If p_i receives NAK, p_i sends NAK to every peer which sends ACK to p_i. Then, p_i finds another satisfiable cut in SC_i.

Next, multiple peers may find different cuts. Suppose a pair of peers p_i and p_j find satisfiable cuts ct_i and ct_j, respectively. p_i and p_j send cut requests ct_i and ct_j to every other peer, respectively, as presented here. Now, suppose p_i receives ACK from every peer and sends ACK to p_j. Here, since p_j may receive ACK from every other peer, p_i sends a *Confirmation* ($Conf$) message ct_i to p_j. If p_j receives NAK from some peer, p_j sends NAK of ct_j to p_i. Here, p_i sends *Agree* of ct_i to every peer and every peer p_j obtains a value v from ct_i. If p_j receives ACK of ct_j from every peer, p_j also sends ($Conf$) of ct_j to p_i. Here, each of p_i and p_j takes either ct_i or ct_j. If $|ct_i| < |ct_j|$, p_i and p_j take the smaller cut ct_i. p_i sends *Agree* of ct_i to every peer. Then, every peer p_j makes an agreement on a value v for the cut ct_i.

6 Concluding Remarks

In human societies, each person may change its opinion in the agreement procedure. By abstracting human behaviours, values in a domain are partially ordered in existentially (E-) and preferentially (P-) precedent relations $\rightarrow_i^E$ and $\rightarrow_i^P$. In this paper, we consider an order-heterogeneous system where every peer has the same domain but some pair of peers has different precedent relations. If each peer just autonomously takes a value by using the precedent relations at each round, the peers may not make an agreement even if there is a value which every peer can take. In order to efficiently make an agreement, we need some coordination mechanisms of multiple peers. In this paper, we discuss a coordination protocol for multiple peers to make an agreement. If values taken by the peers do not satisfy the agreement condition, each peer takes a new value in the *forward* procedure. On the other hand, after some rounds, some collection

of values which peers have taken may satisfy the agreement condition. We define a satisfiable cut which is a tuple of previous values, each of which a peer has taken. Next, we define a recoverable cut. If there is a recoverable cut where each peer can back to the previous round, every peer can make an agreement by backing to a previous round which is satisfiable and where every peer can back. We discuss the protocol for how every peer to back to a previous round shown in a recoverable cut.

References

1. Peer-to-peer, http://en.wikipedia.org/wiki/Peer-to-peer
2. Aikebaier, A., Enokido, T., Takizawa, M.: Checkpointing in a Distributed Coordination Protocol for Multiple Peer Processes. In: Second International Conference on Complex, Intelligent and Software Intensive Systems (CISIS 2008), pp. 48–54 (2008)
3. Ailebaier, A., Hayashibara, N., Enokido, T., Takizawa, M.: A Distributed Coordination Protocol for a Heterogeneous Group of Peer Processes. In: IEEE 21th Conference on Advanced Information Networking and Applications (AINA 2007), pp. 565–572 (2007)
4. Ailebaier, A., Hayashibara, N., Enokido, T., Takizawa, M.: Making an Agreement in an Order-Heterogeneous Group by using a Distributed Coordination Protocol. In: 2nd International Workshop on Advanced Distributed and Parallel Network Applications (ADPNA 2007), CD-ROM (2007)
5. Corman, A.B., Schachte, P., Teague, V.: A Secure Group Agreement (SGA) Protocol for Peer-to-Peer Applications. In: 21st International Conference on Advanced Information Networking and Applications Workshops (AINAW 2007), pp. 24–29 (2007)
6. Ezhilchelvan, P., Morgan, G.: A Dependable Distributed Auction System: Architecture and an Implementation Framework. In: IEEE 5th International Symposium on Autonomous Decentralized Systems (ISADS), p. 3 (2001)
7. Gray, J., Lamport, L.: Consensus on Transaction Commit. ACM Transactions on Database Systems (TODS) archive 31(1), 133–160 (2006)
8. Hurfin, M., Raynal, M., Tronel, F., Macedo, R.: A General Framework to Solve Agreement Problems. In: 18th IEEE Symposium on Reliable Distributed Systems (SRDS), pp. 56–65 (1999)
9. Kling, R.: Cooperation, Coordination and Control in Computer-supported Work. Communications of the ACM 34(12), 83–88 (1991)
10. Lamport, L., Shostak, R., Pease, M.: The Byzantine Generals Problem. ACM Transactions on Programming Languages and Systems 4(3), 382–401 (1982)
11. Lee, P., Lui, J., Yau, D.: Distributed Collaborative Key Agreement Protocols for Dynamic Peer Groups. In: 10th IEEE International Conference on Network Protocols, pp. 322–331 (2002)
12. Sabater, J., Sierra, C.: Reputation and Social Network Analysis in Multi-agent Systems. In: first international joint conference on Autonomous agents and multiagent systems: part 1, pp. 475–482 (2002)
13. Shimojo, I., Tachikawa, T., Takizawa, M.: M-ary Commitment Protocol with Partially Ordered Domain. In: Tjoa, A.M. (ed.) DEXA 1997. LNCS, vol. 1308, pp. 397–408. Springer, Heidelberg (1997)

14. Skeen, D.: NonBlocking Commit Protocols. In: Proceedings of the ACMSIGMOD International Conference on Management of Data, pp. 133–142 (1981)
15. Upadrashta, Y., Vassileva, J., Grassmann, W.: Social Networks in Peer-to-Peer Systems. In: 38th Hawaii International Conference on System Sciences(HICSS-38 2005), CD-ROM (2005)
16. Freenet, http://freenetproject.org/
17. Gnutella, http://en.wikipedia.org/wiki/Gnutella

Performance Evaluation of Two Search Space Reduction Methods for a Distributed Network Architecture

Leonard Barolli[1], Makoto Ikeda[2], Arjan Durresi[3], Fatos Xhafa[4], and Akio Koyama[5]

[1] Department of Information and Communication Engineering
Fukuoka Institute of Technology (FIT)
3-30-1 Wajiro-Higashi, Higashi-Ku, Fukuoka 811-0295, Japan
barolli@fit.ac.jp
[2] Graduate School of Engineering
Fukuoka Institute of Technology (FIT)
3-30-1 Wajiro-Higashi, Higashi-Ku, Fukuoka 811-0295, Japan
bd07001@ws.ipc.fit.ac.jp
[3] Department of Computer and Information Science
Indiana University Purdue University Indianapolis
723 W. Michigan Street SL 280, Indianapolis, IN 46202, USA
durresi@cs.iupui.edu
[4] Department of Languages and Informatics Systems
Polytechnic University of Catalonia
Jordi Girona 1-3, 08034 Barcelona, Spain
fatos@lsi.upc.edu
[5] Department of Informatics, Yamagata University
4-3-16 Jonan, Yonezawa 992-8510, Yamagata, Japan
akoyama@yz.yamagata-u.ac.jp

Abstract. In our previous work, we proposed a distributed network architecture which integrates routing and CAC strategies using cooperative agents. However, in the previous research, we did not evaluated the performance of all agents. In this paper, we evaluate the performance of two search space reduction methods which are used for reduction of the search space of a GA-based routing approach. Thus, the GA can find a feasible route very fast.

1 Introduction

The networks of today are going through a rapid evolution and they are expected to support a wide range of multimedia applications. The requirement for timely delivery of multimedia data raises new challenges for the next generation broadband networks. The key issue is the Quality of Service (QoS) routing. Also, ensuring the QoS demands to traffic flows and groups of flows is an important challenge for future broadband networks, and resource provisioning via Call Admission Control (CAC) is a key mechanism for achieving this.

The purpose of admission control is to support the QoS demands of real time applications via resource reservation. While, the purpose of QoS routing is to find good paths which satisfy user requirements.

Traditional CAC schemes can be classified in equivalent capacity, heavy traffic approximation, upper bounds of the cell loss probability, fast buffer/bandwidth allocation,

M. Takizawa, L. Barolli, and T. Enokido (Eds.): NBiS 2008, LNCS 5186, pp. 49–59, 2008.
© Springer-Verlag Berlin Heidelberg 2008

and time windows. Among proposed CAC schemes, the equivalent capacity gives better results [1].

So far, many routing algorithms have been proposed. The routing strategies can be classified into three classes: source, distributed and hierarchical routing. Source routing algorithms are conceptually simple, but they suffer from scalability problem. Distributed routing algorithms are more scalable, but loops may occur, which make the routing to fail. Hierarchical routing has been used to cope with the scalability problems of source routing in large internetworks. The hierarchical routing retains many advantages of source routing. It also has some advantages of distributed routing because the routing computation is shared by many nodes. But in the conventional hierarchical routing, the network state is aggregate additional and gives some imprecision, which has a significant negative impact on QoS routing [2,3,4].

In order to support multimedia communication over high speed networks, it is necessary to develop routing algorithms which use for routing more than one QoS parameters such as throughput, delay, and loss probability. This is because new services such as video on demand and remote meeting systems require better QoS. However, the problem of QoS routing is difficult, because the distributed applications have very diverse QoS constraints on delay, loss ratio and bandwidth.

To cope with broadband networks the CAC schemes and routing algorithms should be adaptive, flexible, and intelligent for efficient network management. Use of intelligent algorithms based on Fuzzy Logic (FL), Genetic Algorithms (GA) and Neural Networks (NN) can prove to be efficient for telecommunication networks [5,6,7,8,9].

In our previous work, we proposed an intelligent CAC and routing framework using cooperative agents [5]. The proposed framework is based on Distributed Artificial Intelligence (DAI) approach, which deals with design of artificial agents to develop intelligent systems. We introduced two types of agents: simple and intelligent agents. The intelligent agents were based on FL and GA. However, in the previous work we did not evaluate all agents of the proposed architecture. In this work, we evaluate the performance of two agents of the proposed architecture: the Search Space Reduction Agent (SSRA) and Tree Model Network Agent (TMNA). We integrated two agents and implemented two methods: Method1 and Method2.

The paper is organized as follows. In Section 2, we introduce the previous work. The proposed algorithms are presented in Section 3. The simulation environment and results are discussed in Section 4. Finally, conclusions are given in Section 5.

2 Previous Work

2.1 Distributed Network Architecture

The proposed network architecture is a Multi-Agent System (MAS). The agents are distributed and cooperate together. Each Domain Management Agent (DMA) has four agents: the Resource Management Agent (RMA), Precomputation Agent (PA), which includes SSRA and TMNA, Destination Discovery Agent (DDA), and Routing Agent (RA) with its Intra Domain (IntraD) and Inter Domain (InterD) agents. The DMA structure is shown in Fig. 1. The PA includes SSRA and TMNA. We call these two agents PA, because they make the computation before the RA is activated. The computation

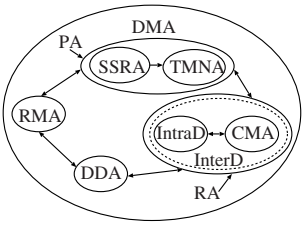


Fig. 1. DMA structure

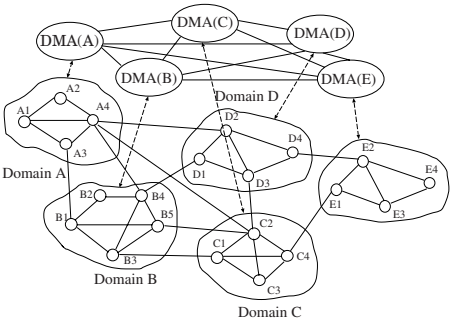


Fig. 2. Distributed network architecture with DMAs

time starts when a new connection makes a request to the network. The RA has the IntraD and InterD agents. In fact, the InterD agent is a composition of IntraD agent and Connectivity Management Agent (CMA), which are activated by an escalation strategy. The distributed network architecture with DMAs is shown in Fig. 2. This architecture can be considered as a hierarchical architecture, where in first level are domains and in the second level are DMAs. We have shown here only five domains. But, this architecture can be scaled-up easily by increasing the number of DMAs and domains in order to deal with the increasing users demands and number of switches.

2.2 RMA

As a RMA we proposed and implemented a Fuzzy Admission Control (FAC) scheme [7]. In difference from the equivalent capacity admission control method [1], which uses only the available capacity as the only variable for the call admission decision, our FAC scheme considers four parameters: Quality of service (*Qs*), Network congestion parameter (*Nc*), Available capacity (*Ac*), and user requirement parameter which is expressed by Equivalent capacity (*Ec*). The output linguistic parameter is the Acceptance decision (*Ad*). In order to have a soft admission decision, not only "accept" and "reject" but also "weak accept", "weak reject", and "not accept not reject" are used to describe the accept/reject decision.

The FAC scheme is shown in Fig. 3. The information for FAC are given by Bandwidth Management Predictor (BMP); Congestion Information Indicator (CII); Quality

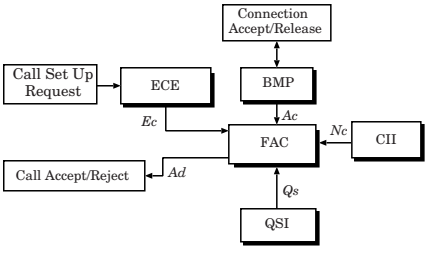


Fig. 3. FAC scheme

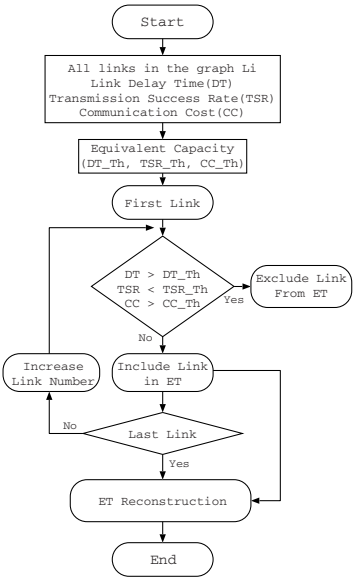


Fig. 4. SSRA flowchart

of Service Indicator (QSI); and Equivalent Capacity Estimator (ECE). The BMP works in this way: if a connection is accepted, the connection bandwidth is subtracted from the available capacity of the network, otherwise, if a connection is released, the connection bandwidth is added to the available capacity of the network. The CII decides whether the network is or isn't congested. The QSI determines whether allowing a new connection violates or not the QoS guarantee of the existing connections.

2.3 PA and DDA

SSRA. By using the SSRA a network with many nodes and links will be reduced in a network with a small number of nodes and links. Thus, the proposed framework is able to cope with more large-scale networks. The flowchart of SSRA is shown in Fig. 4.

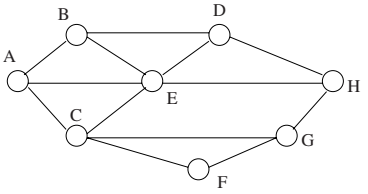


Fig. 5. A network example with 8 nodes

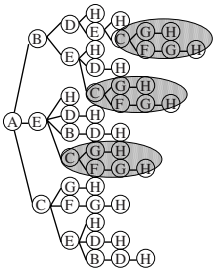


Fig. 6. Network tree model

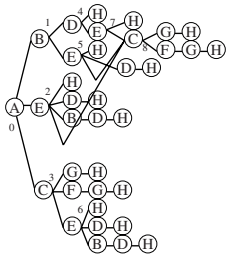


Fig. 7. Reduced network tree model

TMNA. After the execution of SSRA, the effective topology of the network is transformed in a tree model by TMNA. To explain this procedure, we consider a small network with 8 nodes as shown in Fig. 5. Node A is the Source Node (SN) and node H is the Destination Node (DN). All paths are expressed by the tree model shown in Fig. 6. In the shaded areas are shown the same paths from node C to H. Therefore, we further reduce the tree network as shown in Fig. 7. The tree model constructed by TMNA is used by IntraD agent for intra-domain routing. In the reduced tree model, each tree junction is considered as a gene and the path is represented by the chromosome. In this case, two tree junctions were omitted, because they represent the same routes from node C to H.

DDA. After a new connection is accepted, the RMA sends a request to the DDA. The DDA consults a table with node name entries to check whether SN and DN are

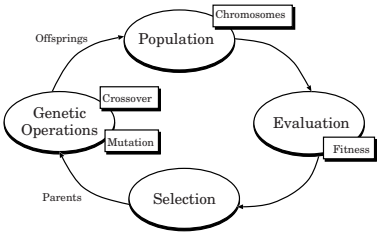


Fig. 8. Genetic cycle

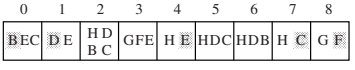


Fig. 9. Gene coding

in the same domain or not. If SN and DN are in the same domain, the DDA of the source domain activates the IntraD agent. Otherwise, if the SN and DN are in different domains, the InterD agent is activated.

2.4 IntraD Agent

The IntraD algorithm is a delay-constraint unicast source routing mechanism and is based on GA [10]. The GA cycle is shown in Fig. 8. The most important factor to achieve efficient genetic operations is gene coding. In order to simplify the genetic operations of the IntraD algorithm, the network is expressed by a tree network and the genes represent the tree junctions. A chromosome example is shown in Fig. 9. The genes in a chromosome have two states "active" and "inactive". A gene is called "active" if the junction is in the path, otherwise the gene is in "inactive" state. The genetic operations are carried out in the "active" genes. Each gene includes information of the adjacent nodes. The paths are represented by chromosomes which have the same length. Therefore, the crossover operation becomes easy.

2.5 InterD Agent

After the DDA finds out that SN and DN are in different domains, the InterD agent is activated. The InterD agent is a composition of IntraD agent and CMA. It use an escalation strategy to make the inter-domain routing. By using the escalation strategy, the information exchange is needed only in domains where the selected path passes. Thus, the information flooding in all domains is not necessary and the network resources can be used efficiently. The InterD agent operates in the following way. After receiving a connection request, a node become a SN. The IntraD agent finds a path inside the domain. The DN of the source domain starts the CMA. The CMA is a simple agent. It finds the best link by using a sorting algorithm based on the inter-domain links parameters. After the CMA decides the best link for connection, the DN of this link becomes a SN and the IntraD agent is activated in the following domain. This procedure is escalated until the DN of the destination domain is found.

3 Proposed Methods

In this section, we will present the proposed methods. We integrated SSRA and TMNA and implemented two methods: Method1 and Method2. The flowchart of the proposed methods is shown in Fig. 10.

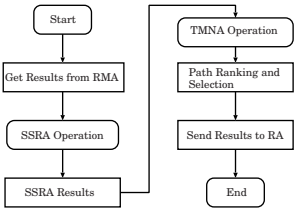


Fig. 10. Flowchart of proposed methods

3.1 Method1

In Method1, after deciding the number of nodes the algorithm starts to connect different nodes in a random way. As link parameters are used: delay, packet loss and bandwidth. These parameters are assign in a random way for each link. Then based on some thresholds the topology of the network is reduced. The threshold is selected more strict when the number of nodes is large. In Fig. 11 is shown an example before using Method1. The reduced network topology is shown in Fig. 12. This method has the following drawbacks:

- the processing time is long;
- when the threshed is very strict, the number of remained paths become close to zero.
- there are many oscillations in the output results.

3.2 Method2

To deal with the problems of Method1, we proposed and implemented a new method called Method2. The Method2 improved Method1 as follows:

- links connecting the source and destination nodes are not considered for reduction of the search space;
- the algorithm is activated only when the number of the remained paths is higher than a predefined value;
- the threshold value is calculated as the sum of a mean value and a correction value.

In Method1, the threshold values were assigned the same for all links. For this reason, when the threshold value was low, the number of output paths was close to zero. In Method2, we reduced the search space of the algorithm only in the paths that did not include the source destination peers. In this way, the number of remained paths is not zero and we can use strict values for the thresholds.

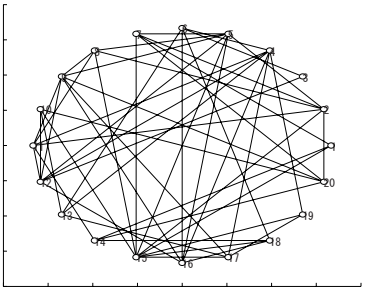


Fig. 11. Generated topology

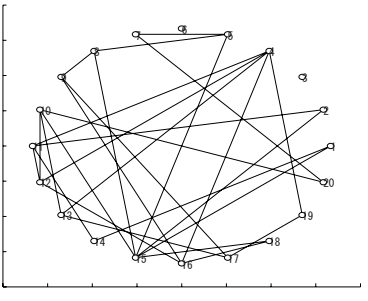


Fig. 12. Topology after network reduction

Table 1. Correction values

Number of Nodes	Correction Value
100	17
90	15
80	13
70	11
60	9
50	5
40	0

In Method1, based on the values of thresholds sometimes the number of output paths is very small. In Method2, when the number of paths is smaller than a predefined value, the process of space search reduction is not carried out.

In Method1 are used fixed values for the thresholds. So, when the parameters are changed in random way, there is a big difference in the number of output values. In Method2, by considering the threshold value as the sum of a mean value plus a correction value, the performance is better than the Method1. The correction values were decided based on our experience as shown in Table 1.

4 Simulation Environment and Results

4.1 Simulation Environment

We evaluated the performance of the Method1 and Method2 by computer simulations. The computer specifications are: CPU: Athlon64 +3200, Memory: 1024MB, OS: Windows XP Pro, and IDE: MATLAB 6.5.

The number of nodes used for simulations in Method1 were 20, 30 and 40, while for Method2 40, 50, 60, 70, 80, 90, and 100. For Method1, we used a small number of nodes because with the increase of the number of nodes the processing time is increased and is very difficult to carry out the simulations. For Method2, we used about 100 nodes, because we have divided the network architecture in different domains and each domain is considered to have about 100 nodes for a better operation of GA.

For simulations, we used 3 parameters: delay, packet loss and bandwidth. The parameters were generated in a random way. We carried out 30 times the simulations and then calculated the average values.

4.2 Simulation Results

Simulation results for Method1 and Method2 are shown in Fig. 13 and Fig.14, respectively. Method2 has a better performance than Method1, because even with increase of the number of nodes, the number of all paths and output paths is not changed too much, while for Method1 we see a big difference between output paths values.

In Table 2 and Table 3, we show also some simulation evaluations for Method1 and Method2, respectively. The number of all paths generated by Method1 is very large, so the algorithm becomes heavy and the processing time is increased. However, in Method2 the number of generated paths is small and we do not carry out the space search reduction process if the number of paths is less than 100. For this reason Method2 is faster than Method1.

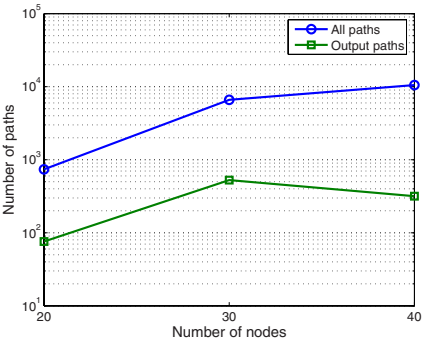


Fig. 13. Method1 simulation results

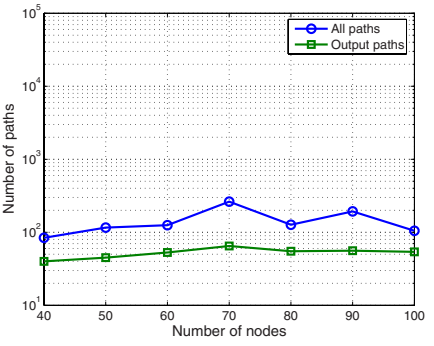


Fig. 14. Method2 simulation results

Table 2. Method1 for 30 nodes

All Paths	Output Paths	Processing Time (s)
5025	346	8.906
358	51	1.516
1563	97	9.906
3033	637	8.421
520	230	3.031
1678	420	5.266
3313	555	5.969
13825	455	20.891
1206	137	2.75
11893	1225	35.266

Table 3. Method2 for 100 nodes

All Paths	Output Paths	Processing Time (s)
19	19	1.938
31	31	1.968
63	63	1.985
95	95	2.109
96	96	2.203
48	48	2.047
78	78	2.078
134	52	2.219
72	72	2.219
132	94	2.281
118	74	2.188

5 Conclusions

In this paper we proposed two methods for space search reduction of GA-based routing in the distributed network architecture. We evaluated the proposed methods by computer simulations. From the simulation results we concluded that the Method2 has a better performance than Method1. For this reason, we will use Method2 for space reduction of the GA algorithm.

As the future work, we would like to implement the proposed agents to carry out the operations in parallel in order to have a faster processing time.

References

1. Guérin, R., Ahmadi, H., Naghshineh, M.: Equivalent Capacity and Its Application to Bandwidth Allocation in High-Speed Networks. IEEE Journal of Select Areas Communications 9(7), 968–981 (1991)
2. Chen, S., Nahrstedt, K.: An Overview of Quality of Service Routing for Next-Generation High-Speed Networks: Problems and Solutions. IEEE Network, Special Issue on Transmission and Distribution of Digital Video 12(6), 64–79 (1998)
3. Baransel, C., Dobosiewicz, W., Gburzynski, P.: Routing in Multihop Packet Switching Networks: Gb/s Challenge. IEEE Network 9(3), 38–60 (1995)
4. Lee, W.C., Hluchyj, M.G., Humblet, P.A.: Routing Subject to Quality of Services Constraints in Integrated Communications Networks. IEEE Network 9(4), 46–55 (1995)
5. Barolli, L., Koyama, A., Yamada, T., Yokoyama, S., Suganuma, T., Shiratori, N.: An Intelligent Routing and CAC Framework for Large Scale Networks Based on Cooperative Agents. Computer Communications Journal 25(16), 1429–1442 (2002)
6. Barolli, L., Koyama, A., Yamada, T., Yokoyama, S., Suganuma, T., Shiratori, N.: A Fuzzy Admission Control Scheme and Its Performance Evaluation. IPSJ Journal 42(12), 3213–3221 (2001)
7. Barolli, L., Koyama, A., Suganuma, T., Shiratori, N.: A Genetic Algorithm Based QoS Routing Method for Multimedia Communications Over High-Speed Networks. IPSJ Journal 44(2), 544–552 (2003)
8. Ikeda, M., Barolli, L., Koyama, A., Durresi, A., De Marco, G., Iwashige, J.: Performance Evaluation of an Intelligent CAC and Routing Framework for Multimedia Applications in Broadband Networks. Journal of Computer and System Science (JCSS) 72(7), 1183–1200 (2006)
9. Barolli, L., Ikeda, M., De Marco, G., Durresi, A., Koyama, A., Iwashige, J.: A Search Space Reduction Algorithm for Improving the Performance of a GA-based Routing Method in Ad-Hoc Network. International Journal of Distributed Sensor Networks (IJDSN) 3(1), 41–57 (2007)
10. Goldberg, D.E.: Genetic Algorithms in Search, Optimization, and Machine Learning. Addison-Wesley, Reading (1989)

Prototype of a Workers' Motion Trace System Using Terrestrial Magnetism and Acceleration Sensors

Nobuyoshi Sato[1], Shouichi Odashima[2], Jun Suzuki[2],
Taiji Ishikawa[2], and Yoshitoshi Murata[1]

[1] Faculty of Software and Information Science, Iwate Prefectural University, Japan
[2] Iwate Plant, Kanto Auto Works, Ltd., Japan
{nobu-s,y-murata}@iwate-pu.ac.jp

Abstract. In quality control of industrial products, it is very important that not only qualities of materials and parts, design, but also a manufacturing work accords to instruction manuals for manufacturing process. Generally, most assembly process such as assembly line and cell production system have processes in everywhere to confirm quality and work in former processes. However, there are some cases that are difficult to be confirmed in later processes. For example, to fix a part by some screws, it is desirable to drive in diagonal order little by little. However, it is not easy to detect in appearance in case of this desirable procedure is violated. Therefore, we are developing a system to improve quality of industrial products. This system deals manufacturing processes as context, and grasp manufacturing process, recognizes context by tracing motions of worker's arm/hand by using terrestrial magnetism sensors and accelerometers. As a prototype of this system, we developed a system that judges attaching process of a particular part is correctly done or not in an accrual vehicle assemble factory. In this paper, we describe outline of the prototype of our workers' motion trace system and algorithms for judgment, and its evaluation.

Keywords: Quality control on industrial products assembly, Motion capture, Terrestrial magnetism sensor.

1 Introduction

Quality control on industrial products is important problems for manufacturers since is may bluster their profits and continuation of their business.

Quality problem in industrial products such as electronic appliances and automobile is divided into three portions: design, quality of parts, and assembly process. For assembly manufacturers, inspection of supplied parts, design and assembly are important factors for quality control because quality of parts and their materials are supplied from other manufacturers. So, every manufacturing sites effort every day to reduce mistakes in manufacturing process such as QC activities.

It is thought that occurrence of defects are mainly caused by violation of defined standard manufacturing process by man and malfunction of industrial robots. In manufacturing works by man, it is difficult to remove all artificial mistakes such as to forget to attach a part, and to attach wrong part. Generally, manufacturing process is

M. Takizawa, L. Barolli, and T. Enokido (Eds.): NBiS 2008, LNCS 5186, pp. 60–70, 2008.
© Springer-Verlag Berlin Heidelberg 2008

organized to detect and correct mistakes in later stages of the process. However, there are some mistakes that are difficult to detect in later stages. A procedure to drive nuts to fix a wheel into a hub is an example of these cases. The correct procedure is to drive nuts little by little in diagonal order. However, if this procedure is not kept, this cannot be detected in appearance of the wheel and nuts. Furthermore, this can cause serious incident such as being off of the wheel.

Therefore, to reduce artificial mistakes, it is effective to detect whether worker's motion is as like manufacturing instruction manual or not. This is realized by recognizing process of manufacture, that is, recognize context of work, and detect motions of arms, hands, and fingers accord manufacturing instruction manual or not. The context can be determined by which part is picked up. Then, this is satisfied by detecting of difference of motion of arms and so on between correct motion and wrong motion. In a system that instruction to pick up a particular part is signaled to worker, it is needed to detect difference between motions which are correct and wrong, including confirmation that designated part is picked up correctly.

We are now developing a workers' motion trace system using terrestrial magnetism sensors and acceleration sensors as a tool for quality control at assembly process [1]. Our system warns if a worker strays predefined standard procedure depending on the worker's context, and if the worker picked wrong parts up. In this paper we describe about the prototype system which is being developed. The prototype system warns in case of the worker's motion deviates the predefined standard motion on corresponding procedure.

Organization of rest of this paper is as follows. In section 2, we will describe related works in motion capture and terrestrial magnetism sensor and acceleration sensors. In Section 3, we will describe about outline of our prototype. In section 4, we will show detailed algorithms to judge motions. In section 5, we show evaluations of our prototype system in a manufacturing process on vehicle assembly factory. Finally, we will conclude.

2 Related Works

GPS realizes trace of a man and object in order of few centimeters. However, utilize GPS is difficult in buildings such as factory because radio signal from satellites cannot be received. Recently there some works on positioning using Wave LAN [2]. These works estimate position of Wave LAN mobile station by using signal strength and delay time from base stations. However, these works cannot be applied to our purpose because they have few meters of errors in positioning. It is conceivable that to trace worker's motion by shooting him/her with video cameras. The paper [3] tries to capture to motion of man who is attached optical makers by using multiple video cameras. There are some works which do not use makers mainly applied to virtual spaces, detection of suspicious man [4]. However, these methods are difficult to apply to our purpose because multiple cameras for a worker cost so heavy.

The others of these works, methods to estimate motion by change of acceleration by moving of man is already introduced. These methods are applied to authentication system by using features of motion of arm [5], and wearable dancing music instruments system [6]. However, since acceleration represents motion in each of moment, so this is

not suitable to estimate continuous motion of a worker in assembly process. Also, since measurement errors of acceleration accumulate to estimate position, correction of absolute positions is needed by another ways than acceleration sensors.

Our proposing system estimates worker's motion by using terrestrial magnetism sensors and acceleration sensors. Terrestrial magnetism can be used everywhere on the earth. In factories, magnetic noises are solicitudes. However, we confirmed that if terrestrial magnetism sensors attached in far position to a certain extent to wires and motors in actual vehicle assembly factory. Also, our prototype uses acceleration sensors auxiliary to detect vibration by screwing and start of worker's motion.

Here, we should pay attention to that terrestrial magnetism sensors measures essentially direction of magnetic north and angle of depression on current location. Therefore, motion such as rotation around elbow and shoulder joints can be detected easily by terrestrial magnetism sensors. However, it is not suitable to detect motions of which changes of terrestrial magnetism are very small such as straight motion.

3 A Prototype of Workers' Motion Trace System

To investigate possibility of realization of our proposed system in actual factory, we implemented a prototype on accrual car assembly factory. The proto-type is intended for a process to attach a fuel tank into vehicle body. In this manufacturing process, a worker mainly attaches and a fuel tank into underbody of a vehicle, and fixes it by using 4 bolts. Since an air impact wrench is used to screw bolts in this process, we attached terrestrial magnetism and acceleration sensors into the air impact wrench. A PC based system judges that whether the worker's action accords to defined standard procedure to attach the fuel tank or not.

We employed Aichi Micro Intelligent Corp.'s AMI601-CG as terrestrial magnetism and acceleration sensor. This sensor device has three components in a package; terrestrial magnetism sensor and acceleration sensor that utilize magneto-impedance effect [7], radio transmitter, and battery. AMI601-CG can measure each of XYZ three axes for both of terrestrial magnetism and acceleration. Measured data is transmitted in each 25 ms via 2.4GHz radio transmitter. MI sensor is recently used as electric compass for portable GPS and GPS equipped mobile phone. Acceleration is

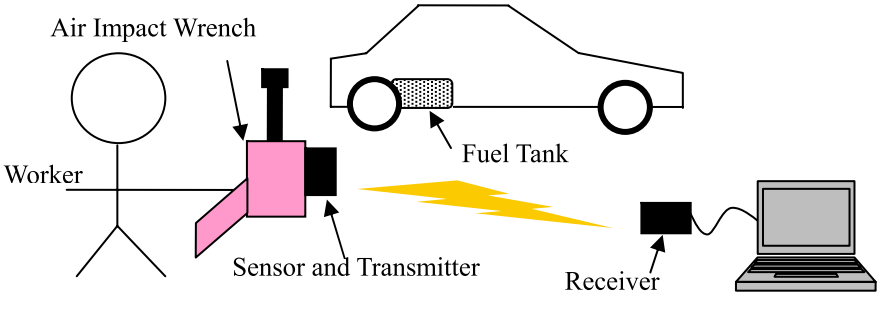


Fig. 1. Structure and modules of a workers' motion trace system

measured by yet another MI sensor for terrestrial magnetism using motion of a moving magnet which is attached through a spring.

Fig. 1 shows that system architecture of the prototype. The prototype consists from following three hardwares:

- Air impact wrench. We assumed that an impact wrench is hold by worker's hand and or hold by instruments which is installed in a factory, and position of the wrench is settled by the worker. For this prototype, we adopted a handgun type air impact wrench. There were no influences to MI sensors because the impact is powered by pressured air and no electrical device is contained in the wrench. Incase of an electrical powered tool is employed, Position that MI sensors to be attached must be considered to avoid influences from electric circuits.
- Terrestrial magnetism and acceleration sensors. As described before, we employed AMI601-CG made by Aichi Micro Intelligent Corp. A plastic package which contains MI sensors and a battery is attached to flat surface of the air impact wrench above by single and both face adhesive tape.
- A PC that a software consists of following modules:
  - Terrestrial magnetism and acceleration sensor interface module. AMI601-CG and its receiver are connected through USB, and MI sensors are seen as a serial port form the software. This interface module reads data through OS's driver and the serial port, and passes the data to upper trace and judgment module.
  - Trace and judgment module. This module compares wave forms from terrestrial magnetism and acceleration sensors and wave forms when a worker's action is correct, and judges the worker's action is co0rrect or not.
  - Standard worker's action database. For different worker, wave forms also differ even if the worker's action is correct for the same process. This database stores features of waveforms of each worker for correct action. This absorbs difference of action in each worker.

Next, we describe about a process to attach a fuel tank that we targeted for this prototype. As shown in Fig.2, the fuel tank is attached in undersurface, rear of a vehicle, on the right under of a rear passenger seat. To reduce possibility that the tank drops out due to vibrations when the vehicle is traveling, 4 bolts must be screwed in correct order in correct torque when the fuel tank is attached. At this time the fuel tank is lifted to regulated position automatically by small elevator, and the worker screws 4 bolts using air impact wrench. Correct order to screw bolts is shown

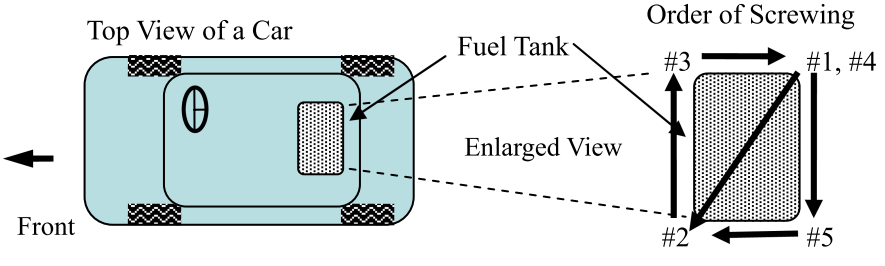


Fig. 2. A fuel tank on car chassis and its correct screwing order to be attached

in Fig. 2. At first, the worker screws a bolt at #1 temporary using electric powered driver. Then, the worker screws 4 bolts tightly at #2 through #5 using air impact wrench in defined correct torque. Order to screw bolts is #2, #3, #4 and #5. Bolt at #1 and #4 is the same.

Our prototype workers& trace system judges the order and place of screwing 4 bolts at #2 through #5 is correct or not. Since torque to screw 4 bolts are managed by air impact wrench, this is out-of-scope in the prototype. In a vehicle assemble factory the prototype targeted, in addition to torques, the number of times the worker pulled trigger of the air impact wrench.

4 Behavior of the Trace and Judgment Module

The trace and judgment module of workers' motion trace system must detect a worker's context. That is, which part of a series of worker's motion to take air impact wrench from a tool chest, pull a trigger to screw a bolt, and put the wench is done by the worker just now. By getting trigger signal from the air impact wrench or being connected to production management system which are installed in the factory, these contexts can be got easily. However, we decided to get worker's contexts only by terrestrial magnetism and acceleration sensors to establish techniques to get context without support from other devices. This technique will be useful when we introduce the prototype system into a factory without production management systems. Of course, we understand advantage to get contexts from another factory devices and systems, and this will be discusses later.

In following, we describe a method to judge a worker's motion accords to determined standard procedure or not.

At first, the judgment module must detect the worker takes air impact wrench from a tool chest, and after motion, the worker puts the wrench to the tool chest. Our workers' motion trace system detects the worker takes / puts the wrench by using changes of terrestrial magnetism sensors outputs. The prototype inferences that the worker takes the wrench when the total sum of changes of terrestrial magnetism sensors outputs in all of X, Y and Z axes overcomes 10mG in 250ms. Also it inferences the worker puts the wrench when the total sum of terrestrial magnetism sensors outputs is less then 5mG in 500ms.

We asked a worker who usually attaches a fuel tank to imitate his usual actions to attach fuel tank into undersurface of a vehicle at a temporary suspended production line in Iwate Plant, Kanto Auto Works Ltd. Then, we got waveform data of terrestrial magnetism sensors and acceleration sensors when the worker attached the fuel tank. Fig 3 is one of terrestrial magnetism and difference of acceleration waveform data from the worker took an air impact wrench to he put the wrench after attaching tank. These waveforms are quarried from long wave forms by the method described above. When the worker is pulling trigger of the air impact wrench, difference of acceleration becomes so bigger due to vibration caused by rotation of the air impact wrench.

Following differences between actual motion when the production line is running and this experiment;

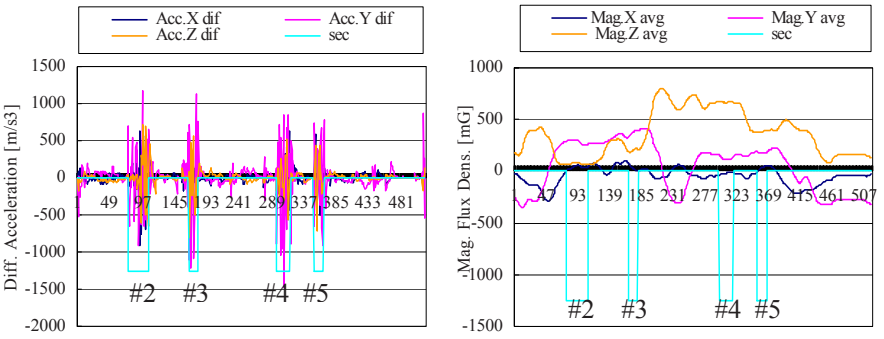


Fig. 3. Differences of acceleration and screwing of 4 bolts by an air impact wrench (left) and Waveforms of terrestrial magnetism sensors when the worker screws bolts in order of #2, #3, #4 and #5 (right)

- Time which the worker pulls the trigger of the air impact wrench sometimes so shorter than actual production. In actual production process, the worker always pulls the trigger more than 10 seconds.
- Since the production line is stopping, location that the worker attaches the fuel tank is always the same, and this location does not move when attaching. In the running production line, since vehicle moves in 6m/s speed, the location the tank is attached moves more than 1m when the worker is attaching it. Waveforms of terrestrial magnetism sensors may be influenced because the worker walks to follow moving vehicle when he is attaching the fuel tank.

Next, the prototype must detect that the worker pulled a trigger of the impact wrench. Our motion trace system judges whether the worker screwed four bolts in correct order or not based on outputs of terrestrial magnetism sensors when the worker pulled the trigger. On the other hand, when the worker is not pulling the trigger, worker's motion and location of the impact wrench is unrelated to screwing bolts. Even if the worker has the impact wrench on his hand, it is unrelated to judgment of order of screwing bolts is correct or not. Our system detects the trigger is pulled by the worker by detecting vibration when a bolts is screw using acceleration sensors. The Following is a method to detect that the trigger is pulled by acceleration:

1. Calculate differences in each of x, y and z axes from when the worker took the impact wrench and he put it onto tool shelf. Let acceleration of each xyz axis at time t be a_{xt}, a_{yt}, a_{zt}, and let these differences be d_{xt}, d_{yt}, d_{zt}. Differences are calculated as following:

$$d_{xt}=a_{xt}-a_{xt-1},\ d_{yt}=a_{yt}-a_{yt-1},\ d_{zt}=a_{zt}-a_{zt-1}$$

2. If at least one of d_{xt}, d_{yt}, d_{zt} overcomes g_{start} and this continues more than t_{start}, our prototype system judges that the worker took the impact wrench. In the prototype, a threshold for acceleration g_{start}=500mg, and threshold for time t_{start}=250ms.
3. The prototype is in state of above 2, if all of d_{xt}, d_{yt}, d_{zt} are less then g_{end} and this continues more than t_{end}, the prototype judges the worker released the trigger. Here, threshold for acceleration g_{end}=500mg, threshold for time t_{end}=500ms.

4. Let corresponding start pulling the trigger and releasing it be a section, and the prototype removes short sections less then t_{min}. Rest sections be sections for judgment s_i (i=1,2,…). Here, t_{min}=500ms in the prototype.

Screwing a bolt in process to attach a fuel tank which the prototype targeted normally requires screwing it continuously 2 or 3 seconds. If the worker after the worker pulled the trigger, about 0.2 sec is needed to vibration overcomes 500mg and to detect rotation of impact wrench by acceleration sensors. On the other hand, differences of acceleration sometimes overcome 500mg by a shock that the worker put the impact wrench into tool shelf. Also, sometimes short sections less then 500ms are shown right after the worker pulled and released the trigger.

Lines marked as "sec" in Fig. 3 shows that sections detected by above method. Sections that marked as #2 through #5 and the lines "sec" indicates 1250 are detected sections by above, and all they corresponds #2 through #5 in Fig. 2. Wave forms of terrestrial magnetism sensors have no largely changes during each of sections except minor changes.

Next, we describe an algorithm to judge the worker's motion accords to correct procedure or not. This algorithm judges based on outputs of terrestrial magnetism sensors agrees in correct rages in bolt screwing position #2 through #5.

1. Calculate averages of terrestrial magnetism sensor outputs m_{xi}, m_{yi}, m_{zi} for each x, y, z axis at section i which the worker pulls the trigger. One or two of x, y, z axes may not be used to judgment if it is / they are thought as unnecessary.
2. Calculate absolute values $|m_{xi}-m_{sxi}|$, $|m_{yi}-m_{syi}|$, $|m_{zi}-m_{szi}|$ of differences between m_{xi}, m_{yi}, m_{zi} and standard values of terrestrial magnetism sensors' outputs at section i for each x, y, z axis. If each $|m_{xi}-m_{sxi}|$, $|m_{yi}-m_{syi}|$, $|m_{zi}-m_{szi}|$ is less than thresholds m_{txi}, m_{tyi}, m_{tzi} at i in each x, y, z axis, corresponding axis is judged as OK which means that worker's motion for this axis and section accords to standard procedure.
3. If there is at least one axis is judged as OK for section i, procedure at section i is judged as it is correct, otherwise, section i is judged as NG which means the worker's motion for this section does not accord to standard procedure.
4. If all of section i are judged as OK, all of motion to attach the fuel tank is judged as OK. Otherwise, all of the motion is judged as NG.

To judge the worker's motion accords to standard procedure using above algorithm, suitable thresholds mtxi, mtyi, mtzi are required. They must satisfy a condition that terrestrial magnetism sensor outputs are out of thresholds if the worker stray standard procedure based on standard terrestrial magnetism sensor outputs in case of correct motion. However, in #4 and #5 on Fig.3 (right), terrestrial magnetism sensors outputs differ only in Z axis. So in this case, Z axis is important to choose thresholds. Oh the other hand, it must be avoided that the prototype system judged NG when the worker did correct procedure. Also, since we now discuss based on data measured in temporary halt production line, choosing threshold may become harsh in the line in operation because vehicle to be the fuel tank is being attached is moving. Furthermore, in the line in operation, it is difficult to enough examination because bad products are made if we let the worker to do wrong assembly motion in intention to examine the prototype. Therefore, a method to choose thresholds must be discussed in our future works.

The prototype we developed this time is not connected with production management system which is already installed in the factory.

Generally, in large factories, production management systems which closely work together with each step of production process are installed. Some of powered tools which utilized in factories connected with these production management systems. That is, there are some items to be managed in impact wrench. The one is the worker taken the wrench or not, the other is haw many times the worker pulled trigger for a process. So, in future, it is thought as enough effective that the prototype systems get these signals from the production management systems so that the prototype system does not misrecognize worker's context. And this will make easier that to judge worker's motion is correct or not. Also, the prototype can detect that mistakes such as the worker pulled the trigger but he didn't set a bolt by combining acceleration sensor's outputs.

5 Evaluations

Here, we describe about procedure to evaluate our prototype system. At first, in a factory of which production line is temporary stopping, we asked a worker to act motions which simulates to attach a fuel tank into undersurface of vehicle, and we measured terrestrial magnetism and acceleration sensors' outputs.

We asked the worker to attach the fuel tank in correct standard procedure and some wrong procedures, and we measured 10 times per each sort of motion. Following is the order of screw bolts in case of correct and wrong motions we asked. Here, the number of combination on order to screw bolts is $\sum_{n=1}^{4} n!$, however, we examined in the cases which seem to occur frequently. The number after # corresponds position of bolts in Fig. 2.

- #2→#3→#4→#5 (standard correct procedure)
- #3→#2→#5→#4 (wrong procedure 1)
- #3→#2→#5 (wrong procedure 2)
- #2→#3→#4→#4 (wrong procedure 3)

Waveform of terrestrial magnetism sensors in case of correct procedure is shown in Fig. 3 (right) before. Fig. 5 shows that one of waveforms in case of wrong procedures.

Next, we calculated standard values of terrestrial magnetism sensor outputs at #2, #3 #4 and #5 m_{syi}, m_{szi} and their thresholds m_{tyi}, m_{tzi} as shown in Talbe 1. Here, we did not employed x axis this time since it has no clear differences in its values at #2, #3, #4 and #5 in all measured waveforms. We used averages of 10 times of motion as m_{syi}, m_{szi} at #2 through #5 in case of correct motion to attach the fuel tank. Also we chosen m_{tyi}, m_{tzi} so that the algorithm described in former section can judge the worker's correct motion as OK, and wrong motion as NG. This is because we cannot conclude 2σ or 3σ is suitable for thresholds based on standard deviation calculated from 10 times of standard correct motions due to they distributes wider than our assumed values. So, in future, we must increase the number of times to measure

Table 1. Table 1 Standard values of terrestrial magnetisms and their thresholds in case of the worker's motion is correct

	m_{syi}	m_{tyi}	m_{szi}	m_{tzi}
#2	260.48	30	75.64	30
#3	390.68	50	259.94	50
#4	126.35	40	665.84	45
#5	162.96	30	384.19	30

correct motions, and we should develop a method which automatically determines thresholds based on measured data.

Based on standard values of terrestrial magnetism sensors outputs and their thresholds calculated by above procedure, we examined our prototype. We gave 10 waveforms for each of the worker's correct motions and wrong motion 1 through 3, and our prototype system. Our prototype system successfully judged in all of correct motions and wrong motions.

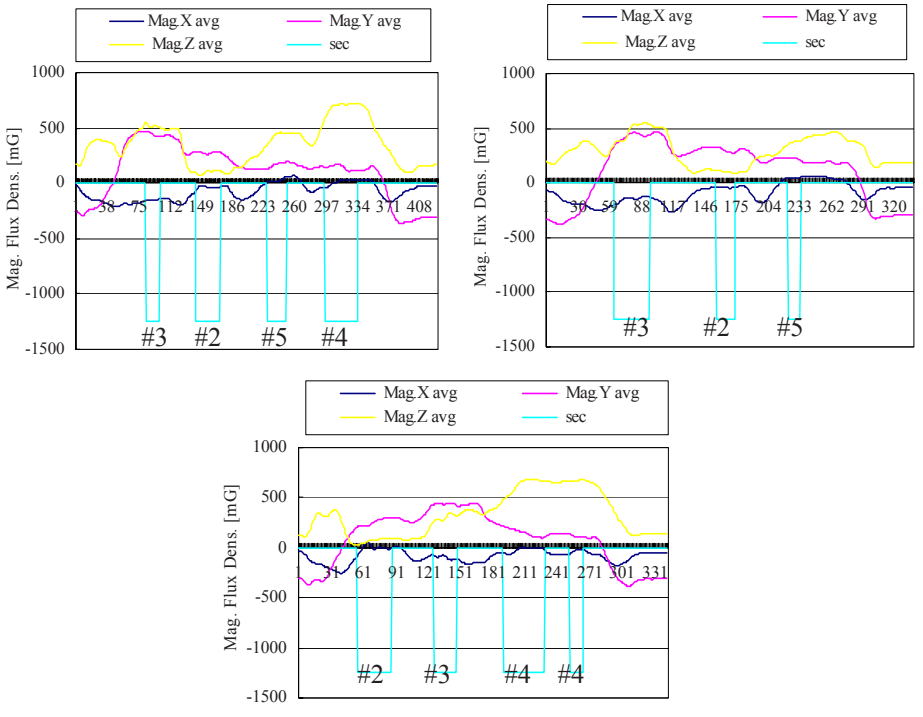


Fig. 4. Waveforms of terrestrial magnetism in various worker's motions. Upper left: #3→#2→#5→#4 (wrong procedure 1). Upper right: #3→#2→#5 (wrong procedure 2). Lower right: #2→#3→#4→#4 (wrong procedure 3).

6 Conclusions

In this paper, we described the prototype of our workers' motion trace system by terrestrial magnetism sensors and acceleration sensors. This system warns if mistake such as order of screwing is detected in assembly process of industrial products. Some mistakes in production process cab be found in later process, however, some mistakes such as wrong order of screwing cannot be founded because it an correct work have the same appearance and correct torque. So order of screwing in some process like attaching tire and wheel into hub may make serious incidents such as dropping out of tire and wheel when a vehicle is traveling.

Our prototype system is developed for fuel tank attaching process at Iwate Plant, Kanto Auto Works Ltd., to verify effectiveness of workers' motion trace system which employs terrestrial magnetism sensors and acceleration sensors. We measured motions data of a worker which imitates a process to attach a fuel tank into undersurface of an in-process vehicle in suspended production line at holiday. Then, we evaluated that the prototype can judge both of correct and wrong motions correctly using the measured data. As the result, our prototype system successfully judged all of wrong motions as wrong motion, and correct motions as correct motions.

As future works, following works are required for practical use of our workers' motion trace system. First, examination of the prototype at operating production line is needed. Influences on terrestrial magnetism sensors outputs in different workers, how to determine standard values of terrestrial magnetism sensors outputs on correct motions, and thresholds for judgment are also matters to be solved. At least, we confirmed that different worker's terrestrial sensors outputs differ even if motions of two workers accord with correct standard procedure. Furthermore, a networked production management system is installed in the factory we targeted, and terminals are installed in each of process to assemble vehicles. So we think we need to make our prototype system to be connected and cooperate with the production management system. Also, we need to make our terrestrial magnetism and acceleration sensor device smaller so that workers easily wear our sensor device.

References

[1] Sato, N., Murata, Y.: Quality Control Schemes for Industrial Production by Workers' Motion Capture. In: Proc. of The 2nd International Workshop on Telecommunication Networking, Applications and Systems (TeNAS 2008) (to be appeared, 2008)

[2] Bahl, P., Padmanabhan, V.N.: RADER: An In-Building RF-based User Location and Tracking System. In: Proc. of IEEE INFOCOM 2000, pp. 775–784 (2000)

[3] Yonemoto, S., Matsumoto, A., Arita, D., Taniguchi, R.: A Real-time Motion Capture system with Multiple Camera Fusion. In: Proc. of International Conference on Image Analysis and Processing (ICIAP), pp. 600–605 (1999)

[4] Trivedi, M.M., Huang, K.S., Mikić, I.: Dynamic Context Capture and Distributed Video Arrays for Intelligent Spaces. IEEE Transactions of Systems, Man, and Cybernetics – Part A: Systems and Humans 35(1), 145–163 (2005)

[5] Ohta, M., Namikata, E., Ishihara, S., Mizuno, T.: Individual Authentication for Portable Devices using Motion Features. In: Proc. of the 1st International Conference on Mobile computing and Ubiquitous networking (ICMU 2004), pp. 100–105 (2004)
[6] Fujimoto, M., Fujita, N., Tsukamoto, M., Tagawa, K.: A Wearable Dancing Musical Instrument System using Acceleration Sensors. In: Proc. of Multimedia, Distributed, Cooperative, and Mobile Symposium, pp. 1215–1222 (2007) (in Japanese)
[7] Mohri, K., Kohzawa, T., Kawashima, K., Yoshida, H., Panina, L.V.: Magneto-Inductive Effect (MI Effect) in Amorphous Wires. IEEE Transaction on Magnetics 28(5), 3150–3152 (1992)

Framework Design Supporting QoS-Power Trade-Offs for Heterogeneous Networked Systems

Christos Antonopoulos, Evangelos Topalis, Aggeliki Prayati, Spilios Giannoulis, Antonis Athanasopoulos, and Stavros Koubias

Applied Electronics Laboratory, Department of Electrical & Computer Engineering
University of Patras, Rio Campus, Greece
{cantonop,topalis,prayati,sgiannoulis,athan,
koubias}@ee.upatras.gr

Abstract. [1]Two of the main research efforts in wireless systems, nowadays, are the Power awareness and Quality of Service (QoS) integration. As frameworks are developed to handle dynamic reconfiguration, the need for a power optimization methodology to investigate alternative cross-layer configurations is critical. However, as networks become more complex and energy savings become crucial, this leads to the consideration of constructs for treating QoS-power trade-offs and adjust to the heterogeneous nature of network systems. In this paper, we propose an interoperable framework design for heterogeneous network systems, where communication is treated transparently and enhancements are proposed to improve QoS by the definition of a framework also supporting dynamic power optimization.

Keywords: Interoperability framework, 802.11x, heterogeneous networked systems, QoS support.

1 Introduction

Heterogeneous wired/wireless networks need an interoperable architecture in order to integrate communication constructs in advanced control applications. As the complexity degree of such applications increases, variant network traffic must be carried through different media in a transparent way. Interoperability is thus intended for a hybrid wireless-wired network system, where communication is treated seemingly homogeneous to the higher layers. Such an interoperable framework architectural design must take into account two different network architectures and make them interoperable, by combining the well known wired Ethernet able to cope with high time constrained demands with the upcoming ad hoc wireless networks, which still lack in all performance parameters compared to the wired Ethernet networks.

[1] The work reported here was performed as part of the ongoing research Program PYTHAGORAS II and funded by the European Social Fund (ESF), in particular by the Operational Program for Educational and Vocational Training II (EPEAEK II).

M. Takizawa, L. Barolli, and T. Enokido (Eds.): NBiS 2008, LNCS 5186, pp. 71–80, 2008.
© Springer-Verlag Berlin Heidelberg 2008

From the user point of view, it is crucial that the whole system functions equally well,independently from any heterogeneity. Moreover, the innovative protocols for the wireless part tend to consider QoS parameters in order to improve network performance. However, the mobile nature of wireless systems requires that the network node life-cycle is extended in terms of power consumption minimization. Power-awareness involves also higher-level parameters and thus a cross-layer model is required to handle power aspects in distributed networks of high heterogeneity. When dealing with a hybrid wired/wireless network, questions arise regarding QoS- and power-awareness issues especially concerning the wireless part of the hybrid network. Integration of QoS and power awareness in wireless networks is nowadays a growing research area as high throughput, timeliness and power efficiency is demanded by several home and industrial applications [3][4]. However, trade-offs especially between QoS parameters and power consumption must be considered to a network with mobile nodes [5][6].

In this paper, an interoperable device architectural model is proposed, supporting network heterogeneity and QoS tenability. In section 2, related work is discussed. In section 3, the characteristics and requirements of the targeted application scenarios are presented, and in section 4 the interoperability framework design is presented. Section 5 illustrates the case study of the proposed framework scheme and conclusions are derived in section 6.

2 Related Work

Proving interoperability among heterogeneous systems comprise a quite extended research field. Consequently, new concepts, approaches and entities were introduced. A newly introduced concept is the Aspect-oriented programming, which facilitates interoperability and provides programmers with appropriate constructs for separating cross-cutting concerns leading to advanced modularization. The main requirement in order to achieve that is to allow the break down of a program to distinct parts that overlap as little as possible [1],[2]. AspectJ is a specific language offering this concept to various research projects [7] and papers [8].

Integration of QoS awareness in wireless networks is a growing research area, as timeliness and application-related constraints must be addressed by the network components. Dynamic variation of CPU parameters like frequency and voltage takes into consideration demands of the OS services and application QoS demands though a flexible framework [9]. An adaptive Point Coordination Function (PCF) algorithm is presented in [10], based on recent polling feedback. The point coordinator uses an AIMD algorithm in order to maintain an active polling list. This architecture results to better medium utilization and successful poll rate.

Related to the aforementioned approaches significant are the respective standardization efforts. The Function Block comprises the main concept defined by the IEC 61499 [11] standard, incorporating descriptions of common field devices' requirements as a generic architecture. An important extension is provided by the IEC 61804 [12] towards the design of modular reusable and distributed Industrial Process and Measurement Control System (IPMCS) using Engineering Support System (ESS).

For some time now, wireless networking devices, proprietary network stacks and solutions in general greatly differ in requirements, characteristics, communication algorithms and important supported capabilities. Under this approach, interoperability is a critical demand and it comes as no surprise that a significant variety of respective proprietary solutions exist. Siemens as a key player in wireless networking presented its proposed solution with PROFInet approach is based on the PROFInet standard [13] incorporating Microsoft's COM/DCOM standard and technologies such as TCP/IP, OLE and ActiveX. On the other hand, the IDA group proposes its own approach [14], which poses as main goal the definition of an infrastructure for open heterogeneous automation application in modular distributed structures. Furthermore, significant research effort is devoted trying to solve the problem of interoperability between different devices incorporating different functionality and at the same time comprise proprietary products [15], [16]. Finally, an aspect worth pointing out is integrating QoS aspects in incompatible systems following various approaches from UDP/TCP improvements [17] to spatial reuse advantages exploitation using directional antennas [18] through appropriate prototype frameworks.

Considering network performance optimization and power consumption minimization the approach that is continuously gaining ground is cross-layer design. The main objective of such designs is to identify and use existing interactions, correlations and dependencies among different network stack layers. A featuring example presented in [19] introduces a new routing protocol that utilizes MAC related information for minimizing excessive power consumption. Even though feasible routes are categorized according to expected power consumption, another option is the consideration of multiple routes and distributed data transmission among them resulting in optimal overall consumption. Power conservation can be achieved by turning off the radio which is one of the most significant sources of consumption. Therefore, the authors in [20] present a respective "on/off" scheduler based on close cooperation between MAC and routing layers. However, strict synchronization demands are imposed as well as the presence of a centralized point of control, while only specific applications are evaluated.

3 Analysis of Network Particularities with Respect to the Targeted Application Scenarios

The network architecture consists of two physical media, one wired and one over the air. The hybrid topology corresponds to the wireless 802.11-based part and to the wired Ethernet-based protocol, as shown in Fig. 1. In order for multimedia traffic to be carried through transparently over the two media, the wireless nodes must transfer the audio packet through the wireless network to the wired Ethernet under end-to-end time and quality constraints. As all multimedia applications the user perceptive quality defines tight QoS requirements that the hybrid network should support.

The system under study imposes the need for transparent multimedia traffic over heterogeneous wireless-wired networks, where end-to-end time constraints must be met and the need for maximizing the system-wide life-cycle is more than imperative. The system consists of time constrained data flows (i.e. audio streams) transferred over a hybrid wireless-wired network topology. The analysis and testing of respective

applications over the hybrid network architecture reveals the following bottlenecks that are treated by the proposed framework:

i. Time constrained data flows transfer over the wireless medium with enhanced routing and MAC protocols provide high throughput and low packet loss.
ii. Time constrained data flows transparent transfer over the wired-wireless connection through the interoperable device architecture.
iii. Trade-off handling between application layer constraints and Network/MAC layer performance.
iv. Power savings with respect to optimal QoS.
v. Cross-layer designs facilitating interoperability as well as network performance.

Apart from inter-network QoS consideration, the proposed enhancements for routing and scheduling algorithms in the wireless network nodes at network and MAC layer respectively, render the wireless network QoS-aware and as such related parameters are treated by the framework for QoS versus power optimization. Therefore the proposed framework must fulfill the demands of such a network architecture and thus it must consider many network parameters such as:

i. Channel conditions (either of the wired or the wireless part of the network).
ii. Optimum route selection under various metric depending on priorities, energy constrains, link quality, avoidance of bottlenecks etc.
iii. Capabilities of node having different technical characteristics.

4 Interoperability Framework Design and Specification

The devices involved in the described scenarios include wireless stations, gateways, Ethernet-based PCs, laptops and PDAs. From the user perspective these devices must work transparently and any communication must be performed efficiently with respect to QoS and heterogeneity. The proposed framework must provide the adequate mechanisms for exposing the different devices particularities in a unified manner to the application.

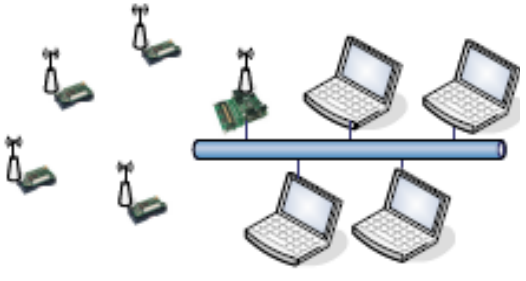


Fig. 1. Network topology

4.1 Framework Architecture

As shown in Fig. 2, the Tunnelling Manager is responsible for monitoring the system, coordinating the system functional components by determining the optimal configuration scenario that minimize power with respect to traffic performance parameters. The optimal scenario is generated based on predefined action rules that investigate alternative configurations for improved performance.

As expected, an efficient framework approach must provide the following three functional units. Firstly, information acquisition must be provided, by any required layer of the network stack. This is very important since various objectives correlate with more than one layers, but also because specific objectives demand combinative use of information provided by different layers. That leads to the next element which is information processing. All the acquired information is used by algorithms and processes aiming to determine the optimum state of operation for the station depending on functional demands as well as network conditions. The role of the final element is to deliver the outcome of considered algorithms and processes. These elements must be able to communicate with all layers since the outcome of any decision making process may influence the operation and characteristic of any layer of even combination of them.

Based on analysis of these requirements, the elementary functionalities are identified and presented. Firstly, it is required that all application communication layers parameters be configurable at real-time. Related to this is the demand of being able to manage the network components parameters at real-time as well. However, at all times and above all, the steady and normal operation of the system must be assured. By incorporating adequate algorithms provision of optimal network functional parameters

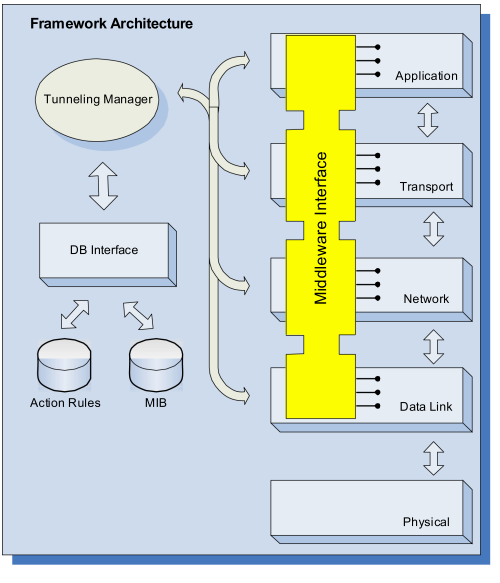


Fig. 2. Framework architecture design

configuration must be assured with respect to power and availability. Additionally an increasingly critical demand for any nowadays network system is QoS support. This demand is met by combining application requirements and historical data on the overall system performance.

Focusing on Fig. 2, the main goal of the Tunneling Manager is to apply optimization algorithms trying to provide the optimum trade-off among network parameters. The advertisement of these parameters is provided by the system sub-layers using the Middleware Interface. Optimum parameter configuration concerning all involved network layers is the result of combining predetermined rules and information taken from the Management Information Base (MIB).

The architecture's main goal is to optimize network performance in terms of power consumption, QoS and time constrains. Especially the latter requirement is of outmost importance since nowadays wireless networks are expected to support not only best-effort communication but soft and hard real time applications as well. The increased complexity in the proposed case study is introduced by added network heterogeneity where respective wired and wireless network capabilities are vastly different. Therefore, in order to evaluate and assure such efficient performance to any point of the network attention must be paid to maximum delay observed as well as packet delay distribution, which are related to hard and soft real time respectively. The previous observation however takes into consideration that different applications have different demands on bandwidth, delay, throughput and various network parameters.

4.2 Generic Device Model Considerations

When it comes to modeling heterogeneous network devices in an interoperable environment, the following issues must be supported:

User interface: the device capabilities must be described in a standardized format i.e. XML and a graphical user interface must support the presentation of different devices' description in a consistent way.

Modularity: the framework must be portable in order to provide network and platform independency.

Interoperability: the devices must be modeled with mechanisms that support the plug & play characteristic, whilst adequate middleware services must provide device technology independency.

Continuity: communication among heterogeneous devices must support seamless handover, service continuity as well as session continuity.

Configurability: the framework must provide services for accessing the device capabilities including static information and dynamic parameter changes. These services may range from simple monitoring mechanisms to control and reconfiguration facilities.

Security: a common and very important characteristic of devices in wireless environments is security provisioning. Therefore, the framework must support services for authorization, user-level intrusion detection and authentication.

Networking: interconnectivity is a basic requirement when talking about heterogeneous devices that must interoperate. Robustness and scalability are two other aspects that

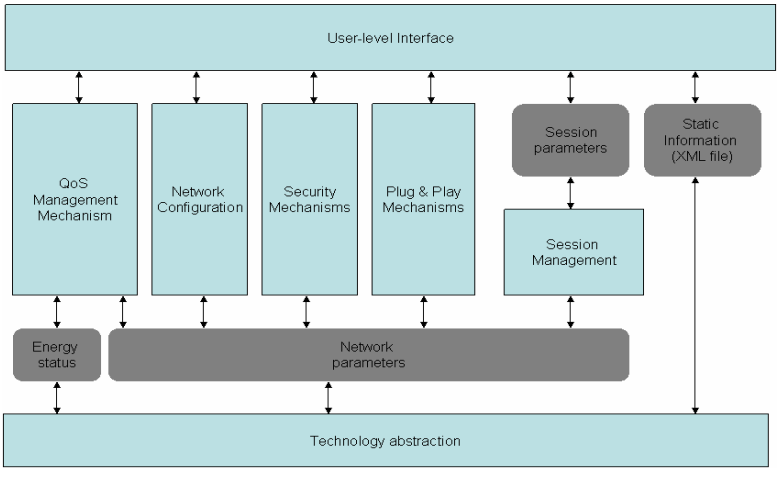


Fig. 3. Device modeling aspects

must be considered by the framework services together with network configuration services.

QoS: Quality of Service provisioning requires optimization of several parameters, mostly related to the networking functionality of the devices. In that sense, these parameters must be exposed to the framework in order to be handled and dynamically configured for ensuring overall system reliability.

As shown in Fig. 3, a device model from the system point of view supports the framework required services, generalizing all the concepts of existing devices implementing interoperable distributed heterogeneous network architectures.

In the lower layer of the device architecture technology abstraction provides the dynamic and static information to be handled by the higher layers mechanisms. Dynamic information consists of the network and energy parameters, whilst static information is related to the device serial number, ID and other manufacturer–provided information. The energy status and the values of the network parameters are exposed by QoS management mechanisms in order to support optimization techniques and priority handling. Networking information includes a wide range of communication parameters set depending on the underlying device technology. These parameters are used by the network configuration services, as well as Plug & play mechanisms. Finally, security mechanisms include authentication, authorization and intrusion detection services. Network parameters are also handled by session-level management covering a higher-level conceptual grouping of parameters ranging from an application session, to end-to-end communication links or even data dedicated to application-specific services. In this sense, the session management mechanisms extract session-related information that is in its turn exposed to the framework services. As far as static information is concerned, this is provided in a standardized format and is available to the framework for supporting any kind of device management mechanism, may this be the off-line application design procedure or even the run-time device re-configuration.

5 Case Study

The applicability of the proposed framework is demonstrated by the mapping of the proposed interoperability architecture to an implemented interoperable heterogeneous system and enhanced with the implementation of the model interfaces. The driver application scenario for our demonstrator is shown in Fig. 4, consisting of a networked system with Ethernet and wireless connection, 2 PCs, a Crossbow gateway and one wireless MICA2 mote.

The application consists of monitoring the temperature sensed by the wireless mica2 mote in the remote wired PC projected in a GUI. The initiator mote generates a data packet and transmits it over the radio to the Stargate (last wireless hop). Using the Ethernet interface, which is provided by the Stargate, the data packet is forwarded over the Ethernet cable until it reaches the final wired destination. In order to also test the

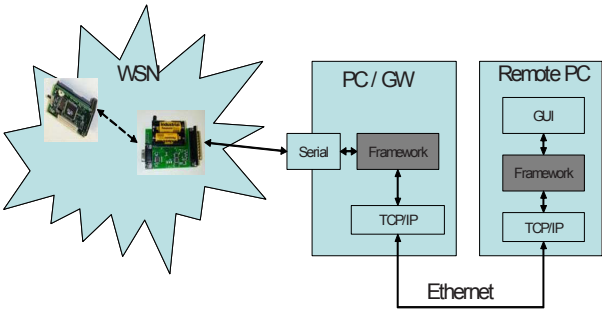


Fig. 4. Demonstrator setup

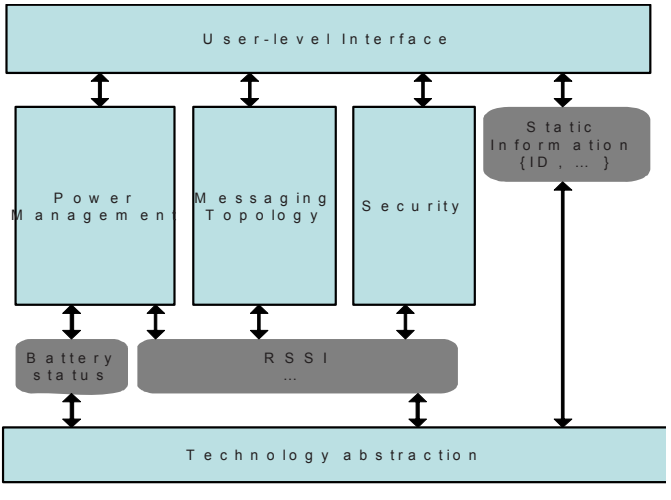


Fig. 5. Device interpretation based on interoperability framework guidelines

reverse communication direction, the LEDs of the wireless mote are controlled remotely by the aforementioned GUI. The Crossbow gateway is programmed in NesC, exposing the services shown in Fig. 4. The wired connection between the two PCs is based on TCP/IP interface, allowing for communication services, as well as QoS handling. The GUI on the remote PC is based on the framework services providing additionally topology and mote static information monitoring.

6 Conclusions

In this paper a framework architecture design has been presented for supporting interoperability in heterogeneous networked systems. The proposed framework addresses the requirements of application running on heterogeneous platforms in order to provide the adequate mechanisms for exposing the different device particularities in a unified way to the higher system layers. A device conceptual architecture was defined for supporting the framework required services, generalizing all the concepts of existing devices implementing interoperable distributed heterogeneous network architectures. Currently, work is being curried in implementing the presented architecture.

The framework architecture allows for future integration of additional modules, should more impact factors of other layers be required. The system design is flexible and the UML model open and adaptable for extension to cover applications of different multimedia particularities or even fulfill the need of multi-dimensional trade-off solving.

References

1. Kiczales, G., et al.: Aspect-Oriented Programming. In: Akşit, M., Matsuoka, S. (eds.) ECOOP 1997. LNCS, vol. 1241, pp. 220–242. Springer, Heidelberg (1997)
2. Eichberg, M.: Component-Based Software Development with Apsect-Oriented Programming. Journal of Object Technology, ETH Zurich, Chair of Software Engineering (2002)
3. Unsal, O.S., Koren, I.: System-Level Power-Aware Design Techniques in Real-Time Systems (Invited paper). Proceedings of the IEEE, Special Issue on Real-Time Systems 91 (July 2003)
4. Van Antwerpen, H., et al.: Energy-Aware System Design for Wireless Multimedia. Panel on Platforms and Tools for Energy-Efficient Design of Multimedia Systems, Design Automization (2003)
5. Conti, M., Gregori, E.: Optimization of bandwidth and energy consumption in wireless local area networks. In: Performance evaluation of complex systems: techniques and tools, pp. 435–462. Springer, Berlin (2002)
6. Takahashi, E.S.C.: Application aware scheduling for power management on IEEE 802.11. In: Proceedings of the 2000 IEEE International Performance, Computing, and Communications-Conference, pp. 247–253 (2000)
7. The TORERO consortium: Deliverable 2.1 Integrative design and development of web-enabled control system design methodology (internal draft) (2003)
8. Tangermann, M., Schwab, C., Kalogeras, A.P., Lorentz, K., Prayati, A.S.: Aspect-Orientation of Control Application Code for Distributed Automation Systems: The TORERO Approach. In: IEEE Proceedings JTRES, 2003 (November 2003)

9. Yuan, Nahrstedt, Adve, Jones, Kravets: Design and Evaluation of a Cross-Layer Adaptation Framework for Mobile Multimedia systems. In: ACM MMCN (2003)
10. Dong, X.J.: Adaptive polling algorithm for PCF mode of IEEE 802.11 wireless LANs. ELECTRONICS LETTERS 40(8) (April 15, 2004)
11. IEC 65/240/CD (61499): Function Blocks for Industrial Process Management and Control Systems Part 1: Architecture Public Available Specification
12. IEC 61804-2 International Standard, Function blocks (FB) for process control – Part 2: Specification of FB concept and Electronic Device Description Language (EDDL) (2004)
13. PROFInet – Architecture Description and Specification, V 1.0, PNO (August 2001)
14. iDA – Architecture Description and Specification, V 1.0 (November 2001)
15. Prayati, A., Koulamas, C., Koubias, S., Papadopoulos, G.: A Methodology for the Development of Distributed Real-Time Control Applications With Focus on Task Allocation in Heterogeneous Systems. IEEE Transactions on Industrial Electronics 51(6) (December 2004)
16. Antonopoulos, C., Athansopoulos, A., Giannoulis, S., Prayati, A., Topalis, E., Koubias, S.: A Framework Architecture Supporting QoS-Power Trade-offs for Heterogeneous Network Systems. In: IEEE International Conference on Emerging Technologies and Factory Automation (ETFA 2005), Catania, Italy, September 19-22 (2005)
17. Lee, S.-B., Ahn, G.-S., Campbell, A.T.: Improving UDP and TCP Performance in mobile Ad Hoc Networks with INSIGNIA. IEEE Communications Magazine (June 2001)
18. Roy, S., Saha, D., Bandyopadhyay, S., Ueda, T., Tanaka, S.: A Network-Aware MAC and Routing Protocol for Effective Load Balancing in Ad Hoc Wireless networks with Directional Antenna. In: Proc. of the Fourth ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc 2003), Annapolis, Maryland, USA, June 1-3 (2003)
19. Safwat, A., Hassanein, H., Mouftah, H.: Optimal Cross-Layer Designs for Energy-Efficient Wireless Ad Hoc and Sensor Networks. In: IEEE IPCCC (2003)
20. Sichitiu, M.L.: Cross-layer Scheduling for Power Efficiency in Wireless Sensor Networks. In: INFOCOM (2004)

Route Cache Based Load Balancing Scheme for Mobile Ad-Hoc Networks

Young-Duk Kim, Jin-Wook Kim, Won-Seok Kang, and Dong-Ha Lee

Dept. of Advanced Industrial Science and Technology
Daegu Gyeongbuk Institute of Science and Technology
Daegu, 700-742, Korea
{ydkim,jwkim,wskang,dhlee}@dgist.org

Abstract. Recently, many on demand routing protocols are suggested to support mobile ad hoc network (MANET) which is self-organized network. The most represented protocols are on-demand routing schemes such as DSR and AODV. These protocols set up routing paths by flooding Route Request (RREQ) packets during route discovery procedure and receive Route Reply (RREP) packets from destination nodes. In order to reduce route discovery latency, every node may use the previous routing information in its route cache and reply RREP instead of the destination. Although the route cache mechanism is simple and efficient approach to enhance the performance, it may result in unnecessary latency when it uses incorrect cache information. In addition, a certain node located around source can be easily congested because it replies every RREQ with RREP and all data flows are concentrated on this node. In this paper, we suggest a dynamic cache monitoring scheme and suppress RREP on heavily congested node to achieve load balancing. To do this, we have defined appropriate queue threshold values and parameters for congestion resolution. We also propose a solution for RREP storm problem which is another side effect of the route cache. Finally throughout the simulation, we have illustrated that proposed scheme shows better performance than the standard algorithm in heavily congested environments.

Keywords: MANET, Route Cache, Load Balancing, Route Reply.

1 Introduction

In Mobile ad hoc network (MANET), all nodes communicate with each other by using wireless channel and have mobility. This means MANET is a self-organized network and there is no infrastructure such as a base station or dedicated cabling. In this environment, all nodes need to find reasonable route and recover the route according to sudden and arbitrary topology changes. In order to support this dynamic MANET, many routing protocols have been suggested such as DSDV (Destination Sequence Distance Vector) [1], DSR (Dynamic Source Routing) [2], AODV (Ad-hoc On-demand Distance Vector) [3]. DSDV is a table driven based protocol and maintains its routing tables with periodic exchanges of control packets. Although DSDV obtains prompt route information from routing tables, it has a serious drawback of high overhead with control packets. DSR and ADOV are on-demand

M. Takizawa, L. Barolli, and T. Enokido (Eds.): NBiS 2008, LNCS 5186, pp. 81–90, 2008.
© Springer-Verlag Berlin Heidelberg 2008

protocols which have been proposed to improve the problems of table driven schemes. They transmit control packets for route discovery only when they want to transmit data packets, which can significantly reduce control message overhead. In addition, these on-demand protocols have a route cache mechanism which offers prompt route information to the source node. The route cache maintains historic entries which are obtained from previous successful data transmissions. By using the route cache, when an intermediate node receives RREQ, it doses not need to rebroadcast RREQ but it can reply with RREP, instead. However, the route cache mechanism may make all source nodes continuously transmit data packets to a certain intermediate node, which leads the node to be seriously bottlenecked. When a node is congested, several problems such as packet loss by buffer overflows, long end-to-end delay of data packets, poor packet delivery ratio, and high control packet overhead to the reinitiate the route discovery procedure can occur. In addition, the congested node consumes more energy to route packets, which may result in network partitions. These problems are continuously growing due to increase of multimedia traffic and real time applications in MANETs.

In this paper, we propose a new efficient scheme which can prevent traffic concentration to a certain node by dynamical suppressing RREP with route cache when an intermediate node detects the congestion symptom. In addition, we also propose an algorithm to define congestion environment with various parameters.

The rest of this paper is organized as follows. In Section II, we review previous works which have suggested load balancing routing and various schemes for route cache management. In Section III, we illustrate detail operations of our proposed protocol. Performance evaluation by simulations is presented in Section IV. Finally, concluding remarks are given in Section V.

2 Related Works

2.1 Routing Protocols with Load Balancing

There are several routing protocols which have considered load balancing as the primary route selection criterion during route discovery execution. In DLAR (Dynamic Load Aware Routing Protocol) [4], the traffic load is defined as total number of buffered packets which pass through the entire route. Then the destination node chooses the least loaded route. And in MCL (Routing Protocol with Minimum Contention Time and Load Balancing) [5], the traffic load is defined as the number of neighbor nodes which content with a source node for channel acquisition. In CRP (Congestion-adaptive Routing Protocol) [6], although the number of packets currently buffered in interface is also defined as network load, the congestion is classified into three statuses, which are red (very likely congested), yellow (likely congested), and green (far from congested). If a node is aware of congestion symptom, it finds a bypass route which will be used instead of the congested route. Although these protocols focus on avoiding congested routes or make backup routes, they do not utilize the route cache mechanism due to accumulation of traffic information for entire routes. However, if the congestion situation does not happen, these operations will result in long route discovery latency by suppressing the route cache mechanism.

Moreover, unconditional RREQ flooding consumes not only high channel bandwidth but also battery energy of all nodes.

2.2 Traffic Distribution with Route Cache Management

Workload Based Load Balancing [7] defines concentrated traffic on a certain node as workload which is the number of buffered packets. It also defines a threshold value to determine congestion symptom. When the workload increases than predetermined threshold value, the bottlenecked node selectively drops RREQ packets to prevent its participation to the route discovery procedure. As a result, the number of flooding highly decreases and data packets can be routed to alternative path, which can achieve load balancing. However, it is not considered the situation of other neighbor nodes and it just drops all RREQ packets when it is congested. The intemperate suppression of RREQ can result in not only route discovery failure but also network partitions. Figure 1 shows an example of the RREP problem. When all nodes, A, B, C try to transmit packets to destination node D, node E should relay all these packets and becomes bottleneck. Then, if another node transmits RREQ packets through node E for route discovery, the load balancing mechanism triggers and drops the RREQ on purpose. However, there is only one route between node E and node F. Therefore, it results in network partition between E and F and the other nodes except A, B, C can not communicate with destination. Moreover, because the bottlenecked node E and its neighbor nodes may transmit RREP with their route cache information, the congestion situation can not be resolved and traffic is more concentrated on node E.

EAODV [8] proposes a selective route cache management scheme for traffic distribution. The main operation of EAODV is to detour RREQ packets by using address pair which is composed of source and destination. Although it can offer an alternative route with traffic distribution, it makes the route longer and suffers from high end-to-end transmission delay.

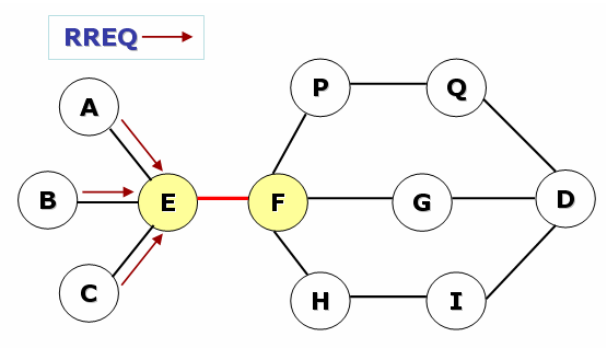


Fig. 1. Example of congestion at relay node

3 Proposed Scheme

3.1 Basic Operation

In general on demand routing protocols, the source node broadcasts RREQ and receive RREP from the destination node to find an optimal path. The RREQ flooding scheme is very reasonable approach when source does not have any information about the route to destination in dynamic network topologies. In order to reduce RREQ flooding overhead and route discovery latency, most on demand protocols use the route cache mechanism. The intermediate node which receives RREQ looks up the cache table. If there is correspondent entry in the route cache, the node transmits RREP immediately.

The proposed scheme basically allows to suppress RREP packets which is obtained from the route cache when it detects congestion symptom. In order to determine whether nodes should suppress RREP or not, all nodes monitor their queue length (Q_{Len}) in network interface and define a queue threshold value ($Q_{Threshold}$) to compare with Q_{Len}. The first step of the proposed scheme is to receive RREQ packets from source node as usual on demand routing protocols. If the receiving node is final destination, it transmits a RREP packet to source node immediately. Otherwise, the intermediate node checks the route cache to find the correspondent entry. Then, when there is no correspondent entry, the node rebroadcasts RREQ packet to continue the route discovery process. If there is a matched cache entry, the node compares measured Q_{Len} with $Q_{Threshold}$ to determine that it should generate the RREP packet or not. At this moment, if Q_{Len} is greater than $Q_{Threshold}$, the node considers that it is congestion situation and suppresses RREP. As a result, the source node can detour the congested intermediate node and prevents packet loss by buffer overflows. Although the intermediate node starts to rebroadcast RREQ packets instead of RREP transmission, it has more latency to receive RREP packet from the destination node due to queuing delay in it. This means another route which excludes congested node has more probability to establish a route to destination. Consequently, proposed scheme can resolve congestion and achieve load balanced routing.

Another advantage of RREP suppression and RREQ retransmission is the prevention of a stale cache problem. In general on demand schemes, the route cache maintains recent information which is obtained from successful packet transmission. However, the cache accuracy decreases especially in dynamic changing topologies due to node mobility and high channel error rate. These problems lead to not only high packet loss rate but also high end-to-end packet delay to reinitiate route discovery process. Moreover, the failure of route discovery procedure is severe overhead to the whole network and it may disturb other data communications between neighbor nodes. Therefore, the operation with RREP suppression and RREQ retransmission is relatively more advantageous to obtain fresh route information.

3.2 Congestion Monitoring Algorithm

When an intermediate node properly suppresses RREP transmission by using the proposed scheme, it should exactly find out whether it is congested or not. Because, although the adequate suppression of RREQ improves to load balancing, immoderate

drops of RREP can not make full use of advantages of route cache mechanism. Therefore, the scheme needs a dynamic detection mechanism to check congestion symptom. To realize this issue, we propose two threshold values. The first threshold is Queue-Threshold ($Q_{Threshold}$) as mentioned last section, which is compared with current Q_{Len} value of interface queue. For example, if $Q_{Threshold}$ is 30 and Q_{Len} is more than 30, the node believes that is it congestion environment now.

The second threshold value is Time-Threshold ($T_{Threshold}$) which is the packet residence time in the interface queue when Q_{Len} is higher than $Q_{Threshold}$. The $Q_{Threshold}$ information alone is not enough to evaluate the actual traffic load status of a node because the queue length can be high for a short period in a temporal situation even though the node is not overloaded due to the characteristics of burst packet transmissions or frame retransmission policies. In addition, due to the various data rate of each node, packet process capacity is also different from each other. When time the packet residence time is greater than $T_{Threshold}$, the algorithm considers that it is a real congestion situation. Thus, the node suppresses RREP packets to distribute traffic load.

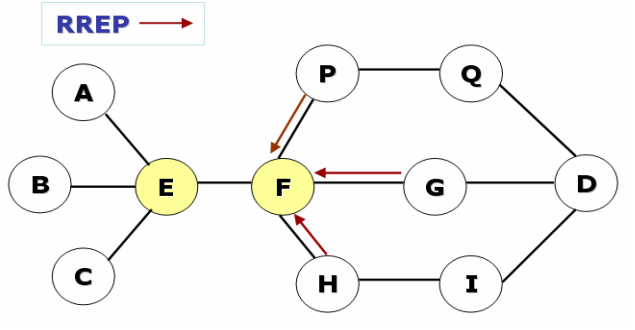


Fig. 2. Example of RREP storm

3.3 Prevention of RREP Storm

Although a source node can detour the congested intermediate node by suppressing RREPs in the route cache, if the intermediate node does not suppress RREP due to low traffic conditions, it still results in a RREP storm problem. RREP Storm usually occurs when there are many neighbor nodes which know the route to the destination node. This situation leads all neighbor nodes to transmit RREP simultaneously, which results in not only high packet collisions but also waste of channel bandwidth and battery life. Figure 2 illustrates an example of this RREP storm problem. If the node F is source or relays RREQ packets for other sources, it can receive three RREP packets simultaneously from node P, G and H which have fresh route information to destination. In this case, several packets collide in a MAC layer and packet retransmissions are inevitable, which consumes additional network resources.

In order to prevent the RREP storm problem, each node do not transmit RREP at the same time. It is a simple solution to postpone the RREP transmission intentionally. However, when the delay is too high, nodes suffer from long packet

delivery latency. On the contrary, when the delay is too short, it can not prevent the RREP storm problem. Although additional timer module for optimal delay value is another solution, memory overhead is very expensive to tiny mobile nodes. Therefore the proposed scheme uses contention window (CW) tuning approach in the MAC layer to reduce memory overhead as well as obtain the reasonable delay value. The expression (1) and (2) describes general CW configuration schemes in CSMA/CA for wireless channel access. Express (1) is set when nodes start to transmit a packet and confirmed successful transmissions. Express (2) is set when a node does not receive an Ack packet from its neighbor node.

$$CW = CW_{min}, \tag{1}$$

$$CW = Min(CW*2,\ CW_{max}) \tag{2}$$

$$Q_{avg(i)} = WQ_{avg(i-i)} + (1-W)Q_{len(i)} \tag{3}$$

$$Adaptive\_CW = Min(CW * (1+\frac{Q_{avg(i)}}{Q_{\max}}), CW_{\max}) \tag{4}$$

If the intermediate node wants to transmit RREP rather than suppression it, the node calculates Adaptive_CW which is a new delayed CW value by using expression (3) and (4). In expression (3), $Q_{avi(i)}$ denotes the average queue length in MAC layer when the node receives i-th packet and $Q_{avi(i-1)}$ is i-1th value which is obtained from previous calculation. The variable W means a weight value based on low-pass filter, which can be properly configured by user policy. Thus, the average is derived from EWMA (Exponential Weighted Moving Average), which can prevent sudden change on measurement and consequently calculate more accurate values.

Throughout expression (4), Adaptive_CW can be adaptively derived in proportion to $Q_{avi(i)}$. In other words, when the average traffic is increase, the additional delay also increases. However, this latency can not higher than CW_{max} because severe delay leads to not only packet retransmissions in MAC layer but also a route failure on upper layers in the worst case.

The reason why an additional delay on RREP is proportion to average queue length is that a node which is less congested node can transmit RREP faster than other nodes. Therefore the source node can establish less congested route. In addition, the bottlenecked node has enough time to process buffering packets and finally we can achieve noble traffic distribution in entire networks.

3.4 Management of Priority Packets

Although an appropriate suppression of RREP transmission is efficient for congestion prevention and load balanced routing, it may degrade the network performance if there is only one route between source and destination and the cache information is highly accurate. Moreover, when a source node needs to process QoS supporting packets which have highly delay sensitive requirements, it should receive immediate responses with RREP. In order to support these priority packets, the proposed scheme adds a priority flag in the option field of the RREQ packet. Thus, when an intermediate node receives RREQ with option flag set, it transmits RREP

immediately assuming that the route cache is accurate. In addition, the node does not apply the queue monitoring algorithm. If the intermediate node suffers from overflows by congestion, it can selectively drop packets according to QoS policy. Because the QoS supporting algorithm is beyond of this paper, we leave this issue for future works.

4 Performance Evaluation

4.1 Simulation Environment

We have studied the proposed scheme which is modified a version of AODV and DSR and compared to original ADOV and DSR routing protocols to validate the performance. For the simulation study, we have used the NS-2 simulator [9]. It was assumed that 50 mobile nodes are randomly located in a 1500m x 300m rectangle network area. For the mobility pattern, we used random waypoint model [10] in which all mobile nodes moved freely. And all nodes have a speed of 10m/sec with the pause time of 50 seconds during the simulation time of 300 seconds. We used IEEE 802.11 using RTS/CTS with 2Mbps for MAC protocol and the radio transmission range for each node was set to 250 meters. The number of data connections was 20 and each pair of source and destination nodes of a connection was randomly selected without duplicate sources. In our simulation, all source nodes generated constant bit rate (CBR) traffic. The reason why we used CBR is that we can easily make congestion environment without TCP congestion control by using UDP. In order to represent different network traffic load, we studied five different packet rates of 5, 10, 15, 20, and 25 packets/sec. The maximum queue size of each node's interface was set to 50 and three different $Q_{Threshold}$ values of 45, 20, 10 and three different $T_{Threshold}$ values of 2, 1, and 0.5 were used for the various simulation studies.

4.2 Simulation Result

Figure 3 shows the averaged number of dropped packets in nodes' interface by buffer overflows. In all circumstances of various packet rates, we can observe that the proposed scheme provides less buffer overflows because during the route discovery procedure it can detour congested intermediate nodes in the network and moderately avoid RREP storm while the other protocols have frequent packet drops by buffer overflows, which eventually leads to route breakdowns. However, when the traffic load is extremely high such as 25 packets/sec, our scheme also can not prevent overflows and suffers congestion. Meanwhile, the number of dropped packets of AODV is higher than that of DSR under low traffic load especially when packet rate is under around 13. However, when the traffic load is high with more than 15 packets/sec, AODV outperforms DSR. This phenomenon is because DSR highly depends on the route cache which generates lots of stale route information using RREP. Thus, DSR consumes more bandwidth and suffer more packets drops under heavy traffic condition.

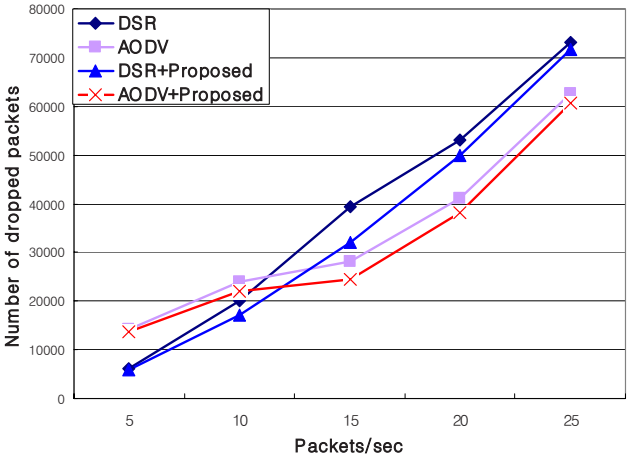


Fig. 3. Number of drooped packets in interface

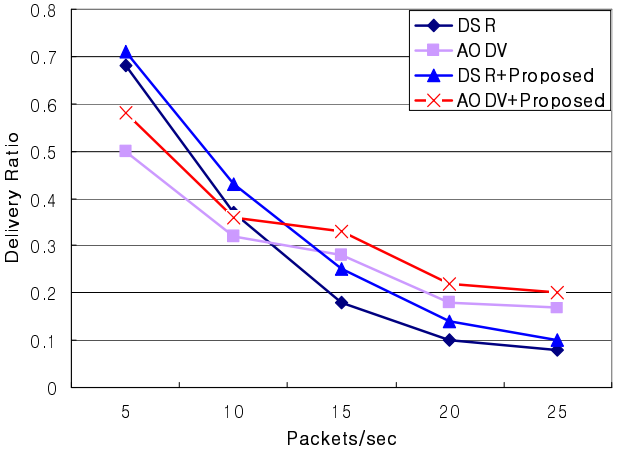


Fig. 4. Data packet delivery ratio

Figure 4 illustrates the packet delivery ratios of proposed scheme and standard protocols as a function of traffic load. The delivery ratio of our scheme is better than those of general DSR and AODV due to less frequent buffer overflows, which is correspondent to figure 3. However, when the packet rate is over 25, delivery ratios of all the protocols are saturated and we can not solve the problem because the entire network is congested.

Figure 5 shows the packet end-to-end delay as a function of traffic load. Usually when the network traffic load increases, the end-to-end delay also increases. However, the delay of proposed scheme does not increase rapidly because it can detour bottlenecked nodes and reduce unnecessary repeated route discovery

Table 1. Various threshold values of proposed scheme with DSR

Threshold		15 packets/sec		
$Q_{Threshold}$	$T_{Threshold}$	Delivery Ratio	End-to-End Delay	Overflow Dropped
45	2	0.24	1.56	38249
20	1	0.25	1.53	32071
10	0.5	0.24	1.55	33892

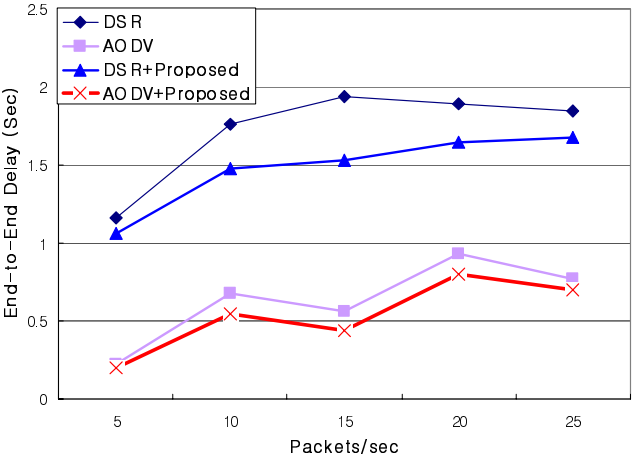


Fig. 5. Average end-to-end delay

procedures. Moreover, when packet rate is around 15, our scheme gets better delay performance with DSR because DSR suffers from more stale cache problem than others and it can not suppress staled RREP. Meanwhile, in general on demand protocols, the end-to-end delay decreases when the packet rate is above 15. When the traffic load is heavily high and most intermediate nodes are congested, RREQ packets are also dropped by buffer overflows, so the congested nodes can not forward RREQ packets as well as data packets to the destination. As a result, it can spontaneously detour the congested nodes during the route discovery procedure.

Finally, Table 1 shows the comparison of the performance with different queue threshold values of proposed scheme with DSR in order to confirm the actual congestion status. Although it is not easy to select the optimal values, we can observe that the queue threshold value affects the protocol's performance by setting differently. Consequently, the new scheme shows the best performance when $Q_{Threshold}$ is 25 and $T_{Threshold}$ is 1.0 sec. However, detailed descriptions with AODV are not given in this paper due to the lack of available space.

5 Conclusion

In mobile ad hoc networks, congestion on a certain node may lead to performance degradation such as extreme packet losses by buffer overflows and long end-to-end

delay. Furthermore, the problem grows more serious especially when most sources transmit their packets to a certain node by using route cache. However, existing load balancing protocols usually consider routing with an alternative route and do not restrain the route cache during congestion situation. In this paper, we have proposed a route cache based load balancing algorithm which can monitor the congested node in route and dynamically suppress RREP transmission during the route discovery session. By using proposed scheme, the source node can detour the bottlenecked nodes. And the bottlenecked node is able to have more time to process buffed packets. We also have defined two queue threshold values to confirm the actual congestion status. Simulation result shows that proposed scheme has a good performance in terms of packet delivery ratio, end-to-end delay, and the number of overflows when the network is heavily loaded.

References

1. Perkins, C.E., Bhagwat, P.: Highly Dynamic Destination Sequenced Distance-vector Routing (DSDV) for Mobile Computers. Comp. Commun. Rev., 234–244 (October 1994)
2. Johnson, D.B., Maltz, D.A., Hu, Y.-C.: The Dynamic Source Routing Protocol for Mobile Ad Hoc Networks (DSR). In: IETF Mobile Ad hoc Networks (MANET) Working Group (Internet Draft)
3. Perkins, C.E., Royer, E.: Ad-hoc on-demand Distance Vector Routing. In: Proc. 2nd IEEE Wksp. Mobile Comp. Sys. App. (1999)
4. Lee, S.-J., Gerla, M.: Dynamic Load-Aware Routing in Ad hoc networks. In: Proceedings of IEEE ICC (2001)
5. Kim, B.C., Lee, J.Y., Lee, H.S., Ma, J.S.: An Ad-hoc Routing Protocol with Minimum Contention Time and Load Balancing. IEEE Globecom (2003)
6. Tran, D.A., Raghavendra, H.: Routing with congestion awareness and adaptivity in mobile ad hoc networks. In: IEEE WCNC (2005)
7. Lee, Y.J., Riley, G.F.: A Workload-Based Adaptive Load-Balancing Technique for Mobile Ad Hoc Networks. IEEE WCNC (2005)
8. Kim, B.C., Lee, H.S., Ma, J.S.: Enhanced Ad Hoc On-demand Distance Vector(EAODV) Routing Protocol with Route Distribution. In: IEEE VTC (2005)
9. McCanne, S., Floyd, S.: NS network simulator, `http://www.isi.edu/nsnam/ns`
10. Bettstetter, C., Hartenstein, H., Perez-Cost, X.: Stochastic properties of the random waypoint mobility model. ACM/Kluwer Wireless Networks (September 2004)

Performance Evaluation of Load-Balancing Multi-path Routing Protocol for Mobile Ad-Hoc Networks

Zomahoun Jean-Eudes[1], Akio Koyama[1], Tomoyuki Tanno[2], Junpei Arai[3], and Leonard Barolli[4]

[1] Department of Informatics, Graduate School of Science and Engineering, Yamagata University, Japan
dsf67302@dipfr.dip.yz.yamagata-u.ac.jp,
akoyama@yz.yamagata-u.ac.jp
[2] Department of Applied Information Sciences, Graduate School of Information Sciences, Tohoku University, Japan
[3] Department of Information and Control Engineering, Yamagata College of Industry and Technology, Japan
j_arai@astro.yamagata-cit.ac.jp
[4] Department of Information and Communication Engineering, Fukuoka Institute of Technology, Japan
barolli@fit.ac.jp

Abstract. An ad hoc network is composed by autonomous, potentially mobile, wireless nodes that operate without the benefit of any existing infrastructure. Each node is can move arbitrary and carries out self-administration and self-organization. For this reason, there are cases in which non-planned route's construction may lead to a decline of the network performance. In this paper, we present a new routing protocol that uses the load balancing and multi-path solution to send data with a probabilistic dispersion. In this research, the total load on a route is evaluated based on the relay node's queue size. Simulation experiments have been made to compare some of the existing protocols with the proposed routing protocol. We have shown by simulations results that our proposed solution efficiently disperses the load and improves the network performance.

Keywords: Ad-hoc networks, Routing, Load balancing, Multi-path routing.

1 Introduction

In recent years, the popularizations of mobile terminals like mobile PC and PDA (Personal Digital Assistant), the miniaturization of mobile computing devices, and the rise of processing power available in mobile computers have resulted in a better computer-based applications into the hands of growing network population.

Nowadays, there are many research works that deal with wireless networks [1] and ad hoc networks. Ad-hoc network is a technology where the communication between terminals is carried out without using any infrastructure such as access points. This is possible by relaying the wireless communication between terminals and the

M. Takizawa, L. Barolli, and T. Enokido (Eds.): NBiS 2008, LNCS 5186, pp. 91–100, 2008.
© Springer-Verlag Berlin Heidelberg 2008

communication with each terminal. The topology of the network always changes, therefore an efficient and stable communication is needed. Until now Mobile Ad-hoc NETworking (MANET) Working Group [2] has proposed many drafts for routing protocols [3]. The fundamental route construction method is to search the route that reaches the destination with the shortest number of hops. However, in the case when all terminals use the shortest path to transmit data, in some terminals the load concentration may occur. This results in decline of performance and increase of the delay and the packet drop rate. For this reason, in ad hoc networks it is very important that the load in the whole network is efficiently dispersed.

In this paper, we propose a novel routing protocol in which the load on each route is evaluated based on terminal's queue length (the packet will wait temporarily in the queue) and then, the load balancing and multi-path will be used to efficiently disperse the load and make the packet transmission. We will also compare our proposed protocol with existing protocols like DSR (Dynamic Source Routing) [4] and LARA (Load Aware Routing in Ad hoc networks) [5].

The paper is organized as follows. In Section 2, we present the basic routing method for ad-hoc networks. In Section 3, we discuss the related work. In Section 4, we present the proposed protocol. In Section 5, we deal with performance evaluation. Finally, conclusions and future work are given in Section 6.

2 Basic Routing Method

For the search of routes in DSR that is the most basic routing protocol, two control packets named RREQ (Route REQuest) and RREP (Route REPly) are used.

At first, the sender node generates a RREQ that contains in its header the sender node number and broadcasts it to all nodes within its wireless range. Other nodes that receive RREQ similarly add their node number to the header of RREQ and broadcast it. If the same RREQ is received again, it is dropped. Thus, the packet overflow can be prevented. The intermediates nodes repeat broadcasting the RREQ until it reaches the destination node. The destination node sends back a RREP for the first arrived RREQ. The RREP inversely follows the route information saved into the RREQ to reach the source node. By the above-mentioned operation, finally the source node knows which nodes to use as intermediate node to be able to reach the destination node and the communication becomes possible.

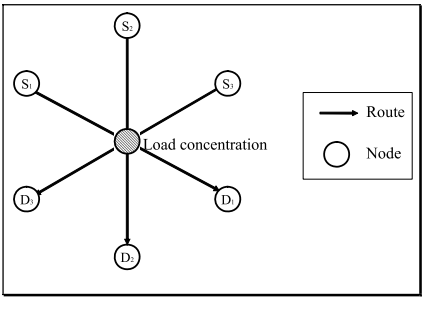


Fig. 1. Load concentration caused by shortest path

In this method only the shortest route candidate is selected. Because of this, as shown in Fig. 1 a specific node is used in several routes. So there is load concentration on that specific node and this may lead to a decrease of the network performance.

3 Related Work

In this section, we introduce LARA and MultiPath Routing protocols.

3.1 LARA

LARA uses the sum of the queue size (packet that are waiting for transmission) of relayed nodes to evaluate the load on each route and the route with the least load is chosen. So LARA is a load balanced routing protocol.

To be able to use a regular size of the queue for the load calculation, the node samples the size of the queue at constant intervals. Let us suppose the average size of node i is q_i, the sample number is N, and the k-th's sample is $q_i(k)$, then q_i can be calculated by Eq. (1).

$$q_i = \frac{\sum_{k=1}^{N} q_i(k)}{N} \tag{1}$$

Since in a wireless communication, it is not possible to communicate when neighbor nodes are communicating with each other, the queue size of neighbor nodes are taken into consideration. So the load Q(i) of a node i is defined by Eq.(2).

$$Q(i) = \sum_{\forall j \in N(i)} q_j \tag{2}$$

N(i) refers to nodes that are at one hop from a considered node. At the route search, the load of the intermediate node is saved in the RREQ while trying to find the destination node. For this reason, when the RREQ reaches the destination node the load evaluation value C(r) of the route is calculated by Eq. (3).

$$C(r) = \sum_{i \in r} Q(i) \tag{3}$$

In LARA, to be able to gather the route information for many candidates, after the first RREP reaches the destination node there are 50 ms of waiting time before the RREP is sent. As shown in Fig. 2, the route with the least load evaluation value is selected from the candidate routes. In this way, LARA realizes a load balancing. However, the evaluation of the load in LARA is a simple sum of the queue size. For this reason it is impossible for LARA to detect a high loaded node. So there are cases where the load balancing is not accurately done. Although much route guidance information is acquired, finally only one route is used for the transmission. This shows that the effect of the load-balancing is low.

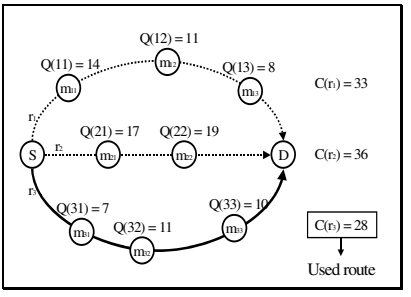


Fig. 2. Route selection in LARA

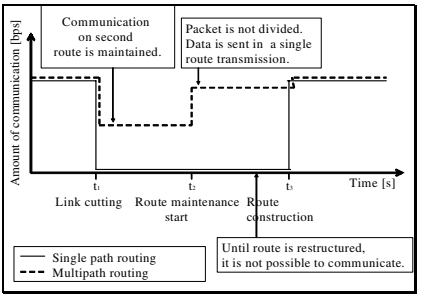


Fig. 3. Avoidance of high drop throughput

3.2 MultiPath Routing

During recent years, there are proposed many multipath routing protocols such as AOMDV (Ad hoc On-demand Multipath Distance Vector) [6] and SMR (Split Multipath Routing) [7]. These protocols consist of alternatively sending packets through two routes. So, the load-balancing is possible and if the link cutting occurs on one of the routes, the communication can be continued by the remaining route. For this reason, the multipath routing protocols have the property of avoiding high drop of throughput.

In Fig. 3 is shown the change of amount of communication in the case of single path and multipath when the link cutting happens. Amount of communication here means the amount of data sent by unit of time from source node to a destination node. At the time t_1, when link cutting occurs in the case of single path routing, the data can not reach the destination anymore. So the amount of data comes down to zero. But in the case of multipath routing, there are two routes available and the amount of transmission of the second route is kept. Besides, in multipath routing at t_2 the source node is aware of the link cutting and the route maintenance starts. Then before the cut link is fixed at t_3, the data is not divided anymore but sent on one route. By doing so, the amount of communication can be restored to a certain level.

However, the conventional multipath routing protocols just divide the transmission data into two routes and send it. So the load condition on the routes is not considered.

4 Proposed Protocol

In LARA the load information of several routes is gathered. However, only one route is consequentially used. For this reason, in the case a large amount of data transmission the load-balancing is not efficient. To solve this problem, we propose a network performance maintained routing protocol in which the route information collected at the route discovery is used without any waste and the load is efficiently balanced around the multiple available routes. The maintenance of the network performance is achieved by concretely using the following techniques:

- evaluation of route load using size of queue,
- detection of extremely loaded nodes,

- optimization of the ratio of the packet transmission based on the evaluation value,
- prevention of high drop of throughput when route is cut by constructing several routes.

4.1 Calculation of Route Load Evaluation Value

The terminal that receives RREQ while the route search is operating calculates its own load evaluation value which is added to the route load evaluation value of the header of the RREQ packet. Let us suppose the load evaluation value of a node i is L_i, the queue size sample is N, the k-th queue size sample is $q_i(k)$, the number of hops from the source node is h, the maximum number of hops is h_M. We assume that h_M is 7. Then, the q_i and L_i can be calculated by Eq. (4) and Eq. (5), respectively.

$$q_i = \frac{\sum_{k=1}^{N} q_i(k)}{N} \tag{4}$$

$$L_i = \left(0.5 + \frac{h-1}{h_M - 2}\right) \times q_i^2 \tag{5}$$

Eq. (4) shows the average of the queue size. This is the same idea with LARA, the queue size varies widely with the time. So, the purpose is to acquire an accurate load as much as possible. In Eq. (5), the L_i is in proportional with the square of the queue size. This is very important to detect a node with a very high load as shown in Fig. 4. Also, based on the number of hops for a node, some weights are considered to give a high priority to the routes with a small number of hops. Finally, when the RREQ reaches the destination node it contains the load value of the whole routes. L(r), the load evaluation value of a route r, is expressed by using L_i (the load evaluation value of node i) as shown by Eq. (6).

$$L(r) = \sum_{i \in r} L_i \tag{6}$$

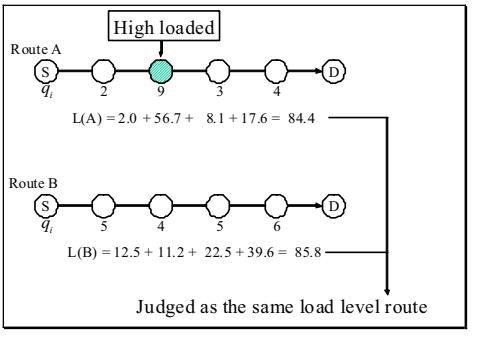


Fig. 4. Detection of a high loaded node

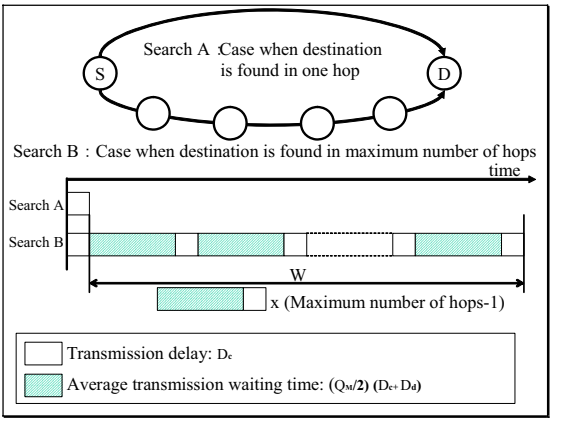


Fig. 5. Maximum difference of RREQ's average delay

4.2 Optimization of RREQ's Waiting Time

After the first RREQ reaches the destination node, the destination node waits for a certain time to gather the following RREQ's information. This is done for gathering the information of many routes. The waiting time W is calculated by Eq. (7):

$$W=(h_M-1)\times\left\{D_c+\frac{Q_M}{2}\times(D_c+D_d)\right\} \tag{7}$$

where h_M is maximum number of hops, D_c is transmission delay of the control packet, Q_M is maximum value of node queue, D_d is transmission delay of the data packet. The search A and search B in Fig. 5 show the search time when the destination node is found in one hop or the maximum number of hops, respectively. The difference is the multiplication of the average transmission waiting time and transmission delay by the (Maximum number of hops - 1). The average transmission waiting time is when the number of control packet and the number of data packet in a node queue are the same for the maximum queue size. Eq. (7) is derived considering the needed time to send all the packets waiting in the queue ($\frac{Q_M}{2}\times(D_c+D_d)$) and the transmission delay D_c (transmission time of RREQ). When the waiting time passes, the destination node sequentially selects from RREQ packets up to four RREQ packets with the smaller load. Then, it selects one RREP packet to reply to the source node by using the smallest load route.

4.3 Calculation of Transmission Ratio

When the RREP reaches the source node, it calculates the ratio of the number of packets to send to each route based on the multiple routes information and the load evaluation values of the routes.

Table 1. Examples of calculation of transmission ratio

j	Example 1		Example 2		Example 3	
	$L(r_j)$	$P(r_j)$	$L(r_j)$	$P(r_j)$	$L(r_j)$	$P(r_j)$
1	9.0	0.4	81.0	0.348	25.872	0.602
2	10.0	0.36	96.5	0.292	68.184	0.228
3	15.0	0.24	142.3	0.198	91.518	0.17
4	none	0	174.1	0.162	none	0

Let us suppose that R is the set of gathered routes information. The probability $P(r_j)$ of packet to be sent by a route r_j is derived by Eq. (8). However, if there is a route with a number of hops equal to 1, this route is unconditionally selected and $P(r_j) = 1$.

$$P(r_j) = \frac{1}{L(r_j)\sum_{k \in R} \frac{1}{L(r_k)}} \tag{8}$$

Table 1 shows examples of calculation of transmission ratio using Eq. (8). We use Example 3 of Table 1 to show how the routes are constructed (see Fig. 6).

At the source node, the route's load evaluated value L(r), the transmission ratio P(r) and the route information obtained from the previously explained route research method are managed for each route. The load evaluation value is maintained in the case of link failure. In this case, it is possible to recalculate the load evaluation value. In order to explain this case, we will use example 1 of Table 1. Let us suppose that the link cutting occurs on the route j=2. Then, it will remain only two routes j=1 and j=3. The transmission ratio will be recalculated based on the route load information L(r) of the remaining routes using Eq. (8). As a result, we have $P(r_1) = 0.625$ and $P(r_3) = 0.375$. Because the route j=2 where link cutting occurred will not be used $P(r_2) = 0$.

In the case the network topology changes and all links fail, the source node initiates the route search and new routes are built again.

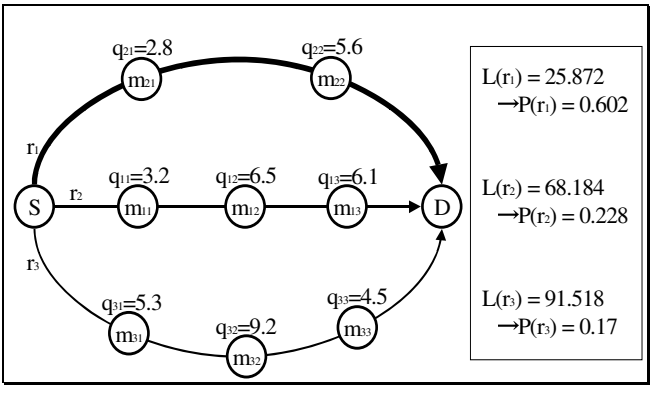


Fig. 6. Load and transmission rate

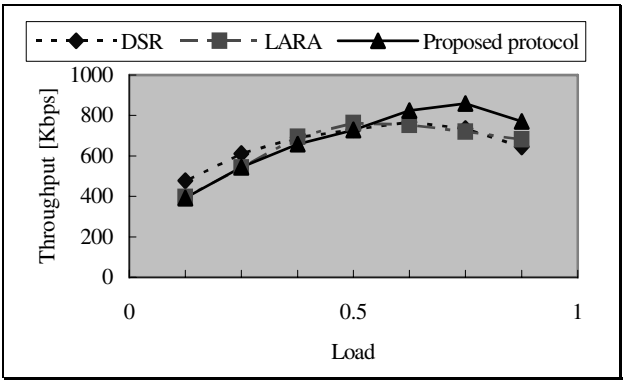


Fig. 7. Throughput results

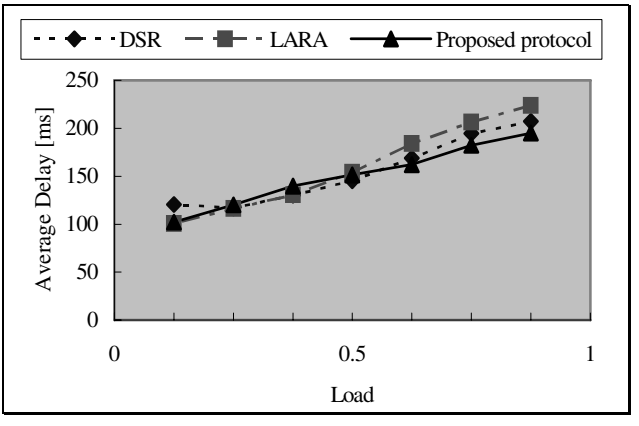


Fig. 8. Average delay

5 Performance Evaluation

To confirm the effectiveness of proposed protocol, we compare its performance with DSR and LARA using computer simulations.

5.1 Simulation Envirnment

To evaluate the performance of proposed and previous protocols, we consider as metrics the throughput, average delay and communication success rate. We change the packet generation interval and investigate how the protocols perform using the measurement metrics.

As mobility model, we used the Random Direction model [8]. We consider a field 500×500 [m^2], the movement speed is from 1 to 2 [m/s] and the movement direction and speed change at a maximum random interval is from 1 to 5 [s]. We assume that the wireless communication is IEEE802.11, the bandwidth is 11 [Mbps], and the wireless range is 100 [m].

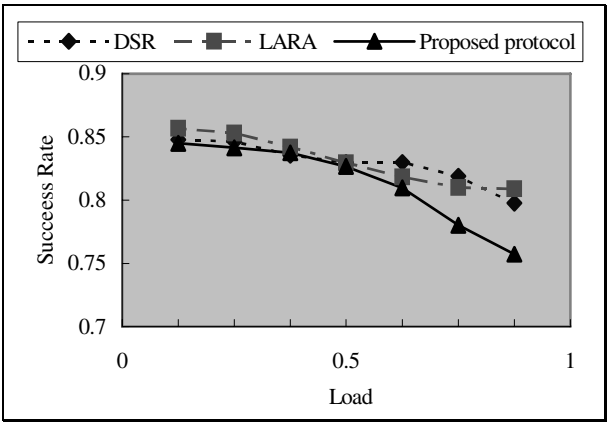


Fig. 9. Communication success rate

The simulation time is 30 [s]. But 3 seconds at the start and end of simulation were not used for measurement of the network performance. So, the simulation time is 24 [s].

We considered that the maximum number of hops is 7, the number of queue size sampling is 10, the queue size sampling interval is 0.1 [s], the control packets size is 64 [byte], and the maximum size of the data packet is 518 [byte].

5.2 Simulation Results and Considerations

In Fig.7, Fig.8 and Fig.9, we show the throughput, average delay and communication success rate obtained from the simulations. The horizontal axis shows the normalized load.

DSR shows a high throughput when the load is low. However, when the load becomes high, the throughput is decreased. This is because when the load is high, collisions occur frequently at neighborhood nodes. So, the packet can not be efficiently forwarded. While in LARA and the proposed protocol, the heavily loaded nodes are avoided for routing, so even in the case of high load a good throughput can be maintained. However, a big difference between LARA and the proposed protocol appears under high load condition. In this case, if high loaded nodes number increases, the load balancing in LARA is not efficient (the algorithm simply consists in adding the average of the queue size). On the other hand, the proposed protocol detects the high loaded nodes and the load is distributed efficiently. So, a good throughput is maintained.

The result of the average delay shows that under low load condition, DSR have the lowest delay and in high load condition the proposed protocol has the lowest delay. LARA and the proposed protocol wait a certain time for gathering RREQ. So there is a disadvantage comparing them with DSR, which does not wait for other RREQs.

Under high load condition, when the frequency of collision is very high and the waiting time at intermediate node increases, the proposed protocol by using avoidance of packet concentration has the lowest delay. The difference between LARA and the proposed protocol shows that the proposed protocol optimizes the RREQ waiting

time. The communication success rate graph shows that when the load becomes high, the success rate of all protocols decreases. This is because the collision can happen easily when the load becomes high. Comparing the proposed protocol with the previous protocols, we can say that the proposed protocol builds several routes, but high loaded routes are influenced by the load balancing parameter. For this reason, the success rate is decreased. However, the proposed protocol is a multipath routing protocol and as we can see from Fig. 8 it has a good throughput.

6 Conclusions

In this paper, we proposed a routing protocol for ad hoc network that efficiently balance the load even under high load conditions. This is done by evaluating the load of each node based on the queue size, balancing and sending the packets on several routes based on the load evaluated value. We also show that the network performance is maintained by balancing the load among different routes. We compared the performance of the proposed protocol with two previous protocols and showed that the proposed protocol has a better performance than LARA and DSR protocols.

As future work, we will use the SMR technique to improve the RREQ operation for searching more candidate routes. Furthermore, a threshold value can be established for selecting the routes. So, the useless routes and high loaded routes can be eliminated. By using these strategies, we can improve the network performance.

References

1. IEEE802.11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications. First Edition (1999)
2. Mobile Ad-hoc Networks (MANET), `http://www.ietf.org/html.charters/manet-charter.html`
3. Toh, C.-K.: Ad-hoc Mobile Wireless Networks: Protocol and System. Pearson Education, London (2002)
4. IETF MANET Internet Draft: The Dynamic Source Routing Protocol for Mobile Ad hoc Networks (2002)
5. Saigal, V., Nayak, A.K., Pradhan, S.K., Mall, R.: Load Balanced Routing in Mobile Ad Hoc Networks. Computer Communications 27(3), 295–305 (2004)
6. Marina, M.K., Das, S.R.: On-demand Multipath Distance Vector Routing in Ad Hoc Networks. In: Proc. of International Conference on Network Protocols (ICNP), pp. 14–23 (2001)
7. Lee, S.J., Gerla, M.: Split Multipath Routing with Maximally Disjoint Paths in Ad hoc Networks. In: Proc. of IEEE International Conference on Communications (ICC), pp. 3201–3205 (2001)
8. Gloss, B., Scharf, M., Neubauer, D.: A More Realistic Random Direction Mobility Model. In: 4th Management Committee Meeting, Germany (2005)

A TCP Enhancement for QoS-Aware Mobile Ad-Hoc Networks

C. Mbarushimana and A. Shahrabi

School of Engineering and Computing
Glasgow Caledonian University
Glasgow G4 0BA, U.K.
{Consolee.Mbarushimana,A.Shahrabi}@gcal.ac.uk

Abstract. Successful deployment of Mobile Ad-hoc Networks (MANETs) is highly dependent on how their scarce resources are used by upper layer protocols. TCP plays a significant role in determining the workload of the network and its variants achieve different levels of performance as reported in literature. With the increase in delay-sensitive applications in today's networks, TCP experiences more spurious timeouts due to blocking by high priority traffic. The resulting retransmissions exacerbate the existing congestion problem in MANETs and misuse the available bandwidth. Our proposed TCP protocol (RE-TCP) achieves resource efficiency by adjusting the retransmission timer according to the medium contention, thus limiting the number of contention-induced retransmissions. In this paper, we first evaluate the impact of different loss recovery mechanisms on TCP performance in presence of voice traffic. We then analyze the resource efficiency of different TCP variants under contention-induced spurious timeouts. Our simulation study reveals a considerable improvement of TCP Reno by our proposed enhancement.

Keywords: MANET, QoS, Spurious, TCP, Variant.

1 Introduction

Although multimedia traffic has increased tremendously over the past few years, today's Internet traffic is still dominated by TCP based applications. However, TCP has been found not to perform well in MANETs as extensively reported in the literature. This is mainly attributed to its inability to effectively respond to frequent topology changes; several solutions were proposed to address this problem [1], [4]. Network congestion, a common occurrence in today's networks, results into large delays, packet loss and blocking of new connections. TCP, which implements flow control to regulate the network traffic, plays a significant role in determining the workload of the network. However, the resource efficiency of TCP is restricted by its excessive access of the medium, not only caused by TCP ACKs, but also by the retransmissions performed by TCP when reacting to timeouts.

TCP timeouts are due to large delays exceeding the computed RTO (Retransmission TimeOut). This delay variability can be attributed to the time-varying quality of the

M. Takizawa, L. Barolli, and T. Enokido (Eds.): NBiS 2008, LNCS 5186, pp. 101–110, 2008.
© Springer-Verlag Berlin Heidelberg 2008

wireless link and frequent route failures, and also to the fierce medium contention due to priority scheduling and pre-emptive services at the MAC layer. The latter reason is becoming increasingly important as future wireless networks become more and more multimedia-oriented, requiring efficient scheduling algorithms to maintain quality of service guarantees. This results into more delay for TCP data and TCP ACKs packets, consequently incurring more sender retransmission timeouts. In addition, the congestion avoidance mechanism is triggered and falsely reduces the TCP window size leading to low link utilization and throughput. An analytical model of TCP Reno sending rate and throughput in presence of spurious timeouts in wireless networks was developed in [2] and several efforts have been put into making TCP robust against spurious timeouts [6], [10].

Although several studies have been carried out to determine which of TCP variants is suited best for MANETs [5], [9], [12], their bandwidth efficiency has not been considered in any of the studies. Furthermore, the behaviour of TCP variants when competing with UDP based delay-sensitive applications in QoS-aware MANETs has not been investigated. In this paper, we evaluate the performance of TCP Reno and NewReno in presence of spurious timeouts due to blocking by high priority traffic. We also measure the improvement achievable by our proposed enhancement to limit spurious timeouts in QoS-aware MANETs (RE-TCP). The idea behind RE-TCP, a cross-layer solution which works conjointly with the MAC layer, is to freeze the retransmission timer when the medium is busy with high priority traffic, and to unfreeze it when it is available for TCP transmission.

The rest of the paper is organized as follows; Section 2 gives a brief review about related work. Section 3 describes the problem of spurious timeouts, and it also gives a brief description of QoS-aware MANETs, RE-TCP and the TCP variants used in this study. The performance evaluation of TCP and RE-TCP variants is carried out in Section 4. Finally, some concluding remarks are given in Section 5.

2 Related Work

The search for a suitable variant for MANETs has been carried out over the past few years. In the study conducted in [12], TCP Vegas is able to perform better than Reno over AODV, but the opposite is observed in DSR networks. Similarly, Kim et al. analyzed the interaction between TCP variants and routing protocols in [5], and their results show that the difference in the performance between TCP variants is not as significant as the difference in routing protocols. It was again shown that TCP Vegas works better with AODV, whereas Reno outperforms Vegas in OLSR networks. Li et al. conducted an analytical study of TCP Reno and NewReno behaviour in IEEE 802.11 based ad-hoc networks with multiple losses in [9]. It was shown that NewReno outperforms Reno in networks with heavy packet loss in terms of throughput and timeout probability.

Some approaches have been proposed to improve TCP resource efficiency in MANETs. Those include the approach to limit the contention-induced losses proposed by Hamadani and Rakocevic in [3]. They proposed a cross-layer solution to dynamically adjust the amount of outstanding data in the network based on the contention experienced by packets as well as the throughput achieved by connections.

Another approach to limit the number of spurious timeouts was proposed by Klein et al. in [7]. This approach proposed to randomly inject additional delay along the communication path in wireless networks. This method however does not take into consideration the status of the medium or the network load. There is therefore a possibility of adding additional delay when it is not needed, which is not desirable when the packet has actually been dropped, a quick timeout is desirable in such a case. The study reported in [13] on the other hand proposed a method to distinguish congestion and route failures losses, and to dynamically adjust the RTO value based on the network conditions, which is indicated by the throughput.

Although some TCP variants are more suitable than others for some particular scenarios in MANETs, they all still perform poorly due to the different characteristics of MANETs, and more improvement is still needed. Several solutions have been proposed to address some of the issues, but the effects of the presence of high priority traffic have not received enough attention. In this paper, we first evaluate how the TCP variants react to spurious timeouts due to blocking by high priority traffic, and we then analyze the improvement achievable by RE-TCP for the different variants.

3 Background

3.1 TCP Variants

All current TCP implementations are based on TCP Tahoe that defines algorithms for slow-start, congestion avoidance and fast retransmit. TCP Reno introduces a new phase, Fast Recovery. Following a fast retransmit, TCP Reno enters fast recovery by setting the threshold and the contention window to half the current size. This will result into a higher sending rate after the loss is recovered. Reno performs very well when the packet losses are small. This is because after retransmitting the first lost segment, Reno typically waits for the transmission timer to expire before retransmitting other segments suspected to be lost.

TCP NewReno is able to detect multiple packet losses and thus is much more efficient that Reno in the event of multiple packet losses. Like Reno, NewReno also enters into fast retransmit when it receives multiple duplicate packets, however if a partial ACK is received, NewReno does not exit fast recovery until all the outstanding data is acknowledged, therefore overcoming the problem faced by Reno of reducing the contention window multiples times.

3.2 QoS-Aware MANETs

There has been a tremendous increase in multimedia applications over the past few years. This type of applications requires QoS guarantees in terms of delay, bandwidth, packet loss and jitter. With the prospects of future MANETs commercial applications, it is desirable to support these services in MANETs as well.

The IEEE 802.11e EDCA provides a priority scheme to differentiate different access categories (ACs) by classifying the arbitration interframe space (AIFS), and the initial (CW_{min}) and maximum (CW_{max}) contention window sizes in the backoff procedures. In addition, each priority level is assigned a Transmit Opportunity

(TXOP). In real life, multimedia traffic like voice and video are assigned higher priority over best effort TCP based applications (e-mail, FTP).

3.3 Delay Spikes and Spurious Timeouts

A delay spike is defined as a sudden and sharp delay increase exceeding by several times the typical RTT of the TCP connection. Delay spikes are not uncommon in MANETs and they can be due to link layer error recovery, route failure, wireless bandwidth fluctuation, and blocking by high priority traffic. Since the data packets are delayed and not lost, the resulting retransmission is unnecessary and the timeout is spurious. Following a timeout, TCP will exponentially increase the RTO value. A study was conducted in [8] to analyze the performance of TCP in wireless networks with delay spikes. The current versions of TCP compute RTO based on the RTT values only. Small values of RTO result into spurious timeout if delay spikes occur, on the other hand, large values make TCP too slow in detecting a proper packet loss and stay idle too long instead of retransmitting the lost packet. It is therefore important to find a RTO value that balances the throughput degradation between both of these cases.

3.4 RE-TCP

In IEEE 802.11e, a node with high priority traffic will start decreasing its backoff earlier than that with low priority traffic. In addition, the backoff counter of high priority AC may count to zero in this interval and transmit the packet, which results in a busy channel due to high priority packet transmission. A station with best effort TCP traffic (both data and ACK packets) will not be allowed to compete for the medium, let alone transmit, as long as there is high priority traffic to be transmitted by one of the neighbouring nodes. For a TCP source node waiting for an ACK, this long sudden delay will results into a retransmission timeout. The node will schedule a retransmission and reduce its congestion window. This retransmission is categorized as spurious, as the packet was neither lost nor damaged.

In our previous work [11], we proposed an approach to reduce spurious timeouts due to blocking by high priority traffic. Our idea, called Resource Efficient TCP (RE-TCP), is based on freezing the retransmission timer when the medium is busy, and resuming the countdown when the medium becomes idle again. This is achieved by defining two interrupts at the MAC layer, which are triggered by a busy and an idle medium. The two interrupts are fed to the transport layer TCP. Using this method, the TCP layer is made aware of the contention status of the wireless channel. On receiving a medium busy interrupt, every active TCP connection enters the retransmission freeze state. TCP will come out this state when it receives the medium idle interrupt. This method implicitly injects an additional delay in the computation of the RTO, but unlike the method used in [7], the delay added reflects the medium contention status. This delay is continuously computed and added every time the medium is busy until the ACK is received or till the maximum RTO is reached. The retransmission timer is then reset.

4 Performance Study

4.1 Simulation Setup

Our simulations were conducted using OPNET Modeller 11.5. We simulated a network consisting of 50 mobile nodes moving in a 1000x1000 m area with a speed of 10m/s and 10 seconds pause time following the random waypoint mobility pattern, and a nominal transmission range of 250m. The MAC layer protocol used is EDCA, the access categories (ACs) are assigned priorities based on the default parameters for an IEEE802.11e physical layer. TCP packet size was fixed at 1460 bytes. The network traffic consisted of long lived FTP file transfers and voice traffic was simulated by establishing G711 CBR connections. Our investigation being on spurious retransmissions due to blocking by high priority traffic, we simulated various scenarios by varying the voice load to emulate different contention conditions.

4.2 Simulation Results

Today's networks are seeing an increase in multimedia traffic. This type of traffic tends to last longer than best effort traffic. While it might take few seconds to download a webpage or FTP file, a voice call can last several minutes. We therefore evaluate the effect of voice connection time on the performance of TCP. This is varied as a percentage of the total simulation time. We also evaluate the effect of the number of voice traffic sources on TCP. The performance is assessed by the number of TCP retransmissions, TCP goodput, bandwidth utilization, TCP segment delay and TCP traffic dropped.

4.2.1 TCP Retransmissions

The TCP retransmissions analyzed in this study include those due to packet loss, packet error and spurious timeouts. For longer lasting voice connections, the network experiences highly loaded conditions for long periods of time. This results into more collisions, hence the increased number of TCP retransmissions as shown on Figure 1(a). Furthermore, TCP timeouts are frequent due to long delays experienced by TCP segments and lack of transmission opportunities for ACK packets. Similarly, the increase in the number of voice conversations results into increased TCP retransmissions as seen on Figure 1(b). Not only does the load increase, but also it is distributed over a large part of the network as the nodes are scattered around the network. Regarding the retransmissions performed by ordinary TCP variants, NewReno is able to retransmit fewer segments than Reno. NewReno congestion avoidance mechanisms are able to detect developing congestion and are therefore more efficient and utilize network resources more efficiently. Due to its modified congestion avoidance and slow start, NewReno incurs fewer retransmits.

It is observed that RE-TCP can indeed reduce the number of TCP timeouts as the average number of retransmissions in the whole network is considerably reduced. The implicit delay equal to the time when the medium is busy with high priority traffic and therefore TCP transmissions opportunities are limited, is continuously injected during the lifetime of a connection in RE-TCP; this results into fewer timeouts and hence fewer retransmissions. We observe a reduction of 20% compared to ordinary TCP in some cases.

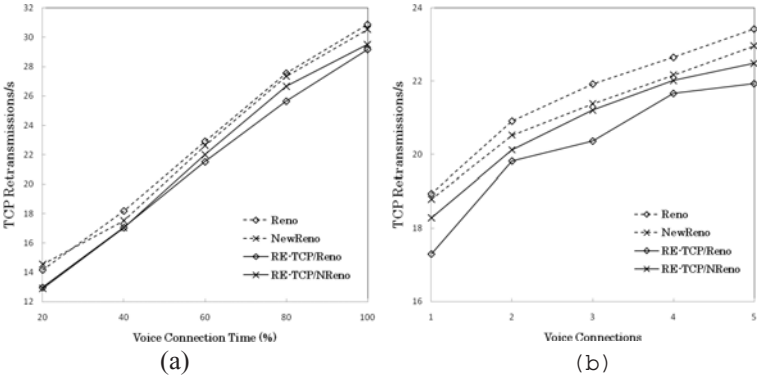


Fig. 1. Effect of (a) Voice Connection Time, (b) Voice Connections on TCP Retransmissions

For longer lasting voice connections, RE-TCP Reno retransmits fewer segments than NewReno. This is attributed to the fact that, during fast retransmit, Reno only retransmits the first segment, and relies on the retransmission timer to expire before it can retransmit any other segments. With RE-TCP increased RTO, the chance of receiving the ACK before the retransmission timer expires is increased; therefore RE-TCP Reno retransmits fewer segments. On the other hand, due to the randomness of the time when the ACK arrive at the source, NewReno which does not rely on the retransmission timer to expire before retransmitting might perform unnecessary retransmissions, when the ACK or data segment was just delayed. In the scenarios where the number of voice connections is increasing, long lasting voice connections were considered, hence the remarkable difference in the number of retransmissions carried out by the different variants.

4.2.2 TCP Goodput

The main purpose of the study is bandwidth optimization, but network users are usually concerned with their achievable goodput. The goodput represents data packets successfully received at the application layer. It therefore excludes the retransmissions.

With the increase in voice traffic, the goodput achievable by the two variants is significantly reduced as shown on Figures 2(a) and 2(b). It can be observed that TCP goodput is almost halved when the voice traffic is constantly present. This is because more transmission opportunities are given to voice traffic while TCP traffic starves. The longer the voice connections last, the poor the TCP goodput becomes. Similarly, the increase in number of VoIP nodes negatively affects TCP performance. With a large number of voice conversations, the probability of TCP sources to win the medium access is reduced.

We explained earlier that delay spikes reduce the overall TCP goodput; due to the fact the spurious timeouts not only induce unnecessary retransmissions but also make TCP enter the slow start phase reducing its congestion window which results into low goodput. Due to the decrease in timeouts as seen above, RE-TCP is constantly able to achieve higher goodput than TCP. An average of 10% gain in goodput is observed; the difference is significant in high contention scenarios. For the same reasons stated

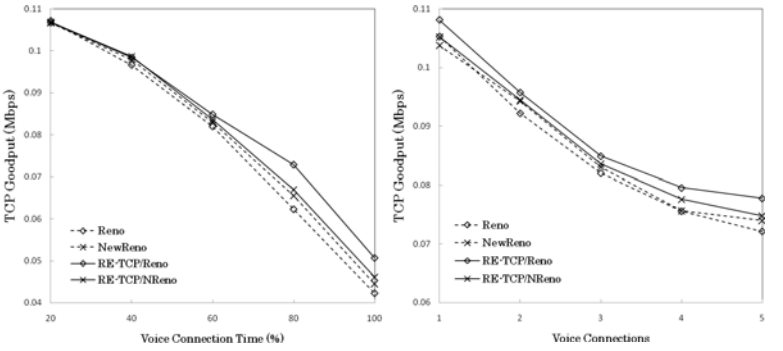


Fig. 2. Effect of (a) Voice Connection Time, (b) Voice Connections on TCP Goodput

above, NewReno is able to achieve better goodput than Reno in ordinary TCP, but the implementation of RE-TCP over Reno outperforms all the other variants. The reduced number of timeouts is favourable to RE-TCP Reno goodput as its congestion window is not often reduced as in the other variants.

4.2.3 Bandwidth Utilization

The network resource utilization is estimated based on the radio transmitter utilization metric. This statistic represents a measure of the consumption to date of an available channel bandwidth, where a value of 1 would indicate full usage. The radio transmitter utilization is a metric that is proportional to the traffic a node injected on the channel. As in the previous case, it is estimated in various load scenarios, increasing both the voice connection time and the number of voice connections as shown on Figures 3(a) and 3(b). As we are interested in TCP resource efficiency in this paper, the graphs depict the average of the channel utilization of the TCP traffic sources.

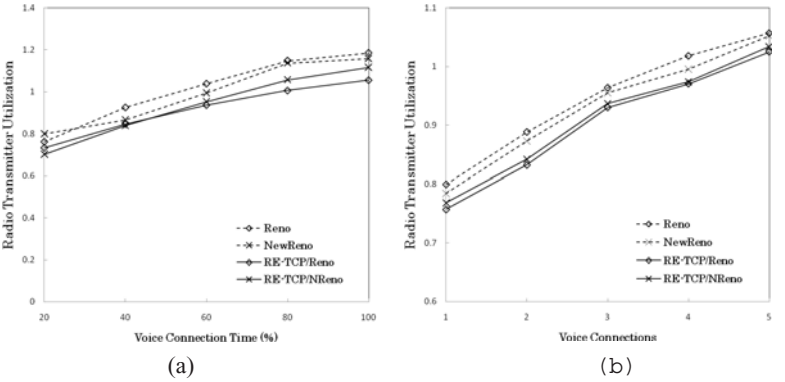


Fig. 3. Effect of (a) Voice Connection Time, (b) Voice Connections on Bandwidth Utilization

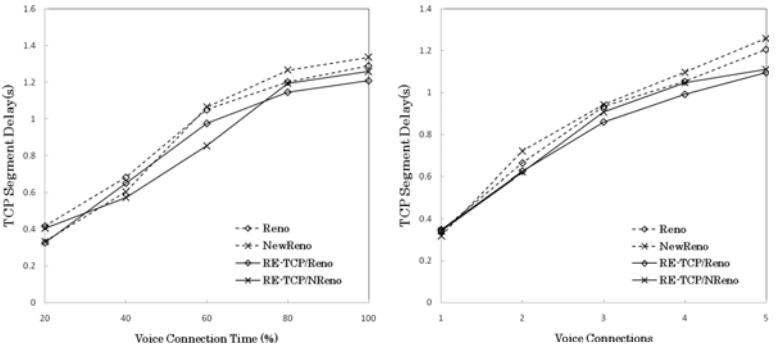


Fig. 4. Effect of (a) Voice Connection Time, (b) Voice Connections on TCP Segment Delay

TCP traffic placed on the channel includes the original TCP segments to be transmitted as well as the segments that need retransmission following a collision or a timeout. RE-TCP shows lower values of channel utilization, due to the fact that it retransmits fewer segments than TCP. However, as its goodput is higher, the difference is not large (about 10% average in these scenarios). Similarly, NewReno utilizes the bandwidth more efficiently than Reno, but the algorithm introduced in RE-TCP is more favourable on Reno than NewReno, hence the lower bandwidth utilization of RE-TCP Reno. It can be observed that this statistic exceeds 1 in some cases, which is an indication of an overloaded network.

4.2.4 TCP Delay

Long sudden delays in data transfers are not typical in wired networks and their effect on the TCP protocol has not been extensively studied. Such delays are not uncommon in WLANs due to link layer level retransmissions and temporal resource pre-emption by high-priority packet traffic such as voice calls. Transmission priority is given to delay-sensitive traffic, while TCP traffic faces long delays. As seen on Figure 4, the segment delay increases with increase in voice load.

Looking at the differences between the protocols, RE-TCP manages to deliver data segments with low delays than TCP, which can be attributed to its efficient resource utilization and congestion avoidance.

4.2.5 TCP Traffic Dropped

Data is sent from the application layer down to the medium access layer to be transmitted. In 802.11e, a queue is held for each access category. The rate at which packets arrive at the MAC layer might exceed the rate at which they get transmitted. This is not uncommon in wireless networks due to the fierce way in which the stations contend for the medium. This would results into overflow of the buffer used to store the ACs data awaiting transmission in which case some of it gets dropped by the MAC layer itself. Moreover, in wireless networks, when the MAC ACK is not received, the source station retransmits the same frame again and again till the MAC ACK is received or it exceeds the limit of transmissions attempts allowed per frame,

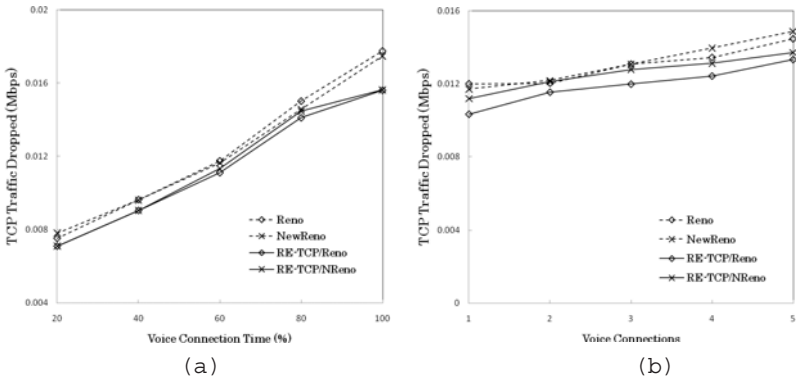


Fig. 5. Effect of (a) Voice Connection Time, (b) Voice Connections on TCP Traffic Dropped

in which case the frame gets dropped and the station moves to the next frame in the queue.

Figure 5 shows TCP data dropped by the MAC layer for the two protocols. Our simulations results show that RE-TCP presents constantly the lowest amount of total traffic dropped compared to TCP, while NewReno drops a larger number of packets than Reno.

5 Conclusion

TCP has been found to perform poorly in the presence of spurious timeouts caused by delay spikes which are more frequent in QoS-aware MANETs due to blocking by high priority traffic. In this paper, we investigated how TCP variants perform in presence of delay-sensitive voice traffic, and how they react to contention-induced spurious timeouts. We also analyzed the improvement achievable by our proposed variant RE-TCP, which is a cross-layer solution that adjusts the retransmission timer based on the medium contention.

TCP NewReno, which does not normally wait for the retransmission timer to expire before retransmitting segments, is more susceptible to spurious retransmissions since in QoS-aware MANETs, TCP segments are often delayed rather than lost. By allowing the source sufficient time to wait for the ACK, Reno, which retransmits subsequent segments after the retransmission timer expires, benefits the most from RE-TCP implementation. Its retransmissions are considerably reduced and bandwidth is better utilized. It also shows better goodput and smaller traffic dropped.

References

[1] Buchholcz, G., Gricser, A., Ziegler, T., Van Do, T.: Explicit Loss Notification to Improve TCP Performance over Wireless Networks. In: Freire, M.M., Lorenz, P., Lee, M.M.-O. (eds.) HSNMC 2003, vol. 2720, pp. 481–492. Springer, Heidelberg (2003)

[2] Fu, S., Atiquzzaman, M.: Modelling TCP Reno with Spurious Timeouts in Wireless Mobile Environment. In: Proc. of IEEE ICCCN (October 2003)

[3] Hamadani, E., Rakocevic, V.: TCP Contention Control: A Cross Layer Approach to Improve TCP Performance in Multihop Ad hoc Networks. In: Proc. of WWIC (May 2007)
[4] Holland, G., Vaidya, N.: Analysis of TCP performance over mobile ad hoc networks. ACM Wireless Networks 8(2), 275–288 (2002)
[5] Kim, D., Bae, H., Song, J., Cano, J.C.: Analysis of the interaction between TCP variants and routing protocols in MANETs. In: Proc. of ICPP (June 2005)
[6] Kesselman, A., Mansour, Y.: Optimizing TCP Retransmission Timeout. In: Lorenz, P., Dini, P. (eds.) ICN 2005. LNCS, vol. 3421, pp. 133–140. Springer, Heidelberg (2005)
[7] Klein, T.E., Leung, K.K., Parkinson, R., Samuel, L.G.: Avoiding Spurious TCP Timeouts in Wireless Networks by Delay Injection. In: Proc. of the IEEE Global Telecommunications Conference (GLOBECOM) (December 2004)
[8] Lassila, P., Kuusela, P.: Performance of TCP on low-bandwidth wireless links with delay spikes. European Transactions on Telecommunications (April 2007)
[9] Li, X., Kong, P.Y., Chua, K.C.: TCP Performance in IEEE-802.11-Based Ad Hoc Networks with Multiple Wireless Lossy Links. IEEE Transactions on Mobile Computing 6(12), 1329–1342 (2007)
[10] Ludwig, R., Katz, R.: The Eifel algorithm: Making TCP robust against spurious retransmissions. ACM Computer Communication Review 30(1), 30–36 (2000)
[11] Mbarushimana, C., Shahrabi, A.: Resource Efficient TCP: Reducing Contention-Induced Spurious Timeouts in QoS-Aware MANETs. In: AINA 2008 (accepted, March 2008)
[12] Papanastasiou, S., Ould-Khaoua, M.: Exploring the performance of TCP Vegas in mobile ad hoc networks. International Journal of Communication Systems 17(2), 163–177 (2004)
[13] Touati, H., Lengliz, I., Kamoun, F.: TCP Adaptive RTO to improve TCP performance in mobile ad hoc networks. In: Proc. of the Sixth Annual Mediterranean Ad Hoc Networking Workshop (June 2007)

Experimental and Simulation Evaluation of OLSR Protocol for Mobile Ad-Hoc Networks

Makoto Ikeda[1], Leonard Barolli[2], Giuseppe De Marco[3], Tao Yang[1], and Arjan Durresi[4]

[1] Graduate School of Engineering
Fukuoka Institute of Technology (FIT)
3-30-1 Wajiro-Higashi, Higashi-Ku, Fukuoka 811-0295, Japan
bd07001@bene.fit.ac.jp, bd07003@bene.fit.ac.jp
[2] Department of Information and Communication Engineering
Fukuoka Institute of Technology (FIT)
3-30-1 Wajiro-Higashi, Higashi-Ku, Fukuoka 811-0295, Japan
barolli@fit.ac.jp
[3] Department of Engineering, Toyota Technological Institute
2-12-1 Hisakata, Tenpaku-ku, Nagoya 468-8511, Japan
demarco@toyota-ti.ac.jp
[4] Department of Computer and Information Science
Indiana University Purdue University Indianapolis
723 W. Michigan Street SL 280, Indianapolis, IN 46202, USA
durresi@cs.iupui.edu

Abstract. In this paper, we evaluate the performance of Optimized Link State Routing protocol by experimental and simulation results. The experiments are carried out by using our implemented testbed and the simulations by using *ns-2* simulator. We also designed and implemented a new interface for the ad-hoc network testbed in order to make more easier the experiments. The comparison between experimental and simulation results shows that for the same parameters setting in the simulation we did not have packet loss but in the experiments we experienced packet loss because of the effects of the environment and the traffic interference.

1 Introduction

In recent years, ad-hoc networks are continuing to attract the attention for their potential use in several fields. Most of the work has been done in simulation, as in general simulator can give a quick and inexpensive understanding of protocols and algorithms. However, experimentation in the real-world are very important to verify the simulation result and to revise the models implemented in the simulator. A typical example of this approach has revealed many aspects of IEEE 802.11x, like the gray-zones effect [1], which usually are not taken into account in standard simulators, as the well-known *ns-2* simulator [2].

In this paper, we concentrate on the performance analysis of a small testbed of seven computers acting as nodes of a wireless ad-hoc network. We use Optimized Link State

M. Takizawa, L. Barolli, and T. Enokido (Eds.): NBiS 2008, LNCS 5186, pp. 111–121, 2008.
© Springer-Verlag Berlin Heidelberg 2008

Routing (OLSR) which is a pro-active protocol, and it has been gaining great attention within the scientific community. Furthermore, the *olsrd* [3] software we have used in our experiments is the most updated software we have encountered.

In our previous work, we proved that while some of the OLSR's problem can be solved, for instance the routing loop, this protocol still have the self-interference problem. Moreover, there is an intricate inter-dependence between MAC layer and routing layer which can lead the experimenter to misunderstand the results of the experiments. For example, the horizon is not caused only by IEEE 802.11 MAC, but also by the routing protocol.

In this work, we deal with experimental and simulation evaluation of OLSR protocol. The simulation results showed that there is not any packet loss for our settings. However, in the experiments we experienced packet loss because of experimental environment and traffic interference.

The structure of the paper is as follows. In Section 2, we present the related work. In Section 3, we give a short description of OLSR. In Section 4, we present the testbed and simulation system description. In Section 5, we introduce the hidden node problem. In Section 6, we discuss experimental and simulation settings. In Section 7, we present experimental and simulation evaluation. Finally, conclusions are given in Section 8.

2 Related Work

In [4], the authors analyze the performance of an outdoor ad-hoc network, but their study is limited to reactive protocols such as AODV and DSR. The authors of [5], performs outdoor experiments of non standard pro-active protocols. Other ad-hoc experiments are limited to identify MAC problems, by providing insights on the one-hop MAC dynamics as shown in [6].

The closest work to ours is that in [7]. However, that work was concerned with the analysis of TCP parameters in an indoor scenario only. Moreover, it looks like that the authors did not care about the routing protocol. In [8], the disadvantage of using hysteresis-based selection of routes is presented through simulation and indoor measurements. Our experiments are concerned with the interaction of transport protocols and routing protocol, for instance OLSR. Furthermore, we compare the performance of the testbed with simulation results in an indoor scenario.

3 OLSR Overview

The OLSR protocol is a pro-active routing protocol, which builds up a route for data transmission by maintaining a routing table inside every node of the network. The routing table is computed upon the knowledge of topology information which are exchanged by means of Topology Control (TC) packets. The TC packets in turn are built after every node has filled its neighbors list. This list contains the identity of neighbor nodes. A node is considered a neighbor if and only if it can be reached via a bi-directional link. OLSR checks the symmetry of neighbors by means of a 4-way handshake based on the so called `HELLO` messages. This handshake is inherently used to compute the packet loss probability over a certain link. This can sound odd, because

packet loss is generally computed at higher layer than routing one. However, an estimate of the packet loss is needed by OLSR in order to assign a weight or a state to every link.

In OLSR, control packets are flooded within the network by electing special nodes, called Multi Point Relays (MPRs), to the role of forwarding nodes. By this way, the amount of control traffic can be reduced. These nodes are chosen in such a way that every node can reach its neighbors 2-hops far away. In our OLSR code, a simple RFC-compliant heuristic is used [9] to compute the MPR nodes. Every node computes the path towards a destination by means of a simple shortest-path algorithm, with hop-count as target metric. In this way, a shortest path can result to be also not good, from the point of view of the packet error rate. Accordingly, recently OLSRd has been equipped with the Link Quality (LQ) extension, which is a shortest-path algorithm with the average of the packet error rate as metric. This metric is commonly called as the Expected Transmission Rate (ETX), which is defined as $\mathrm{ETX}(i) = 1/(NI(i) \times LQI(i))$. Given a sampling window W, $\mathrm{NI}(i)$ is the packet loss probability seen by a node on the i-th link during W. Similarly, $\mathrm{LQI}(i)$ is the estimation of the packet loss seen by the neighbor node which uses the i-th link. When the link has a low packet error rate, the ETX metric is higher. The LQ extension greatly enhances the packet delivery ratio with respect to the hysteresis-based technique [10].

4 Testbed and Simulation System Description

4.1 Testbed Description

Our testbed is composed by six laptops and one gateway machine as shown in Fig. 1. The operating system mounted on these machines is Linux with kernel 2.6, suitably modified in order to support the wireless cards. These IEEE 802.11g cards are all from the same vendor, except for two nodes, the gateway (GW) and the first hop from the gateway (clio), which use a LynkSys card. However, the chipset is the same and is based on the Broadcom series. The driver can be downloaded from the Web sites in references [11,12]. This wireless USB card comes with an external antenna. We verified that the external antenna improves the quality of the first hop link, which is the link connecting the ad-hoc network with the GW.

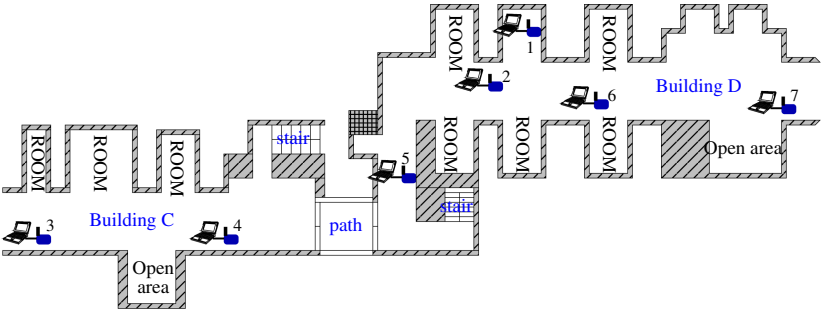


Fig. 1. Experimental model with 7 nodes

The GW machine serves as DNS system and Internet router for the nodes in the ad-hoc network. This feature are provided by the `iptables` mechanism, readily available under Linux machines. By this way, the GW can be accessed ubiquitously from anywhere. Moreover, the GW hosts also all the routines used to coordinate the measurement campaign, as well as graphical tools to check network connectivity.

In our testbed, we have two systematic background or interference traffic we could not eliminate: the control traffic and the other wireless APs interspersed within the campus. The control traffic is due to the `ssh` program, which is used to remotely start and control the measurement softwares on the GW machine. The other traffic is a kind of interference which is typical in an academic scenario.

We used the UDP protocol and studied the Mesh Topology (MT). In the MT scheme, the MAC filtering routines are not enabled. We collected data for four metrics: the throughput, Round-Trip Time (RTT), jitter and packet loss. These data are collected by using D-ITG [13], which is an open-source traffic generator. D-ITG computes the packet loss as the number of lost packet divided by the effective number of sent packets.

Every experiment is repeated 50 times. In the experiments, we suppose only one single test-flow. The source is always located at the GW and the destination is changed along the experiments. By this way, we fairly focus the study on the performance of the routing protocol and not on other phenomena, such as congestion or throughput saturation.

4.2 Testbed Interface

Until now, all the parameters settings and editing were done by using command lines of Linux Operating System (OS), which resulted in many misprints and the experiments were repeated many times. In order to make the experiments easier, we implemented a testbed interface. The interface is shown in Fig. 2. For the Graphical User Interface

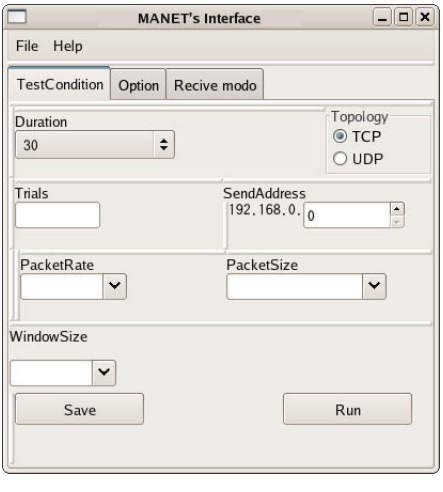


Fig. 2. GUI interface

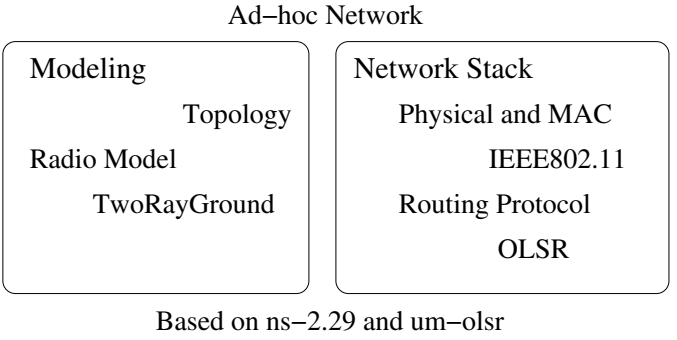


Fig. 3. Simulation model

(GUI) we used wxWidgets tool and for GUI Builder the wxGlade. The definition for each button was done by wxGlade and each operation were implemented by Perl language.

We implemented 6 parameters in the interface: transmission duration, trial number, source address, packet rate, packet size and window size. We can save the data for these parameters in a text file and can manage in a better way the experimental conditions.

4.3 Simulation System Description

The simulation system is shown in Fig. 3. We used the same parameters setting as the testbed. We consider in the simulation system the modeling and the network stack. In the modeling, we implemented the topology and the radio model. We implemented two topologies: Linear Topology (LT) and MT, but in this simulation system, we used MT. As the radio models, we considered TwoRayGround and Shadowing models, but for the present simulation system we used TwoRayGround model. For the physical and MAC layer we considered IEEE 802.11 and as routing protocol for evaluation the OLSR. In the future, we will implement also other protocols.

5 Hidden Node Problem

The 802.11 MAC is a CSMA/CA access protocol based on carrier sensing with collision avoidance. The contention of the radio medium is controlled by means of a back-off timer which in turn depends on the value of a contention window. In multi-hop network, the DCF 802.11 is used. It is well known that the throughput per node scale as $O(\frac{1}{n})$. This can be easily explained by an example. Consider a linear network with step R m, as in Fig. 4. We suppose that based on the transmission range every node can be listen up to a distance of $2R$ m.

Node A can hear Node B and node C. The node B can listen the nodes A, C and D. In general, every node can be interfered by up to 4 nodes. Therefore, the available throughput scales as $1/n$ of the total available capacity. For this configuration, the throughput at node i is $1/(i+2)$, e.g. for node A it is $1/3$ and at node C it is $1/5$. The IEEE 802.11 has another well know problem which limits the sustainable throughput. It is caused

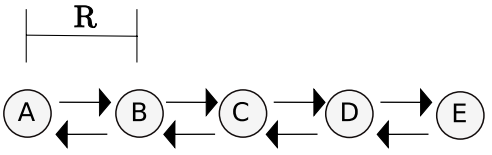


Fig. 4. Horizon effect of IEEE 802.11 when radio range is $2R$ m

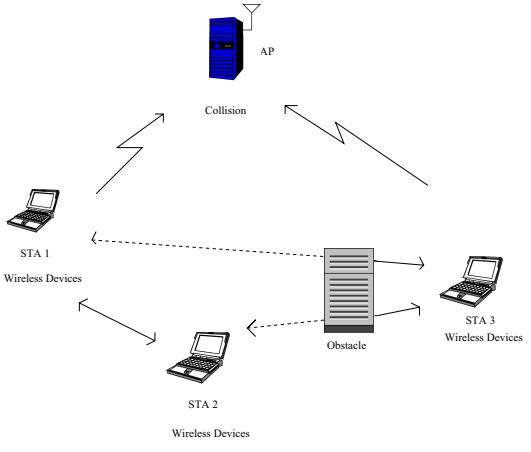


Fig. 5. Hidden node problem

by the hidden node problem which causes packet collisions. In general, this problem is more deleterious than the packet loss induced by carrier sense.

The hidden node problem is showing in Fig. 5. As we can see, because between ST1 and ST3, and ST2 and ST3 there is an obstacle, ST1 and ST3 can't hear each other because of high attenuation (e.g., substantial range), but they can both communicate with the same Access Point (AP). Because of this situation, ST2 may begin sending a frame without noticing that ST3 is currently transmitting (or vice versa). This will very likely cause a collision between ST2 and ST3 to occur at the AP. As a result, both ST2 and ST3 would need to retransmit their respective packets, which results in higher overhead and lower throughput.

In order to solve in some extent the hidden node problem we use the RTS (Request to Send) / CTS (Clear to Send) mechanism. By enabling RTS/CTS on a particular node, it will refrain from sending a data frame until the node completes a RTS/CTS handshake with another node, such as an AP. A node initiates the process by sending a RTS frame. The AP receives the RTS and responds with a CTS frame. The station must receive a CTS frame before sending the data frame. The CTS also contains a time value that alerts other nodes to hold off from accessing the medium while the node initiating the RTS transmits its data.

The RTS/CTS handshaking provides positive control over the use of the shared medium. The primary reason for implementing RTS/CTS is to minimize collisions among hidden nodes. This occurs when users and APs are spread out throughout

the facility and a relatively high number of retransmissions occur on the wireless environment.

By activating RTS/CTS, the collision will not happen. Before transmitting, ST2 would send a RTS and receive a CTS from the AP. The timing value in the CTS (which ST3 also receives) will cause ST3 to hold off long enough for ST2 to transmit the frame. Thus, the use of RTS/CTS reduces collisions and increases the performance of the network if hidden nodes are present.

6 Experimental and Simulation Settings

Our testbed is composed of 6 laptops and one desktop machine acting as a GW. The GW provided us the possibility of control the testbed from anywhere within the campus. All machines run Linux Fedora with kernel 2.6. The wireless network cards are from Lynksys, and are usb-based cards with an external antenna of 2dbi gain as shown in Fig. 6.

In previous experiments [14], we realized that an external antenna improves radio signal reception. The transmission rate of the data flows is 122 Kpkt/s = 499.712 Kbps, i.e. the packet size of the payload is 512 bytes. All experiments have been performed in indoor environment, within our departmental floor of size roughly 100 m. All laptops are in radio range of each other. We use the same method of analysis as our previous work.

Every experiment lasted 10 s and it has been repeated 50 times. The injection of traffic has been carried out by means of D-ITG, which is a traffic generator for IP networks [13]. We measured the throughput for UDP, which is computed at the receiver. We estimate the packet loss to compute the link quality metrics, e.g. LQ. For OLSR, $wT_{\mathrm{HELLO}} < T_{\mathrm{Exp}}$, where T_{Exp} is the total duration of the experiment, i.e., in our case, $T_{\mathrm{Exp}} = 500$s, and T_{HELLO} is the rate of the HELLO messages. However, the testbed was turned on even in the absence of measurement traffic. Therefore, the effective T_{Exp} was much greater.

For the simulations, we used *ns-2* simulator with the same settings as the testbed. The experimental and simulation parameters are shown in Table 1.

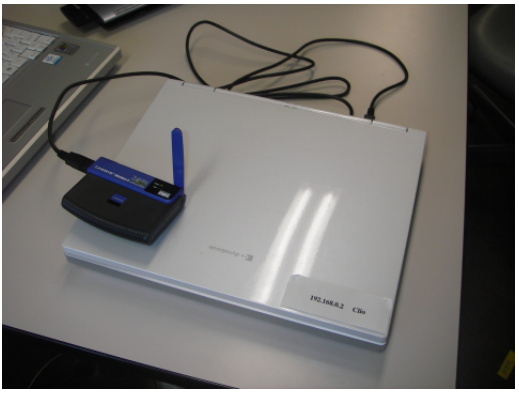


Fig. 6. Hardware settings

Table 1. Experimental and simulation parameters

Number of Nodes:	7	MAC:	IEEE 802.11b
Packet Rate (Kpkt/sec):	122	Number of Trials:	50
Packet Size (Kbytes):	512	LQ Window Size:	100
Protocol:	OLSR	RTS Thresholds:	64, 128, 256, off

7 Experimental and Simulation Results

We show the experimental results in Figs. 7, 8 and 9. In Fig. 7, we show the experimental results for throughput. We use different RTS/CTS thresholds: 256, 128, and 64. The off status means that the RTS/CTS mechanism is not activated. The better throughput is for the threshold 256. When the RTS/CTS is not activated there are a lot of oscillations.

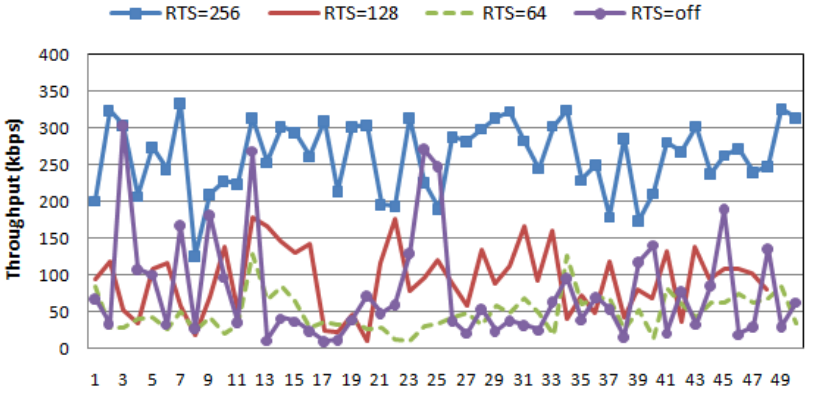


Fig. 7. Throughput

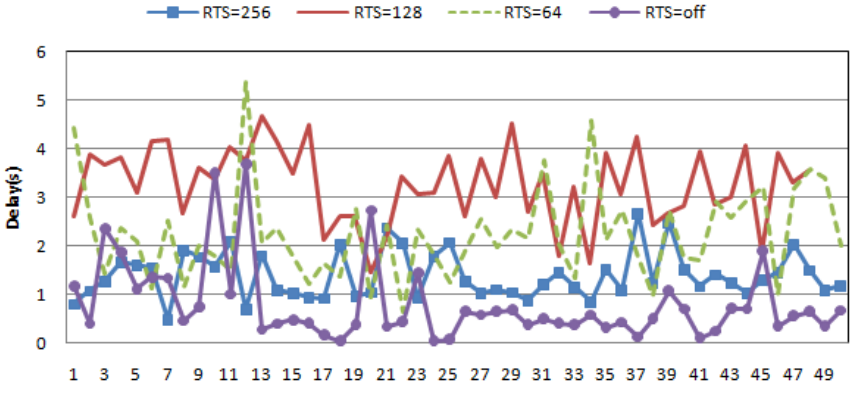


Fig. 8. RTT

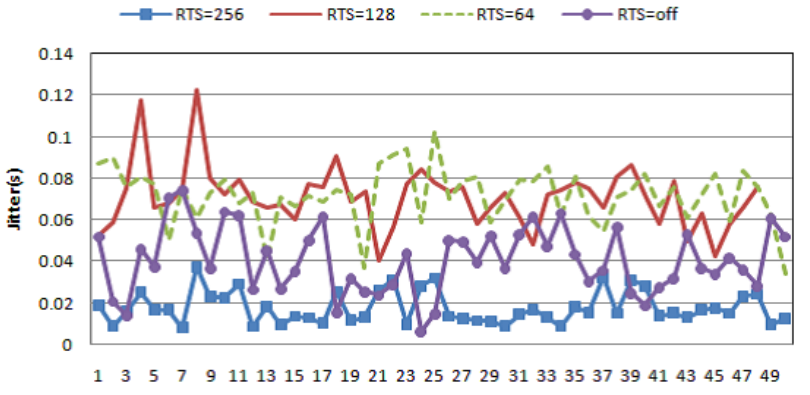


Fig. 9. Jitter

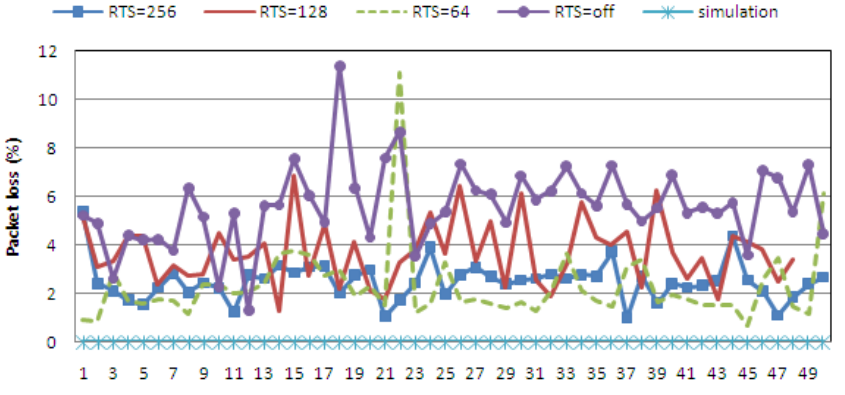


Fig. 10. Packet loss

It should be noted that an increase in performance using RTS/CTS is the net result of introducing overhead (i.e., RTS/CTS frames) and reducing overhead (i.e., fewer retransmissions). If we have small thresholds, then we have more overhead so the throughput is decreased.

In Fig. 8 are shown the experimental results for RTT. As can be seen from the figure, the RTT when the RTS/CTS mechanism is off is smaller then the other cases. However, as we show in the Fig. 8 the packet loss is higher when the RTS/CTS is off. So we need to consider some trade-off relations when using the RTS/CTS mechanism.

In Fig. 9 we show the experimental results for the jitter parameter. The case when the threshold is 256 has the smaller jitter than other thresholds and when RTS/CTS mechanism is off. This shows that by a good selection of the threshold value we can achieve a better performance of RTS/CTS mechanism.

In Fig. 10 we show the experimental and simulation results for the same settings. As can be seen from the figure, there are not packet loss during simulations but we experienced packet loss during experiments because of the effects of the environment

and the traffic interference. This shows that the experiments with ad-hoc networks are a must for getting good data to implement them in real environments.

8 Conclusions

In this paper, we carried out experimental and simulation results for a small wireless ad-hoc network with 7 nodes. We used OLSR protocol for experimental and simulation evaluation. We considered four parameters for performance evaluation: throughput, RTT, jitter and packet loss.

In order to deal with hidden node problem, we implemented RTS/CTS mechanism. The experimental results shows that we should be careful when implementing RTS/CTS. We got better results when the threshold value was 256.

For the same settings, we did not get any packet loss by simulations, but we experienced packet loss during the experiments because of the effects of the environment and the traffic interference.

In order to make easy settings of the testbed we implemented a new interface. In the future work, we would like to improve the interface and increase the number of nodes.

Acknowledgment

This work is support by a Grant-in-Aid for scientific research of Japanese Society for the Promotion of Science (JSPS). The authors would like to thank JSPS for the financial support.

References

1. Lundgren, H., Nordstro, E., Tschudin, C.: Coping with Communication Gray Zones in IEEE 802.11b Based Ad Hoc Networks. In: Proc. of the 5th ACM International Workshop on Wireless Mobile Multimedia, pp. 49–55 (2002)
2. NS-2: Network Simulator (Ver. 2), LBL, http://www.isi.edu/nsnam/ns/
3. Tønnesen, A.: OLSRd: Implementation Code of the OLSR (2006), http://www.olsr.org
4. Maltz, D.A., Broch, J., Johnson, D.J.: Lessons from a Full-Scale Multihop Wireless Ad Hoc Network Testbed. IEEE Personal Communications 8(1), 8–15 (2001)
5. Gray, R.S., Kotz, D., Newport, C., Dubrovsky, N., Fiske, A., Liu, J., Masone, C., McGrath, S., Yuan, Y.: Outdoor Experimental Comparison of Four Ad Hoc Routing Algorithms. In: Proc. of the 7th ACM International Symposium on Modeling, Analysis and Simulation of Wireless and Mobile Systems (MSWiM 2004), pp. 220–229 (2004)
6. Anastasi, G., Borgia, E., Conti, M., Gregori, E.: IEEE 802.11b Ad Hoc Networks: Performance Measurements. Cluster Computing 8(2-3), 135–145 (2005)
7. Kawadia, V., Kumar, P.R.: Experimental Investigation into TCP Performance over Wireless Multihop Networks. In: Proc. of the 2005 ACM SIGCOMM Workshop on Experimental Approaches to Wireless Network Design and Analysis (E-WIND-2005), pp. 29–34 (2005)
8. Clausen, T.H., Hansen, G., Christensen, L., Behrmann, G.: The Optimized Link State Routing Protocol, Evaluation Through Experiments and Simulation. In: Proc. of IEEE Symposium on Wireless Personal Mobile Communications, 6 pages (2001)

9. Clausen, T., Jacquet, P.: Optimized Link State Routing Protocol. IETF RFC3626 (2003)
10. De Couto, D., Aguayo, D., Bicket, J., Morris, R.: A High-throughput Path Metric for Multi-hop Wireless Routing. In: Proc. of ACM MobiCom 2003, pp. 134–146 (2003)
11. Ndiswrapper, http://ndiswrapper.sourceforge.net
12. The rt2x00 Project, http://rt2x00.serialmonkey.com
13. D-ITG: Distributed Internet Traffic Generator, http://www.grid.unina.it/software/ITG
14. De Marco, G., Ikeda, M., Yang, T., Barolli, L.: Experimental Performance Evaluation of a Pro-active Ad-hoc Routing Protocol in Out- and Indoor Scenarios. In: Proc. of IEEE AINA 2007, pp. 7–14 (2007)

A Multi-Source Streaming Model for Mobile Peer-to-Peer (P2P) Overlay Networks

Alireza Goudarzi Nemati[1], Tomoya Enokido[2], and Makoto Takizawa[1]

[1] Seikei University
{alireza.gn,makoto.takizawa}@computer.org
[2] Risho University
eno@ris.ac.jp

Abstract. In peer-to-peer (P2P) overlay networks, multimedia contents are in nature distributed to peers by downloading and caching. Here, a source peer transmits a multimedia content to a receiver peer. A peer is realized in a process of a fixed mobile computer. A peer on a mobile computer moves in the network. Thus, not only receiver peers but also source peers might move in the network. In this paper, we would like to discuss how source peers deliver multimedia contents to receiver peers in a streaming model so that enough quality of service (QoS) is supported in change of QoS of network and peer, possibly according to the movements of the peers. If a current source peer is expected to support lower QoS than required, another source peer takes over the source peer and starts sending packets of the content. The receiver peer is required to receive packets of the content with enough QoS, e.g. no packet loss even if the source peer is being switched with a new source peer.

1 Introduction

Peers holding objects like multimedia contents can support other peers with some types of objects like a part of the object in peer-to-peer (P2P) overlay networks. Thus, objects are in nature distributed to peers with various ways by downloading and caching. Peers are realized in types of computers including mobile devices [1,2]. In addition, peers may be realized in mobile agents [3,4]. Thus, some peer is moving in a network. QoS (quality of service) supported by a peer and communication link among peers is dynamically changing due to not only congestions and faults but also movements of peers. Suppose there is a receiver peer p_r which would like to listen to a music content c. The receiver peer p_r finds a source peer p_1 which can support with the music content c in a network by using types of discovery algorithms [5,6,7,8,9]. p_r starts receiving packets of the music c from p_1. p_r or p_1 may be moving in the network, QoS of a content c which p_r receives from p_1 is changing. If p_r finds QoS supported by p_1 to be degraded, another source peer p_2 is detected before QoS obtained from p_1 gets less qualified than required. p_2 starts transmitting packets of the content c. The receiver peer p_r has to receive packets of the content c with higher QoS than required even in a transition state where the current source peer is being

M. Takizawa, L. Barolli, and T. Enokido (Eds.): NBiS 2008, LNCS 5186, pp. 122–131, 2008.
© Springer-Verlag Berlin Heidelberg 2008

switched. The new source peer p_2 has to be synchronized with the current source peer p_1 on what packets of the content c to send to p_r. In this paper, we discuss a protocol for multiple source peers to support a receiver peer with a multimedia streaming service of enough QoS in change of QoS of network and source peer and in presence of movements of peers.

A multi-source steaming model (MSS) is discussed in papers [10]. A multimedia content is divided into a segment and each of the source peers sends a segment. Thus, a set of multiple source peers in parallel send segments to a receiver peer. In addition, data in a content is redundantly distributed to segments so that a receiver peer can receive every data in the content even if some number of packets are lost and some number of source peers are faulty. In this paper, one source peer sends packets of the content to a receiver peer but another source peer takes over the source peer if the source peer cannot provide enough QoS.

In section 2, we discuss the movement of the peer. In section 3, we discuss the MSS. In section 4, we present how a receiver peer receives packets from source peers. In section 5, we evaluate the MSS model.

2 Movement of Peer

A system is composed of *physical* and *overlay* layers. The physical layer shows an underlying network which is composed of nodes interconnected in networks. A node is shows fixed or mobile computer. A mobile node communicates with another node in a wireless network [11]. We assume each moving node can communicate with every other node by using protocols like mobile IP [11]. However, the communication link between a pair of nodes may not necessarily support enough and invariant QoS due to congestions and faults. An overlay layer shows an overlay network on the physical layer, which is logically composed of peer processes (*peers*) interconnected in logical communication links. A peer p_i is realized in a *supporting* node D_s. A mobile peer can manipulate a multimedia content in a supporting node. QoS supported by a mobile peer depends on the supporting node. We assume that every pair of peers can communicate with one another. If a supporting node of a peer p_i is moving away from the access point, p_i may lose some packets due to noise. Thus, QoS obtained by a peer may change due to not only congestions and faults in the underlying network but also movements of peers. In this paper, we model the *movement* of a peer to mean the change of QoS. It is critical to discuss how much qualified multimedia content a receiver peer can obtain from a source peer in a network. QoS in the underlying network is shown by parameters, bandwidth, delay, and packet loss ratio [12,13]. QoS of the multimedia content at the overlay layer are frame rate, resolution, number of colors, quality of sound, and so on depending on types of contents. A *QoS instance* is thus denoted by application level QoS parameters $\langle q_1, \ldots, q_m \rangle$. In this paper, we assume that for a pair of QoS instances Q_1 and Q_2, we can decide which one is higher than or equal to the other for simplicity. Here, "$Q_1 > Q_2$" means that Q_1 is higher than Q_2.

Let QC_{ic} be *quality of content* (QoC), which is QoS of a multimedia content c which a source peer p_i holds. Even if the content c itself supports higher QoS, enough QoS cannot be obtained from p_i if p_i is overloaded. Let QP_i indicate the *peer* QoS of a peer p_i. QoS QCP_{ic} of the content c can be obtained at p_i, $QCP_{ic} = min(QCP_{ic}, QP_i)$. Let QN_{ij} show QoS of a network link between p_i and p_j. The peer p_j finally obtains the *service* QoS Q_{ijc} from p_i. The server QoS Q_{ijc} is given to be $min(QN_j, QCP_{ic})$. In this paper, we assume the peer QoS QP_i of a peer p_i and the content QoS QC_{ic} of a content c are invariant for simplicity. Here, only the network QoS QN_{ij} among p_i and p_j changes.

3 A Multi-source Streaming Model

Let p_r be a receiver peer and p_i^c be a source peer which holds a full replica c_i of a multimedia content c ($i = 1, \ldots, m_r$). The receiver peer p_r first finds source peers of a target multimedia content c from which p_r can obtain enough QoS. Let RQ_{rc} be the minimum QoS of a content c required by p_r. Let SP_{rc} be a set of source peers which support a content c of higher service QoS than RQ_{rc}, i.e. $SP_{rc} = \{p_i^c \mid Q_{ric} \geq RQ_{rc}\}$. Suppose p_r receives the content c from a source peer p_i^c. Here, the *current* source peer p_i^c starts transmitting packets of the content c to p_r. If p_r or p_i^c may move in a network, Q_{ric} which p_r obtains from p_i^c changes. If $Q_{ric} < RQ_{rc}$, p_r cannot obtain enough QoS from the source peer p_i^c. Hence, p_i^c has to be switched with another source peer p_j^c where $Q_{rjc} \geq RQ_{rc}$. Thus, one of source peers $p_1^c, \cdots, p_{m_r}^c$ necessarily supports p_r with the streaming service of the content c of enough QoS even while source peers are switched.

A multimedia content c sent by a source peer p_i^c at the overlay layer is transmitted in a sequence of bits which are in fact carried by packets in a network at the physical layer. Let BS_{ic} be a bit sequence of a replica c_i which a source peer p_i^c transmits to the receiver peer p_r. We assume that every source peer p_i^c translates a content c to a bit sequence BS_{ic} in the same algorithm. Let $|BS_{ic}|$ be the total bit length of the bit sequence BS_{ic}. BT_{ic} indicates the total time to play the content c obtained from BS_{ic}. Here, $BT_{ic} = BT_{jc}$ if a pair of the

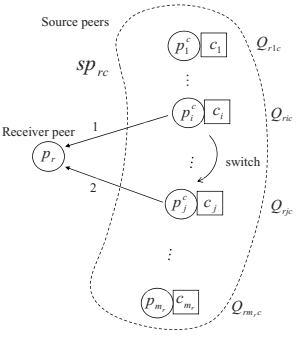


Fig. 1. Multi-source streaming (MSS) model

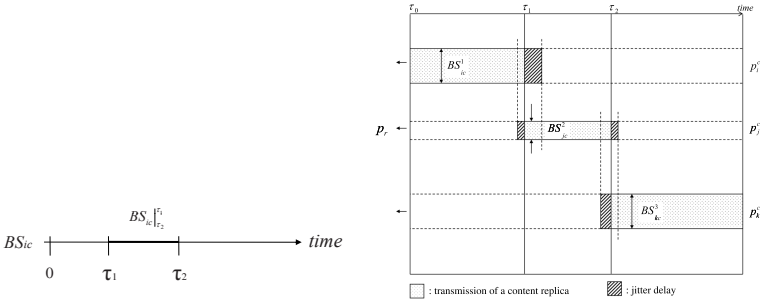


Fig. 2. Multi-source model

content replicas c_i and c_j have the same structure and support the same QoS. Let BR_{ic} show the bit rate of the content replica c_i, $BR_{ic} = |BS_{ic}|/BT_{ic}$. Let $BS_{ic}|_{\tau_2}^{\tau_1}$ be a subsequence of BS_{ic} to be played from time τ_1 to τ_2 as shown in Figure 2. A current source peer p_i^c is taken over by another source peer p_j^c if $Q_{ric} < RQ_{rc}$. Suppose p_i^c starts transmitting a bit sequence BS_{ic} at time τ_0. At time τ_1, another source peer p_j^c takes over p_i^c. Let BS_{ic}^1 be a subsequence $BS_{ic}|_{\tau_1}^{\tau_0}$ of BS_{ic}. The other source peer p_k^c takes over p_j^c at time τ_2 as shown in Figure 2. Let BS_{jc}^2 and BS_{kc}^2 be the bit sequences $BS_{jc}|_{\tau_2}^{\tau_1}$ and $BS_{kc}|_{\tau_3}^{\tau_2}$, respectively.

Suppose the network bandwidth QN_{ir} between p_r and p_i^c and QN_{kr} between p_r and p_k^c are so high that a fully colored movie can be transmitted to p_r. However, the network QoS QN_{jr} is too slow from p_j^c to transmit the fully colored movie to p_r. The source peer p_j^c can transmit only a monochromatic version c_1 of c. The movie c should be played without losing frames at p_r.

As shown in Figure 2, the receiver peer p_r receives a bit subsequence BS_{ic}^1 from one source peer p_i^c. Then, p_r receives the bit sequence BS_{jc} from another source peer p_j^c. BS_{ic}^1 and BS_{jc}^1 have to be received by p_r in a seamless manner. We have to synchronize the bit sequence BS_{ic}^2 with BS_{ic}^1. Since it takes time to synchronize the bit sequences, i.e. jitter delay, p_i^c sends the bit sequence BS_{ic} for a longer time than $\tau_1 - \tau_0$, i.e. $BS_{ic}|_{\tau_1+\delta}^{\tau_0}$. The other source peer p_j^c starts sending BS_{jc} earlier than time τ_1. That is, p_j^c sends a bit subsequence $BS_{jc}|_{\tau_2+\delta}^{\tau_1-\delta}$ from the $\tau_1 - \delta$ to $\tau_2 + \delta$. p_r receives redundantly bits of the content c in $BS_{ic}|_{\tau_1}^{\tau_1-\delta}$ and $BS_{jc}|_{\tau_1+\delta}^{\tau_1}$ from p_i^c and p_j^c. The duration from $\tau_1 - \delta$ to $\tau_1 + \delta$ is *overlapping time* of p_i^c and p_j^c. In the overlapping time, p_j^c has to synchronize the transmission of the content c with p_i^c so that p_r can receive the content c with required QoS.

4 Peer-to-Peer Communication Protocols

4.1 Receiver-Source Communications

A receiver peer p_r finds a source peer p_i^c which can support p_r with the streaming service of a multimedia content c with enough QoS, i.e. $Q_{irc} \geq RQ_{rc}$. While receiving packets of the content c from p_i^c, p_r monitors Q_{cp} and QN_{ir}

supported by p_i^c and the network, respectively. Then, p_r obtains service QoS $Q_{irc} = min(Q_{cp}, QN_{ir})$. If Q_{irc} is degraded, p_r finds another source peer p_j^c which can support enough service QoS Q_{jrc}, i.e. $Q_{jrc} \geq RQ_{rc}$.

We discuss how to coordinate a receiver peer p_r and multiple source peers $SP_r = \{p_1^c, \ldots, p_{m_r}^c\}$ $(m_r \geq 1)$ so that p_r can receive packets of a multimedia content c from one of the source peers. First, p_r finds source peers p_1^c , ..., $p_{m_r}^c$ which support the content c in a P2P overlay network by using some P2P discovery algorithm. The receiver peer p_r selects one source peer, p_i^c whose Q_{irc} satisfies the requirement QoS RQ_{rc}, $Q_{irc} \geq RQ_{rc}$. Then, p_r sends a *Content* request *Creq* to p_i^c, which includes parameters $\langle \tau, RQC_{ic}, RQ_{rc} \rangle$ where RQC_{ic} is the required QoC of the content c and RQ_{rc} is the service QoS which p_r would like to receive from p_i^c. τ shows from what bit in the bit sequence BS_{ic} p_i^c should send. Here, $\tau = 0$, i.e. p_i^c sends from the first bit of BS_{ic}. If p_i^c agrees on serving the content c, p_i^c sends ACK with QP_i and QC_{ic} to p_r. Then, p_r calculates Q_{ric} from QP_i, QC_{ic}, and QN_{ir} and sends a *Start* message with start time τ_s and end time τ_e to p_i^c. Then, p_i^c starts transmitting packets of the content c. Here, τ_s is 0 and τ_e is time from the beginning to the end of the content. The receiver peer p_r starts receiving packets from p_i^c. If p_i^c sends every packet of the content c. from time τ_s to τ_e, p_i^c sends an *End-of-Content (EoC)* message to p_r. Then, p_r sends *Disconnect* to p_i^c.

During the transmission of packets from a source peer p_i^c, the service QoS Q_{ric} might be degraded, e.g. due to the movement of the receiver peer p_r. First, suppose p_r finds the degradation of Q_{ric} from p_i^c. The receiver peer p_r calculates Q_{irc} each time p_r receives a packet from p_i^c. Even if $Q_{ric} \geq RQ_{rc}$, p_i^c is *suspicious* if $Q_{ric} \leq TRQ_{ri}$. $TRQ_{ri} = \alpha * RQ_{rc}$ where $\alpha (\geq 1)$ is a constant. If p_i^c is suspicious for some time units or $Q_{ric} < SRQ_{ri}$, p_i^c is *dangerous*. If $Q_{ric} < RQ_{rc}$, p_i^c is *faulty*. If p_i^c gets dangerous, p_r finds another source peer p_j^c. It takes time T to switch a source peer since p_r has to find another source peer p_j^c and do negotiation with p_j^c. At each time t, p_r estimates *breaking time* bt_t when p_r could not receiver enough QoS from p_i^c if Q_{irc} is being degraded. Hence, $|bt_t - t| > T$ is required to be satisfied. Otherwise, the current source peer is switched with

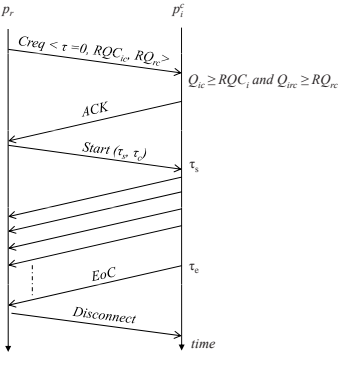


Fig. 3. Receiver source communication

another source peer without obtaining enough QoS. Here, let $Q_{irc}(t)$ show QoS Q_{irc} which is supported by p_i^c at time t. The threshold value TRQ_{ri} is obtained as $TRQ_{ri} = Q_{irc}(t_1),\ t_1 > t,$ and $(bt_t - t_1) \geq T$.

Next, if p_i^c could not support p_r with enough QoS, p_i^c sends a *Retire* message to p_r. On receipt of *Retire*, p_r tries to find another source peer. If p_r finds another peer p_j^c and p_j^c starts sending packets, p_r sends *Disconnect* to p_i^c.

4.2 Receiver Peers

Suppose that a receiver peer p_r is receiving packets of a multimedia content c from a current source peer p_i^c. Both p_r and p_i^c monitor QoS obtained from p_i^c through the network and QoS which p_i^c is supporting, respectively. Suppose p_r detects that Q_{irc} is expected to be under the required QoS ($Q_{irc} < RQ_{rc}$) some time units τ later. Let δ_{ri} be delay time between p_r and p_i^c. τ has to be longer than $2 * \delta_{ri}$. In fact, if Q_{irc} gets smaller than some QoS value TQ_r, p_r considers p_i^c to be suspicious. If p_i^c is suspicious for λ time units, p_r starts finding another source peer p_k^c. Here, p_r finds another source peer p_k^c of the content c and sends *Change* to p_k^c. On receipt of *Change*, p_k^c sends ACK to p_r if p_k^c could send the content c where $RQ_{rc} \leq Q_{kc}$. Then, p_r sends *Start* to p_k^c.

A multimedia content c is realized as a sequence $P(c)$ of packets, $pk_1, \cdots, pk_q$ ($q \geq 1$) to be transmitted in the network. Each packet carries a subsequence of the bit sequence BS. Each packet pk is identified in the sequence number $pk.seq$. In p_r, a variable REQ shows the sequence number seq of a packet of the content c which p_r has most recently received from a source peer p_i^c. p_r sends a *Start* request with sequence number REQ to p_k^c. On receipt of *Start*, p_k^c calculates the delay time δ_{rk} to p_r and the minimum transmission rate $minTR_{rk}$ which is decided by QoS of the content c. Then, the sequence number seq of a packet which p_k^c starts transmitting is calculated as $seq = REQ + minTR_{rk} * \delta_{rk}$. p_k^c starts transmitting packets from a packet pk_{seq} as shown in Figure 4.

On entering a suspicious state, {
  $count = 0$; $SREQ = REQ$; $SEQ = REQ + minTR_{rk} * \delta_{rk}$;
  p_k^c = *Find-New-Source* (c); **send** start **to** p_k^c }
On receipt of a packet pk **from** p_k^c, {
if *current* $\neq p_k^c$, {
  if $REQ > pk.seq$, **discard** pk;
  else if $REQ = pk.seq$, { **receive** pk;
    $REQ = REQ + 1$; $count = count + 1$;
    if $REQ - SREQ > minPN$ **and** count$/(REQ - SREQ) \geq 0.9$,
    {**send** *Disconnect* **to** *current*; $current = p_k^c$; } }
  else discard pk; }
} **else if** $REQ = pk.seq$, { **receive** pk; $REQ = REQ + 1$; }
  else if $REQ < pk.seq$ { $REQ = pk.seq + 1$; **discard** pk; }
  else discard pk; }

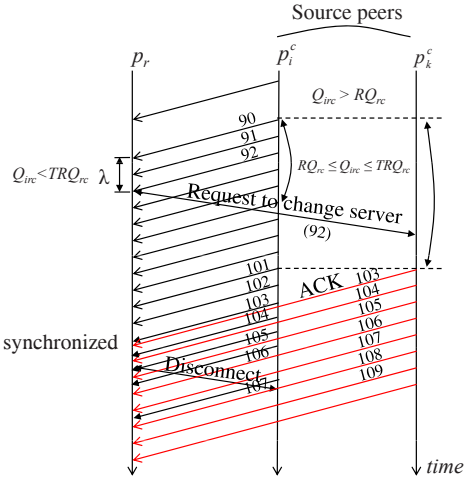


Fig. 4. Change of source peer

minPN shows the number of packets. After receiving the number *minPN* of packets since sending a *Start* message to the source peer p_k^c if p_r receives packets in order from p_k^c, p_r changes the current source peer with p_k^c.

4.3 Source Peers

If the receiver peer p_r decides on changing the current source peer p_i^c, p_r sends *Disconnect* to p_i^c. On receipt of *Disconnect*, p_i^c stops transmitting packets to p_r.
[**Receiver peer** p_r] p_r finds another source peer p_i^c and sends a *content* request with QC_c and RQ_r to p_i^c

On receipt of an ACK message **from** p_i^c, {
  $current = p_i^c$; $REQ = 1$; **send** *Start* **to** p_i^c; }
On receipt of a NACK message **from** p_i^c, {
  find another source peer; **go to** the beginning of this procedure; }
[**Source peer** p_i^c]
On receipt of a *Content* request *Creq* with QC_{ic} and RQ_{rc} **from** p_r, {
  if $Q_{ic} > RQ_{rc}$, { **send** ACK **to** p_r; $SEQ = 0$;}
  else send NACK **to** p_r; }
On receipt of a *Start* request **from** p_r, {
  $SEQ = SEQ + 1$; **send** a packet pkt_{SEQ} **to** p_r; }
On receipt of a *Disconnect* request **from** p_r, {
  stop sending packets **to** p_r; }

5 Evaluation

In the evaluation, there are a moving receiver peer p_r and fixed source peers $p_1^c, \cdots, p_n^c$ $(n \geq 1)$ of a multimedia content c. Then, we consider environment

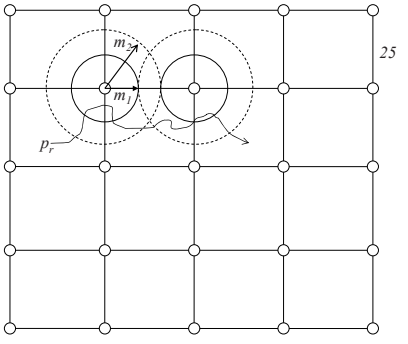


Fig. 5. Mesh

where the source peers $p_1^c, \cdots, p_n^c$ are uniformly distributed in an area where p_r moves. Here, we make the following assumptions:

1. Totally 25 source peers are uniformly deployed in a 100×100 mesh as shown in Figure 5. The mesh interval is 25 [unit].
2. The receiver peer p_r knows where each source peer p_i^c exists in the network.
3. The receiver peer p_r moves with velocity v_r in one direction to a location $\langle x,\ y \rangle$ for one time unit and then randomly changes the direction to a location $\langle x + v_r sin\theta,\ y + v_r cos\theta \rangle$ which angle θ is randomly calculated. Here, $v_r = 0.01$ [unit/sec]. One unit time shows 100[msec.].
4. The bandwidth between the receiver peer p_r and a source peer p_i^c is considered as the service QoS Q_{irc}. Q_{irc} depends on the distance d_{ir} between p_r and p_i^c. The maximum bandwidth is 1 which peer p_r can obtain if p_r exists on the same location as a source peer p_i^c.
5. The required bandwidth RBW is 0.56. $TRQ_{ri} = 0.6 + 0.4 \cdot RBW = 0.824$ and $SRQ_{ri} = 0.3 + 0.7 \cdot RBW = 0.692$.

$BW(d)$ shows the bandwidth which a receiver peer p_r can receive at distance d from a source peer. $BW(d) = 1$ for $0 \leq d \leq m_1$. For $m_1 < d \leq m_2$, $BW(d)$ is one randomly selected from 1 to 0.05. When $m_2 > 2 * m_1$, $BW(d)$ is approaching to 0 from 0.05 for $d > m_2$. The larger m_1 is, with the stronger radio a source peer sends messages to p_r. The assumptions mean that one unit shows 10[m] where a person is walking with around 4[km/hour].

In our multi-source streaming (MMS) model, a current source peer p_i^c transits to a suspicious state in a receiver peer p_r. Here, p_r finds another source peer p_j^c and negotiates with p_j^c before the current source peer p_i^c would not support enough QoS. On the other hand, in the traditional (T) model, the receiver peer p_r starts finding another source peer if QoS supported by the current source peer is degraded. Since it takes time to find and negotiate with another source peer, the receiver peer p_r may not receive enough QoS. If a receiver peer p_r does not receive enough QoS, p_r is referred to as *faulty*. Figure 6 shows the average *fault* ratios of the multi-source streaming (MSS) model and the traditional (T) model

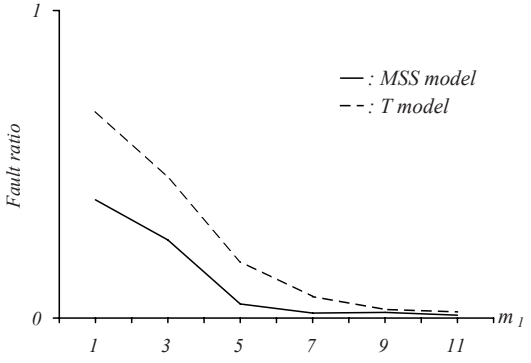


Fig. 6. Fault ratio

that p_r is in faulty state for m_1. As shown in Figure 6, the *MSS* model impress the less fault ratio than the T model. For example, the fault ratio of the *MSS* model is 4.59% while 18.35% in the T model for $m_1 = 5.5$.

6 Concluding Remarks

In this paper, we discussed how to support a receiver peer with enough QoS of the multimedia steaming service by multiple source peers. Not only a receiver peer but also source peer is moving in a network. Here, QoS supported by the source peer is changing according to the movement of the receiver peer and source peer. In this paper, a receiver peer switches a source peer with another source peer if the source peer might not support enough QoS. We discuss the multi-source streaming protocol which a receiver peer can receive packets of a multimedia content from multiple source peers. We evaluated the multi-source streaming protocol and compared with other approaches.

References

1. Shimoura, H., Tenmoku, K.: Development of elemental algorithms for future dynamic route guidance system. In: Proc. Vehicle Navigation and Information Systems Conference, August 31– September 2,1994, pp. 321–326 (1994)
2. Zijderhand, F., Biesterbos, J.: Functions and applications of socrates: a dynamic in-car navigation system with cellular-radio based bi-directional communication facility. In: Proc. Vehicle Navigation and Information Systems Conference, pp. 543–546 (1994)
3. Shiraishi, S., Enokido, T., Takizawa, M.: Fault-tolerant mobile agent in distributed objects systems. In: Proc. of the 9th IEEE International Workshop on Future Trends of Distributed Computing Systems (FTDCS 2003), pp. 145–151 (2003)
4. Tanaka, Y., Hayashibara, N., Enokido, T., Takizawa, M.: Fault-Tolerant Transactional Agent Model for Distributed Objects. In: Proc. of 9th IEEE International Symposium on Object and component-oriented Real-time Distributed Computing (ISORC 2006), pp. 279–286 (2006)

5. Watanabe, K., Enokido, T., Takizawa, M., Kim, K.: Charge-based Flooding Algorithm for Detecting Multimedia Objects in Peer-to-Peer Overlay Networks. In: Proc. of IEEE 19th Conference on Advanced Information Networking and Applications (AINA-2005), vol. 1, pp. 165–170 (2005)
6. Watanabe, K., Nakajima, Y., Enokido, T., Takizawa, M.: Ranking Factors in Peer-to-Peer Overlay Networks. ACM TASS 2(3), 11:1–11:26 (2007)
7. Rowstron, A., Druschel, P.: Pastry: scalable, distributed object location and routing for large-scale peer-to-peer systems. In: Proc. of IFIP/ACM International Conference on Distributed Systems Platforms (Middleware), pp. 329–350 (2001)
8. Stoica, I., Morris, R., Karger, D., Kaashoek, F., Balakrishnan, H.: Chord: A Scalable Peer-to-Peer Lookup Protocol for Internet Applications. IEEE/ACM TON 11(1), 17–32 (2003)
9. Zhao, B.Y., Kubiatowicz, J., Joseph, A.D.: Tapestry: An Infrastructure for Fault-resilient Wide-area Location and Routing. Technical Report UCB/CSD-01-1141, University of California, Berkeley (2001)
10. Itaya, S., Hayashibara, N., Enokido, T., Takizawa, M.: Distributed coordination protocols to realize scalable multimedia streaming in peer-to-peer overlay networks. In: Proc. of ICPP 2006, pp. 569–576 (2006)
11. Perkins, C.: Mobile ip. Communications Magazine 35, 84–99 (1997)
12. Jung, J.I.: Quality of service in telecommunications.ii. translation of qos parameters into atm performance parameters in b-isdn. Communications Magazine 34(8), 112–117 (1996)
13. Jung, J.I.: Quality of service in telecommunications. i. proposition of a qos framework and its application to b-isdn. Communications Magazine 34(8), 108–111 (1996)

Building a Linux Grid on a Virtual Machine Using a Windows Grid

Kenichi Tanaka, Minoru Uehara, and Hideki Mori

Department of Open Information Systems, Toyo University, Japan
{gz0700187,uehara,mori}@toyonet.toyo.ac.jp

Abstract. In recent years, computing demands have increased greatly. Many researchers have studied on grid computing in order to satisfy the demands. Several researches focus on virtual computing environment known as cloud computing. In cloud computing, computing resources are virtualized, divided into small units and re-organized as arbitrary heterogeneous platform. However, it is difficult to develop the application because cloud computing needs massive resources. In this paper, we propose virtual grid for cloud computing, in which Linux grid is constructed on Windows grid. This utilizes many Windows PCs as grid platform and realizes all functions of Linux grid.

Keywords: Grid, Virtual Machine.

1 Introduction

With the improvements of hardware in recent years, PCs for home users now offer substantial computing resources. However, in most cases these resources are not fully utilized and can be considered as dead resources. Grid Computing is a technique that enables these dead resources to be incorporated into a huge Virtual Computer. A grid is more flexible than a cluster, because the grid can adjust computing resources dynamically according to the requested services.

Research into Grid Computing is on-going. In Japan, MEXT (the Ministry of Education, Culture, Sports, Science and Technology) has instituted the National Research Grid Initiative (NAREGI) Project [5]. There is increased interest in the use of grids, as the computational demands increase, not only with respect to scientific computing, but also in the field of business and graphics technology.

Software support for grids is mainly limited to UNIX / Linux, making it difficult to use the computing resources of Windows PCs. As Windows PCs are the most popular computers (with a market share of 91%). If grid middleware support Windows, many more computing resources would be available for the grid. For example, business software like Systemwalker CyberGRIP from Fujitsu makes use of Windows PCs as Computational nodes, although it is not able to use a Windows PC as the Manager. We have previously described a grid created and used with Windows in [1] [2]. Grid middleware for Windows has many restrictions, and most necessary functions for creating a grid are available only for UNIX / Linux Platforms.

M. Takizawa, L. Barolli, and T. Enokido (Eds.): NBiS 2008, LNCS 5186, pp. 132–141, 2008.
© Springer-Verlag Berlin Heidelberg 2008

One way of creating a grid using Windows, is to develop special middleware that supplements existing middleware by providing the necessary functionality. Another way is to create a Linux Grid on virtual machines. The latter is the simpler method for creating a grid on Windows without overhead of virtualization, but it does introduce the problem of decreased performance due to virtualization. A solution to this is given in [3] and therefore grids on virtual machines are viable. The target of this research uses the latter method.

In this paper, we propose a system for controlling a grid on virtual machines using a Windows Grid. This system is composed of a Virtual Machine control service, service management service, and virtual disk image management service. Related techniques are mentioned in Chapter 2. We describe an overview of the proposed system in Chapter 3. Results of the evaluation of the system are given in Chapter 4, while Chapter 5 discusses the issues arising from this evaluation. In Chapter 6, we present the conclusions, benefits and future tasks.

2 Related Techniques

2.1 Grid Middleware

Grid Middleware is software for building grids that reduces the burden of development. The mainstream Grid Middleware is the Globus Toolkit by Globus Alliance.

Globus toolkit is global-standard Grid Middleware developed by Globus Alliance [4]. We used this toolkit in our research.

This toolkit builds grid services as web services based on the WS-Resource Framework [6], which means that the web services can be released without security risks. The Globus Toolkit uses the knowledge of web service development and also, cooperates with Apache Tomcat instead of the internal service container.

Core functions of this toolkit are written in Java, while data management functions, job management functions, etc. are written in C, and are thus only supported in UNIX / Linux.

2.2 Virtual Machine

Virtual machine technology allows the creation of virtual hardware on a real computer. Multiple virtual machines can be executed on a high-end computer. Currently there are two main types of virtual machine; one is the GuestOS, while the other is the Virtual Machine Monitor.

GuestOS. Virtual machines created with GuestOS emulate hardware with software and therefore result in reduced performance of the virtual machine. Instead of this, we can use the existing OS on a virtual machine of this type. Such a machine can be used whenever a virtual machine instead of a real machine is called for, and can test software or the OS itself. As examples, VMware Workstation / Player [10] and Microsoft Virtual PC [11] are GuestOS type virtual machines. We used VMware Player in this research as it has a corresponding implementation in Linux.

Virtual Machine Monitor. In this case, the monitor inserts virtual hardware between the OS and real hardware. The host OS does not exist; instead virtual machines that operate by special execution authority perform this job. When compared to the GuestOS type of virtual machine, there are fewer overheads at the time of execution, as the virtual machine is created without using the host OS. A typical virtual machine monitor is Xen [12].

2.3 Cloud Computing

In recent years with the wider availability of always-on connections to the Internet, processes previously executed by the client are now performed by a server. The conventional desktop application has moved to a server, and data has been moved to a data center. The concept of this kind of environment is called Cloud Computing. Calculation resources beyond the Internet and the Internet itself are often called "clouds." In Cloud Computing, services can be accessed independently of devices by using suitable access "clouds." Software existing in the server can be used with data saved at whichever data center, without taking into account the internal structure.

Amazon EC2. Amazon Elastic Compute Cloud (Amazon EC2 [7]) is a virtual server environment offered by Amazon as one of their web services. The virtual environment known as an instance is used in Amazon EC2, and the operation thereof is performed via a web service.

Whilst many resources can be obtained by starting two or more instances simultaneously, all unnecessary instances should be terminated to keep down the overhead. However, if the data in an instance is required and since it has not been saved, a virtual environment like the Amazon Machine Image (AMI) needs to be created, and the data can then be saved using Amazon Simple Storage Service (Amazon S3).

3 Virtual Machine Grid Service

In this research we have developed a virtual machine grid service (Virtual Machine Grid Services, VMGrid Services), VMGridMasterService for collecting VMGridService(s) that manage the state of virtual machines and the information, and for performing package management. It consists of VMGridDiskService that manages the server holding the disk image of a virtual machine, and a service monitor which is an interface to a management service. A map of this system is shown in Fig. 1.

Fig. 1 shows the relation between VMGridMasterService, VMGridService, a disk server and a virtual machine. The direction of the arrows in the figure represents the flow of data in terms of a request and a response. Although not shown in the figure, VMGridDiskService exists as a service for the virtual machine disk server management. This is a service which VMGridService uses through VMGridMasterService, and offers access to a disk server.

Although communication for each service is fundamentally performed by the request / response of SOAP, when the state of the service is changed, or when it changes dynamically, the change is made known by WS-Notification [9]. Also in this

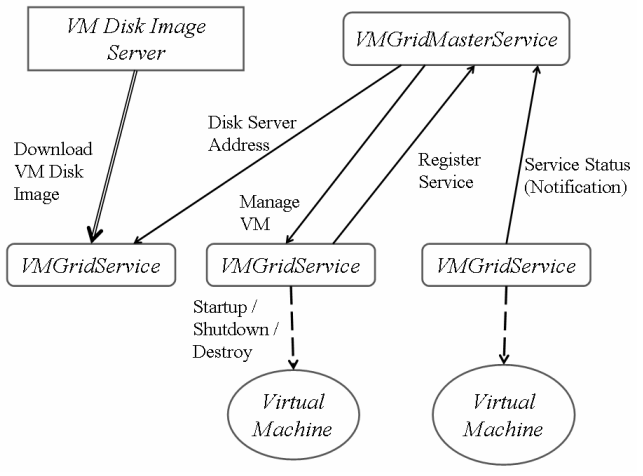


Fig. 1. Key map of VMGrid Services

system, a service monitor supervising Notification is notified by VMGridMasterService and VMGridService, and updates the state of a monitor. In addition, it reports the static setup of each service to a Java property file.

3.1 VMGridService

VMGridService is a service which performs deployment and execution control of a virtual machine. It is created automatically under management of the VMGridMasterService and registers its own URI with the VMGridMasterService at startup time.

Prepare. Prepare operation initializes a virtual machine, can be performed when a virtual machine disk image does not exist. The process performed is the following.

1) Acquire the address of VMGridDiskService from VMGridMasterService.
2) Acquire the address of Disk Server from VMGridDiskService.
3) The information from the archive file of a virtual machine image is acquired.
4) A virtual machine image is downloaded to a temporary directory allocated by the OS, and is deleted once communication terminates.
5) An archive file is created and deployed in the working directory of the service. Moreover, when the VMware configuration file is discovered in the archive, its path is added as a property to the service, and it is used for subsequent virtual machine control.

If a virtual machine image already exists, nothing further is done if it is in a position to be started. Moreover, in order to avoid a timeout, this is performed by another thread, and it returns from the request after only a short time.

Execution control can be performed after the deployment of a virtual machine image has been completed. Execution control includes the processing of 1) Startup, 2) Shutdown, and 3) Destroy.

Startup. This operation starts a virtual machine. A run state is acquired by the process object. However, in order to return without waiting for the end of the process, an independent method is needed that can ascertain whether the starting situation of the grid on a virtual machine etc. can be used. Moreover, although it is known when the startup of the VMware process has failed, when it terminates unexpectedly as in the case when it freezes during startup etc., you have to acquire a state again.

Shutdown. The operation for terminating a virtual machine safely is Shutdown. However, since the method for ending a process safely does not exist in the Process class, a WM_CLOSE message is transmitted to the window of a virtual machine using JNI (Java Native Interface [8]).

Destroy. When a virtual machine cannot be terminated by the Shutdown operation in the case where loading of the JNI library fails etc., this is the operation for compulsory termination of a process. Since the process ends shortly after performing this operation, the state of a virtual machine may be affected.

3.2 VMGridMasterService

VMGridMasterService is a management service that performs deployment and execution control of a virtual machine to all VMGridService(s) that are attached to the relevant VMGridService node, and management using the notice by WS-Notification.

Self is initialized at the time the service container starts, and the notice by Notification from VMGridService is adhered to. If the VMGridService is registered, the URI of the service newly connected to the service monitor will be notified, because self publishes a Notification. One of the functions of the VMGridMasterService is the acquisition of a URI list of VMGridService under management, and another is demand transmission to two or more VMGridService(s).

The same Startup / Shutdown / Destroy interface as in VMGridService is defined for VMGridMasterService, and this allows processing of the same name to all

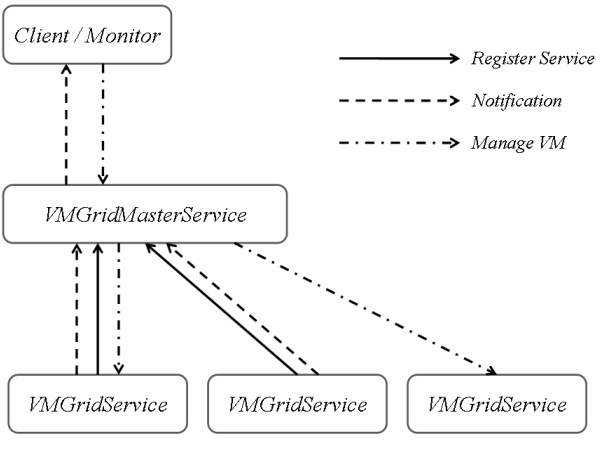


Fig. 2. Key map of Communication model

registered VMGridService(s). Fig. 2 shows the communication model between VMGridMasterService and VMGridService.

3.3 VMGridDiskService

This is needed to give access to a server for deployment of a virtual machine. Although the service opens a connection with a server according to the connection demand, if it needs to change the connection according to the load at the time etc., deployment of a virtual machine image can perform it efficiently.

A disk server is a thread called from VMGridDiskService. A socket is opened for every connection and processing of two or more demands is enabled by generating the thread written in there. However, since it operates within the same process as a service, performance is not good.

3.4 Service Monitor

A service monitor is the client to VMGridMasterService, and executes by notice according to the state of the master service to acquisition or Notification. The interface of a service monitor implemented in Java Swing is shown in Fig. 3.

The state of the virtual machine on each VMGridService can be ascertained from the service monitor. Deployment and control of a virtual machine can be individually performed through the interface prepared for every instance of VMGridService. All registered VMGridService(s) are also controllable with the buttons located in the lower part.

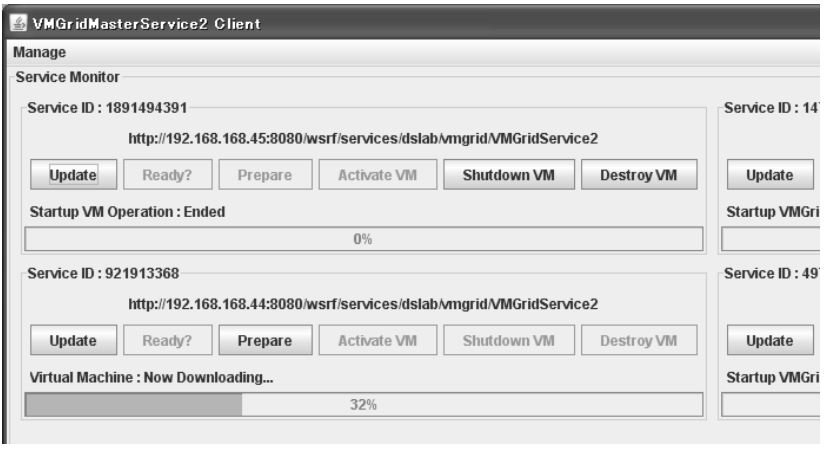


Fig. 3. Service Monitor Interface

4 Evaluation

The evaluation was performed with respect to the deployment time of a virtual machine. Deployment time is the time taken to download a virtual machine image and deploy the archive file of the image.

The following two patterns comprised the evaluation

A) Deploy to 1 service.
B) Deploy to 4 services at one time.

Pattern A is executed first to ascertain the performance of the fundamental system. Thereafter Pattern B is executed to determine the decline in performance while arranging more than one set simultaneously.

Table 1 describes the environment used for evaluation of the virtual machine deployment time of the system. Table 2 contains information on the archive file of a virtual machine image used for evaluation. In addition, the network connecting the host machines is a 100Mbps LAN.

Table 3 shows the results of evaluating deployment time. The download time of an archive file increases in proportion to the number of services in Table 3. Since each service performs to deploy an archive, there is not much difference between the average deployment time for 4 services and the deployment time for 1 service.

Table 1. Specification of Host machines

Service	VMGridMasterService	VMGridService
OS	Windows XP Professional x64	Windows XP Home
CPU	AMD Athlon X2 Dual Core Processor 3800+, 2.00 GHz	AMD Athlon X2 Dual Core Processor 4600+, 2.41 GHz
Memory	3.93 GB RAM	1.93 GB RAM
Java	JRE 1.6.0 Update 3 (Server VM)	JRE 1.6.0 (Client VM)

Table 2. Archive file of Disk image

File Formats	ZIP	
File Size [MB]	After compression	2047
	Before compression	5725
	Compression ratio	0.357686626

Table 3. Deployment Time

#VMGS	Service #	Download [ms]	Expand [ms]	Total [ms]
1		188078	298563	492484
4	1	728188	291765	936704
	2	729609	251234	987110
	3	726594	327500	1059922
	4	728390	412563	1148282
	Average	728195.25	298265.5	1033004.5

5 Considerations

Table 4 shows the download speed *Vd*, the reading speed at the time of deployment *Ver*, and the corresponding writing speed is *Vew*. The unit of speed used is MB/s. These services are the same as VMGS 1-4 in Table 3.

Table 4. Average speeds

	Vd [MB/s]	*Ver* [MB/s]	*Vdw* [MB/s]
1 Service	10.88804667	6.858860746	19.17561421
4 Services	2.812160669	6.865702004	19.19474061

Table 4 shows a big decrease in download speed. The speed for the deployment of one set is equal to Vd_A / Vd_B = 0.25827963 when the speed in case of 1 service assumed to Vd_A, and the speed in case of 4 services assumed to Vd_B. This ratio is about 1 / 4. Therefore, it is shown that the performance of the disk server offered by VMGridDiskService is inversely proportional to the number of virtual machines which will be arranged simultaneously in this system. In other words, if multiple VMGridServices handle prepare requests which require deployment of virtual machines, the speed of downloads falls remarkably with the performance of the disk server. But, the deployment of archive files is carried out by each service, so deployment speeds shown are almost constant. Deployment speed does not depend on the number of downloads.

The time *T* for the deployment of a virtual machine is the sum of the download time *Td* and the deployment time *Te*. These times are expressed respectively in terms of the archive size of a virtual machine S and the speed *Vd*, and the compression ratio *C* and speed *Vew*.

$$\mathrm{Td} = \frac{\mathrm{S}}{\mathrm{Vd}} \tag{1}$$

$$\mathrm{Te} = \frac{\mathrm{S}}{\mathrm{CVew}} \tag{2}$$

Therefore, the time *T* for deployment of a virtual machine is as follows.

$$\mathrm{T} = \mathrm{S}\left(\frac{1}{\mathrm{Vd}} + \frac{1}{\mathrm{CVew}}\right) \tag{3}$$

It turns out that deployment time is proportional to archive file size from this formula. Therefore, to arrange a virtual machine efficiently, it is preferable to use a smaller archive file.

The download speed of the disk image which poses a problem can be improved by the arrangement, number, etc. of a server. The server operates as a thread in a service container using the Socket connection generated within VMGridDiskService. Thus the problem of decreased performance with two or more connections can be prevented, if the download process has exclusive use of an FTP server for example. Moreover, a snapshot etc. can be used. The amount of data transmitted at the time of renewal of a virtual machine is reducible by taking the method of offering only the difference of a disk. Shortening of deployment time is expected as a result.

Apart from download speed, three further problems still need to be solved. The first is the processing of the Shutdown operation in VMGridService. The Shutdown operation is realized by calling the Win32API through JNI. Therefore, Shutdown cannot be published if the JNI library is not able to be loaded. Moreover, at the time

of the shutdown of a virtual machine, a WM_CLOSE message is sent to the window of the relevant virtual machine. In this case, the target virtual machine is specified by searching for the title character string from the window generated by the VMware process. However, if the character string acquired as the title matches that of another process, there is a possibility that an unrelated process may be terminated.

The second problem is that the address of the service container of the virtual machine started by VMGridService cannot be obtained. However, if the IP address assigned to the guest OS can be acquired from the host machine side, it can effect a possible solution, possibly using RARP [13].

The third problem exists in the starting situation of the service container on a virtual machine. It is meaningless if the service container itself has not started, even if it is able to obtain the address of the service container. A solution to this problem lies in the ability to start the service container as a daemon. If it is a daemon, it will have started, before a login prompt is shown.

6 Conclusion

In this paper, a grid system was created that performs automatic deployment and execution control of a required virtual machine for the purpose of building a Linux Grid using virtual machines and building the grid environment with a full set that exploits Windows resources. The proposed system enabled the use of a virtual machine by downloading it, and deploying a disk image.

A future task is to make the deployment of a virtual machine image smoother. At present, deployment is monitored manually. If possible, manual intervention should be eliminated by specifying the container start time and enabling it to perform this automatically. Finally, efficiency needs to be improved when more than one connection is made.

References

1. Tanaka, K., Uehara, M., Mori, H.: Parallel Computing of CG using Open Source Grid. Forum on Information Technology (2007)
2. Tanaka, K., Uehara, M., Mori, H.: Parallel Computing of CG using Open Source Windows Grid. In: DPS Workshop (2007)
3. Tanaka, K., Uehara, M., Mori, H.: A Case Study on Linux Grid using Virtual Machine, DPS133 (2007)
4. The Globus Alliance, http://www.globus.org/
5. Center of GRID Research Development, http://www.naregi.org/index_e.html
6. Globus: WSRF - The WS-Resource Framework, http://www.globus.org/wsrf/
7. Amazon Elastic Compute Cloud, http://en.wikipedia.org/wiki/Amazon_Elastic_Compute_Cloud
8. JDK 6 Java Native Interface, http://java.sun.com/javase/6/docs/technotes/guides/jni/index.html
9. OASIS Web Services Notification (WSN) TC, http://www.oasis-open.org/committees/tc_home.php?wg_abbrev=wsn

10. VMware: Virtualization, Virtual Machine & Virtual Server Consolidation - VMware, http://www.vmware.com/
11. Microsoft Virtual PC (2007), http://www.microsoft.com/windows/products/winfamily/virtualpc/default.mspx
12. Xen.org, http://www.xen.org/
13. A Reverse Address Resolution Protocol, http://www.ietf.org/rfc/rfc0903.txt

The Similarity Computing of Documents Based on VSM

Qinglin Guo[1,2]

[1] North China Electric Power University
Department of Computer Science and Technology
102206 Beijing, China
[2] Peking University
Department of Computer Science and Technology
100871 Beijing, China
qlguo88@sohu.com

Abstract. The precision and efficiency of the similarity computing of documents is the foundation and key of other documents processing. In this paper, the DF and TF-IDF algorithms are improved. First, DF's time complexity is linear which suits mass documents processing, but it has the fault that exceptional useful features may be deleted, so we make up that by adding the count of the words at the important places. Second, we rectify the weight of feature by the result of feature selection phase. In this way, we improve the precision of documents similarity without adding much time and space complexity.

Keywords: documents similarity, feature selection, TF-IDF, VSM.

1 Introduction

The way people get information has been completely changed as the popularization of computer and internet and the number of the electronic documents increases on an unprecedented scale. Facing such large quantity information, how to get the information we need quickly is become more and more crucial. If there is not efficacious organizing and extracting information method, people will spend much more time on seeking the information than on learning it that people can not bear. Document similarity is a measure of matching level of two or more documents. The larger the similarity level of two documents is, the higher the similarity level. On the contrary, the similarity level is lower. The efficiency of similarity computing of documents is the foundation and key of Documents Clustering, Information Retrieval, Question Answering System, Document Classing, and so on.

2 VSM

VSM (Vector Space Model) put forward by Salton [1] is one of the models that are widely used and with good effect in information retrieval spheres. Its basic thought is: suppose that the words are Independent, and each document is

M. Takizawa, L. Barolli, and T. Enokido (Eds.): NBiS 2008, LNCS 5186, pp. 142–148, 2008.
© Springer-Verlag Berlin Heidelberg 2008

expressed in a vector, that simplify the complexity relationship of the words and make the model have calculability [2] . In VSM, each document is composed by terms (T_1,T_2,T_3,...,T_n) which are independency, and every term $T_i$has a weight $W_i$which is assigned a value according to the importance. The terms (T_1,T_2,T_3,...,T_n) are regarded as the coordinate axis in N-dimensional coordinate system and the tem (W_1,W_2,W_3,...,W_n) is the corresponding value. In this way, the orthogonal term vector which is decomposed by (T_1,T_2,T_3,...,T_n) composes a document VSM [3].

3 Feature Selection

Feature selection is the process of selecting subset from the original feature set according to some standards. Its task is selecting the features which are useful for the similarity computing of documents, and the features which are selected should be able to express the theme of the document.

3.1 Common Used Algorithms of Feature Selection

Feature selection [4] algorithms are divided into supervised and unsupervised classes [5]. The supervised one includes IG (information gain), CHI, MI (mutual information) and so on, and the unsupervised one is comprised of DF (document frequency), TS (term strength) [6], etc.

(1) IG (information gain)
The IG of a feature is the amount of information of forecasting class if the feature appears in the document [7]. In other words, it's the difference of entropy between the sets without considering any of the features of the document sets and taking into account the feature. The formula is as below:

$$\begin{aligned} IG(t) = &- \sum_{i=1}^{m} P(c_i) \log P(c_i) + P(t_k) \sum_{i=1}^{m} P(c_i|t_k) \log P(c_i|t_k) \\ &+ P(\underline{t}_k) \sum_{i=1}^{m} P(c_i|\underline{t}_k) \log P(c_i|\underline{t}_k) \end{aligned} \tag{1}$$

$P(C_i)$is the probability of the class C_i, $P(t_k)$ is he probability of feature t_k in the training set, $P(C_i|\ t_k\ )$is the probability of document includes t_k when it belongs to classC_i, $P(C_i|\ \underline{t}_k\ )$ is the probability of document without t_k and it belongs toC_i. It will provide important information for classification that whether the term is in the document. We should select a certain number of features on the top order by their IG.

(2) CHI (χ2 Statistics)
χ2 Statistics is the relationship between the feature and the class [8]. The formula is as below:

$$\begin{aligned} \chi^2(t,c) &= \frac{N \times (P(t,c) \times P(\overline{t},\overline{c}) - P(t,\overline{c}) \times P(\overline{t},c))^2}{P(t) \times P(\overline{t}) \times P(c) \times P(\overline{c})} \\ &\approx \frac{N \times (AD-BC)^2}{(A+B) \times (C+D) \times (A+C) \times (B+D)} \end{aligned} \tag{2}$$

$t$is the feature, A is the count of documents which includes t and belongs to class c,B is count of the documents which includes t but does not belong class c, C is the count of documents which belongs to c and doesn't include$t$, D is the count of the documents which doesn't include t and is not belongs to class c, and N is the number of all documents in the training set. We can also compute the average and max CHI of t.

$$\chi^2_{avg}(t) = \sum_{i=1}^{m} P(c_i)\chi^2(t, c_i) \tag{3}$$

$$\chi^2_{\max}(t) = \max_{i=1}^{m} \{\chi^2(t, c_i)\} \tag{4}$$

$M$is the number of classes; the CHI of one feature compares the contribution of a feature to classes, it also indicates the impact of the feature of classification. The larger of CHI is, the more interrelated between the feature and the class.

(3) MI (mutual information)
The formula of the MI of feature t and class$c$ is as below:

$$I(t, c) = \log \frac{P(t, c)}{P(t) \times P(c)} \tag{5}$$

In a set containing $m$classes, the MI of feature t and a , class c has two forms of definition as below:

$$I_{avg}(t) = \sum_{i=1}^{m} P(c_i)I(t, c_i) \tag{6}$$

$$I_{\max}(t) = \max_{i=1}^{m} I(t, c_i) \tag{7}$$

The $I_{avg}(t)$ represents the average MI between t and$c$, and the $I_{max}(t)$ is the max MI of them. The MI tokens the level of correlation between the feature and the class. The MI is higher when a feature only belongs to one class. The MI is 0 when the feature and the class are independent, and the MI of them is a negative when a feature appears in a class hardly. We should select the features which MI is higher.

(4) DF (documents frequency)
DF is the ratio of the count of the documents which includes the feature t_k to the count of the documents which doesn't include t_k. DF supposes that the low frequency feature hardly has efficacious information for classification, or even is the noise point. It is the simplest algorithm of feature selection and its time complex icy is linear, so it's easy to extend to large data sets and the effect is good in practical application [9].

(5) TS (term strength)
We can compute the TS based on the probability of a feature which appears in one of a couple related documents and also in the other one. TS gets the

document pairs which similarity exceeds the threshold value in the training set. d_i and d_j are related and different documents and t is the feature, the formula of TS is as below:

$$\begin{aligned} \mathrm{TS}(t) &= P(t \in \mathrm{d_j} | t \in \mathrm{d_i}), \mathrm{d_i}, \mathrm{d_j} \in \mathrm{D} \\ &\cap \mathrm{sim}(\mathrm{d_i}, \mathrm{d_j}) > \beta \end{aligned} \tag{8}$$

β is the threshold value. The time complexity of TS is $O(N_2)$, N is the number of documents.

3.2 Improved DF Formula

In DF formula, the feature will not be selected if its DF is below a certain threshold or above other one. In this way, some features with much useful information may be deleted. In order to make up the fault, we add n to the frequency Ni of the words which are at the important place, so we can reduce the probability of deleting the useful features. The formula of improved DF is as below:

$$\mathrm{w_i} = \frac{T_i}{T} * \frac{N_i}{N} \tag{9}$$

$W_i$is the weight of a feature i, $T_i$is the number of documents which includes i. N_i is the number of feature i, and N is the number of all features.

4 The Similarity Computing of Documents

Vector Space Model and Set Operations Model are in common used in computing similarity of objects. Because the limitation of the latter, the most used is VSM [10].

4.1 TF-IDF Based on VSM

TF-IDF is efficacious and used commonly in the algorithms based on VSM. It considers the different words' frequency in all documents and the ability to distinguish the documents, and is used in similarity computing of documents widely. In VSM, every vector composed by the feature and its weight represents a document. The similarity of documents can be expressed by the angle or distance between them, and the smaller the angle or distance is, the higher of the documents similarity.

TF represents Term Frequency and IDF is Inverse Document Frequency. The formula is as below:

$$W_{t,d} = TF_{t,d} * IDF_t \tag{10}$$

$W_{t,d}$ is the importance of the feature t in the document c, $TF_{t,d}$ is the count of $t$in document d, IDF_t is like this:

$$IDF_\mathrm{t} = long(\frac{N}{n_t}) \tag{11}$$

N is the number of all documents in set. $n_t$is the frequency of t, that is the number of t in all documents of set. IDF reflects the distribution of the feature in all the set, so it can reflect the ability of distinguishing classes. TF reflects the distribution of a feature in a document. TF-IDF can exclude the words with high frequency but low distinguish, so TF-IDF is an effective algorithm for computing feature's weight.

4.2 Improved TF-IDF

In the normal algorithms of similarity computing of documents, the first step is selecting features terms from the training set, then computing the weight of each feature in the set, and the workflow is separate. In other words, once the features are selected in the stage of feature selection, they have the same weight. But in practical applications, the training set and the application set are always interrelated. Not only DF, but also TF-IDF is based on statistics, so the effect will be better if there are more documents in the set. In order to use such feature, we put forward that: in the stage of computing the weight of features in application set, we take into account the weight of the feature in feature selection stage. The improved TF-IDF is as below:

$$W_{t,d} = \mathrm{W_t} * TF_{t,d} * IDF_t \tag{12}$$

W_t is the weight of feature t in the training set. We can adjust W_t based on specific situation.

4.3 Computation of Documents Similarity

In this paper, we used the algorithm of cosine which is commonly used to compute the similarity of the documents [11]. The formula is as below:

$$\mathrm{sim}\,(T_i, T_j) = \frac{\sum_{t=1}^{n} T_{i,t} * T_{j,t}}{\sqrt{\sum_{t=1}^{n} T_{i,t}*}\sqrt{\sum_{t=1}^{n} T_{j,t}}} \tag{13}$$

T_i is the vector represented documents, $T_{i,t}$ is NO. t vector in the document T_i.

5 Result and Analysis

The Lancaster corpus of mandarin Chinese (LCMC) [12] includes 500 documents, 839009 words, 15 classes and the documents have been segmented and expressed in the form of XML. Also, the words' properties have been tagged. The corpus we used in the experiment includes three classes: The reporting of news, a leading article of news, and the commanding of news. There are 88 documents, 147771 words. We select 350 words as the features and 8 documents to compute their similarity. The result is as table 1 and fig. 1:

$\mathrm{sim}(\mathrm{D}_i,\mathrm{D}_j)$is the similarity of documents $i$and j.

From fig.1 we can see that the improved TF-IDF is closer to the result of export assessment, so it can reflect the similarity of documents better.

Table 1. The Result of Experiment Contrast

Order number	sim(D_i,D_j)	TF-IDF	Improved TF-IDF	expert
1	D1 D2	0.56	0.73	0.78
2	D1 D3	0.47	0.45	0.50
3	D2 D3	0.43	0.29	0.35
4	D4 D5	0.36	0.33	0.37
5	D4 D6	0.43	0.69	0.65
6	D5 D6	0.23	0.25	0.32
7	D7 D8	0.44	0.53	0.60

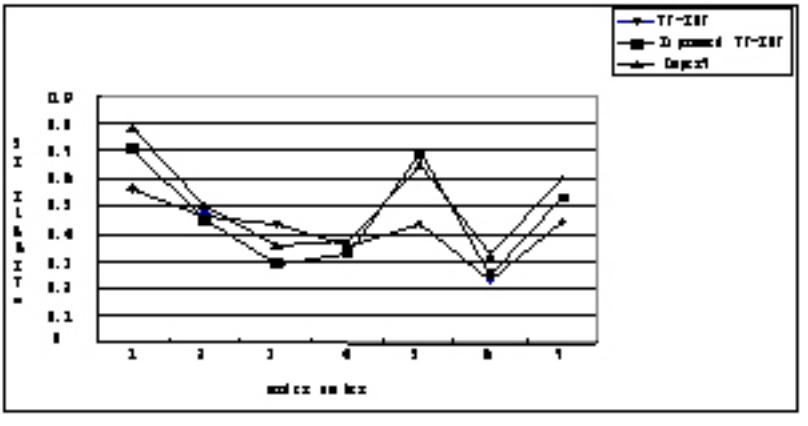


Fig. 1. The result of experiment contrast

6 Conclusions

The similarity computing of documents is the base and key of information mining and is attracting more and more attention. Its efficiency and exactness will influence the effect of the following stages. Not only DF, but also TF-IDF is based on statistics, so the effect will be better if there are more documents in the set. In order to use such feature, we put forward the method: in the stage of computing the weight of features in application set, we take into account the weight of the feature in feature selection stage.So, In this paper, we improved the DF and TF-IDF, and enhanced the exactness of similarity of documents. Experiments do prove that it is feasible to use the method to develop a similarity computing system of documents, which is valuable for further study in more depth.

Further in-depth research should be undertaken in the following areas. First, the methods to analyze sentence with the help of semantic block and sentence model and in-depth analysis of Chinese sentence analyzing theory based on semantic block and sentence model on the basis of semantic web should be explored. Second, construction of large-scale domain ontology should be studied. Third, transplant research outcome into similarity computing system of large-scale documents should be researched.

Acknowledgments

We would like to acknowledge the support from the National Natural Science Foundation of China (90412010, 60573166), Foundation of Post-Doctor in China, the National High Technology Research and Development Program (863 Program in china: 2004AA1Z2450), HP Labs China under "On line course organization".

References

1. Salton, G., McGill, M.G.: Introduction to Modem Information Retrieval. McGraw-Hill, New York (1983)
2. Wang, X.J.: Research on several problems in text retrieval. University of posts & telecommunications, Beijing. Dissertation of the degree of doctor (2006)
3. Song, B.: Study on Chinese text similarity computing based on word segmentation, Tianjin University of Finance & Economics, Tianjin. Dissertation of the degree of master(2006)
4. Dash, M., Liu, H.: Feature selection for classification. International Journal of Intelligent Data Analysis 3, 131–156 (2007)
5. Liu, T.: An evaluation on feature Selection for text clustering. In: Proceedings of ICML 2006, Washington DC (2006)
6. Yang, Y.: Noise reduction in a statistical approach to text categorization. In: Proceedings of ACM SIGIR 2005 (2005)
7. Yang, L., Pedersen, J.O.: A comparative study on feature selection in text categorization. In: Proceedings of 24th International Conference on Machine Learning, pp. 412–420. Morgan Kaufmann, San Francisco (2007)
8. Galavotti, S.F., Simi, M.: Feature selection and negative evidence in automated text categorization. In: Proceedings of KDD 2005, Boston, MA (2005)
9. Yang, L.L.: A class-based feature selection algorithm for test clustering. International J. Computer Engineering and Applications 12, 144–146 (2007)
10. Song, L., Zhang, Z.J.: The study on the Comprehensive Computation of Documents Similarity. International J. Computer Engineering and Applications 1, 1160–1163 (2006)
11. Zhou, Y.X., Wang, G.S., Zhao, H.J.: Text Similarity Computing Based on Hanning Distance. International J. Computer Engineering and Applications 6, 109–116 (2006)
12. The Lancaster corpus of mandarin Chinese (LCMC), http://www.ling.lancs.ac.uk/corplang/lcmc/

Case Study on the Recovery of a Virtual Large-Scale Disk

Erianto Chai, Minoru Uehara, and Hideki Mori

Dept. of Open Information Systems, Toyo University Graduate School, Japan
eriaji@gmail.com, {uehara,mori}@toyonet.toyo.ac.jp

Abstract. With the recent flood of data, one of the major issues is the storage thereof. Although commodity HDDs are now very cheap, appliance storage systems are still relatively expensive. As a result we developed the VLSD (Virtual Large-Scale Disk) toolkit in order to construct large-scale storage using only cheap commodity hardware and software. We also developed a prototype of the large-scale storage system by using the VLSD to collect free disk space on PCs. However, the reliability of this storage depends on the MTTR (Mean Time to Repair). In this paper, we evaluate the MTTR of our prototype and then discuss its efficiency.

Keywords: Virtual Large Storage, Fault Tolerance.

1 Introduction

Recently, the demand for mass storage has increased. In a conventional application, the DBMS is used to share data. However, recently, SOA based applications have increased due to spreading Web services. In such applications, services provided by different providers are interleaved, with the result that they have to use a storage service in order to integrate interfaces as services. As an example, Amazon provides the Simple Storage Service (S3).

Video sites such as YouTube and NikoNikoDoga work as the journalistic complement to mass communication such as TV. For example, in the USA YouTube is used during the presidential elections. On such sites, mass storage is needed to store the large number of videos.

In life logs such as MyLifeBits, a user tries to record every activity in his own life. For this purpose, the system requires at least several terabytes of storage.

Corporations need to store logs for the purpose of internal control. A log which is gathered from all organizations in the corporation can be very large. In order to manage it efficiently, information life cycle based 3 tier storage is usually used.

Demand for these storage services has improved HDD technology with the result that cheap, mass storage disks are now available. However, the HDD used by an appliance file server is 10 times more expensive than a commodity HDD as specific hardware is required. The current cost of such a storage system is unacceptable. We can however, reduce this cost by using commodity hardware and specific software.

We have developed the VLSD (Virtual Large-Scale Disk) toolkit for constructing large-scale storage. A VLSD is independent of platform because it is written in 100%

M. Takizawa, L. Barolli, and T. Enokido (Eds.): NBiS 2008, LNCS 5186, pp. 149–158, 2008.
© Springer-Verlag Berlin Heidelberg 2008

pure Java. Using a VLSD, we constructed a 70TB storage system consisting of 484 PCs, where each PC provides 180GB free disk space. This system realizes an acceptable MTTF using RAID66, which is 2 layered RAID6.

The availability of the storage as a whole is also dependant on the MTTR. For example, as the number of disks increases, the capacity increases and the MTTF (Mean Time to Failure) decreases. Therefore, we need to minimize the MTTR, which comprises disk exchange time and data repair time. Disk exchange time is the time needed to change a faulty disk, whereas data repair time is the time needed to restore data. In order to minimize the MTTR, both times should be reduced. However, disk exchange time can be ignored if there is a hot spare in the system.

In this paper, we evaluate the availability of our system constructed as a VLSD by measuring repair time.

This paper is organized as follows. Section 2 summarizes related works. Section 3 describes a VLSD, and Section 4 explains how repairs are carried out in a VLSD. Section 5 evaluates the availability, while Section 6 highlights some considerations. Finally, we present our conclusions.

2 Related Works

2.1 RAID

RAID (Redundant Arrays of Inexpensive Disks) [1][2] is usually used to improve the reliability of large-scale storage. RAID allows data to be striped onto several disks in a distributed fashion with data redundancy and thereby achieves both performance and fault tolerance. RAID is classified into 6 levels according to how the redundant data is striped.

RAID0 stripes data without redundancy. It can expand the capacity of storage and can also access data quickly. RAID1 is used to mirror data. It cannot expand the capacity of storage but increases the reliability thereof. RAID2 stripes data with error correcting code based on redundancy. RAID2 has no advantage over parity based RAID and is therefore not usually used. RAID3 stripes data by byte or bit. It corrects errors using parity data, which is usually stored on a separate disk. RAID4 stripes data by block and can also store parity data on a separate disk. This may cause a bottleneck. RAID5 also stripes data by block, but stores parity data on distributed disks, thus avoiding a bottleneck. In a RAID5 system, a read/write operation may access a functioning disk instead of a faulty disk. In this case, two disks fail. RAID5 cannot recover when more than two disks fail. Thus, RAID6 is often employed instead of RAID5. RAID6 [5] is more reliable than RAID5 because it is resilient to 2 failures and can recover data in the case of 2 failed disks. RAID6 uses two parities: parity P is generated by xor-ing data on all disks, while parity Q is more complex. Q is computed using GF (Galois Field) operators. A GF is a set of finite elements; for example GF(2^8) has 2^8 elements, 0,.., 2^8-1. When 2 disks fail in RAID6, the different combinations are classified as follows: P block + Q block, P block + data block, Q block + data block, and 2 data blocks. RAID6 can recover data in all these cases.

RAID arrays are classified into software (SW) and hardware (HW) systems. HW RAID is predominant, while SW RAID is considered to have a lower performance

than HW RAID and adds load to the CPU. However, the latest CPUs have higher performance than a HW RAID controller. CPU control of SW RAID is effective in machines such as file servers that can concentrate on the RAID operation. Moreover, SW RAID is the only option for RAID across a network.

2.2 Reliability of RAID

The calculation of the average time until the system crashes is often used as a measure of the reliability of RAID. According to [1], the MTTF of RAID1 can be expressed by the following formula

$$\frac{MTTF_{disk}^{2}}{N \times MTTR_{disk}}$$

where N = 2. When RAID1 is generalized, the formula becomes

$$\frac{MTTF_{disk}^{N}}{N! \times MTTR_{disk}^{N-1}}$$

The MTTF of RAID3, RAID4 and RAID5 is given by

$$\frac{MTTF_{disk}^{2}}{N \times (G-1) \times MTTR_{disk}}$$

where $MTTF_{disk}$ is the MTTF of one disk, N is the number of disks in the disk array, G is the number of data disks in the group not contained on the parity disk, and $MTTR_{disk}$ is the MTTR of one disk. The MTTF of RAID6 is calculated as follows.

$$\frac{MTTF_{disk}^{3}}{N \times (G-1) \times (G-2) \times MTTR_{disk}^{2}}$$

The performance and reliability can be raised by replacing single RAID with multiple RAID (hierarchy RAID). For example, performance will be increased if RAID0 is changed to a higher ranking layer. Reliability is improved when changing to RAID1, RAID5 or RAID6 in the subordinate position layer. In this research, as reliability is considered important, RAID55 and RAID66 are used. The MTTF of RAID55 can be expressed by the following formula.

$$\frac{MTTF_{RAID5\_disk}^{2}}{N \times (G-1) \times MTTR_{RAID5\_disk}}$$

The MTTF of RAID66 can be calculated as following.

$$\frac{MTTF_{RAID6\_disk}^{3}}{N \times (G-1) \times (G-2) \times MTTR_{RAID6\_disk}^{2}}$$

$MTTFRAID5_{\_disk}$ and $MTTFRAID6_{\_disk}$ are the MTTF of the $RAID5_{\_disk}$ and $RAID6_{\_disk}$, respectively.

In terms of reliability, the MTTR also plays a significant role. The MTTF decreases as the MTTR increases. Therefore, it is necessary to reduce the MTTR in order to increase the MTTF, which is the average time to a system crash.

3 VLSD

VLSD is a toolkit for constructing large-scale storage systems. It is written in pure Java and can therefore be used on any platform on which Java runs, such as Windows, Linux, etc. VLSD can even be used if there is no native NBD server in the OS, as the toolkit includes software for RAID and NBD. VLSD consists of typical RAID classes and NBD client/server software, thus allowing us to combine any RAID and NBD devices with one another.

3.1 Class of VLSD

From a software perspective, VLSD [3] consists of the following classes.

NBDServer: The NBD server is called by the client, and offers empty capacity to the storage system as a virtual disk file. The OS of the client can be either Windows or Linux. Since the NBD server is mounted using Java, it is platform independent. Moreover, two or more disks may be added to the array by a single client. FAT32 may also be used despite its 4GB maximum file size restriction. Because a virtual disk created using 120GB drives cannot be used as a single file, a virtual disk is created by bundling two or more files and combining them with RAID0 or JBOD as described later.

DiskServer: Is an interface to the disk server.

DiskServerImpl: Is an implementation of the disk server, and a remote disk using RMI.

Disk: Is an interface to a virtual disk.

AbstractDisk: Is a class of an abstract virtual disk that defines the constants and methods that are used in low order classes.

DiskArray: RAID1 is mounted by the base class of the disk wrapper that consists of two or more disks.

RAID: The base class for RAID. Mounting of RAID1 is inherited from DiskArray.

SingleDisk: Is the base class of the disk wrapper which consists of a single disk.

PagedDisk: Is a wrapper for an arbitrary disk accessed via disk paging. For the wrapped disk, read/write is per page only. The fraction on each page is disregarded.

VariableDisk: Is a changeable capacity disk that does not direct resources beforehand, and having resources secured dynamically if necessary.

RemoteDisk: Is a remote disk, where the disk server is accessed.

RAID0: RAID0 is used to increase capacity. It differs from the below-mentioned JBOD in that striping is performed. Though the performance is good, adding

additional disks only adds capacity equal to that of the smallest drive in the array. For example, striping drives of 100GB, 120 GB, and 160 GB results in 300GB only – 100GB x 3. It is better to use JBOD when aiming purely to increase capacity, although RAID0 can be expected to improve performance. In some file systems, the super block which manages i-nodes concentrates disk activity in a specific area. In such a case, striping has the effect of distributing this load. RAID0 performs per byte striping.

RAID 5: Parity is distributed and stored on each disk. The disk where parity is stored is different for each block.

RAID 6: The RAID6 class allows generation of the GF table, block unit striping, and distributed parity to be performed.

JBOD: This class does not provide redundancy like RAID0 and is used to increase capacity only. Because striping is not performed, capacity is simply summed. For example, the total capacity is 380GB if 100GB, 120 GB and 160 GB drives are connected. JBOD, though, lacks the load-balancing effect of RAID0 as described above. Since cashing effectively creates a certain amount of scale, RAID0 performance may be superior.

3.2 Trial Production of the Large-Scale Storage Using VLSD

In this section, we discuss the design of a 70TB distributed storage system in the environment described in Section 1. This system comprises 484 client PCs each of which provides 180GB free disk space. The total free space is 85TB. However, a client PC may be inadvertently shut down, making the distributed storage system no more reliable than a conventional file server. We overcame this problem by using hierarchical RAID, RAID66, which is two layered RAID6 (Fig. 1).

This system includes a 64 bit file server and the disk server as shown in Fig. 2. A virtual disk that consists of OS's such as Linux and Windows for the disk servers provides read/write access to the disk via the Java RMI. The file server connects with the prepared disk and constructs RAID66. RAID66 is RAID6 on two hierarchies. The NBD Server waits for access by an NBD Client. After the NBD Client is started, it is formatted with XFS. Windows clients access the file server through Samba (NFS is used for Linux).

Since the NBD protocol lacks security, applying it on a network is dangerous. However, since NBD is used in our system for inter-process communication only, it can be safely applied. Actual communication between C/S is realized by a protocol based on RMI with the relevant security considerations implemented.

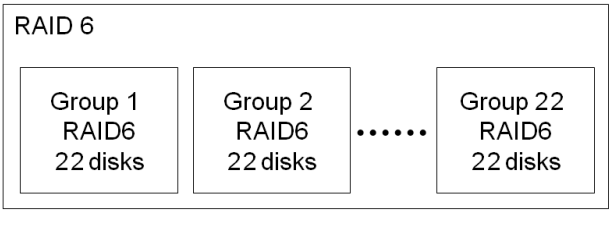


Fig. 1. Composition of RAID66

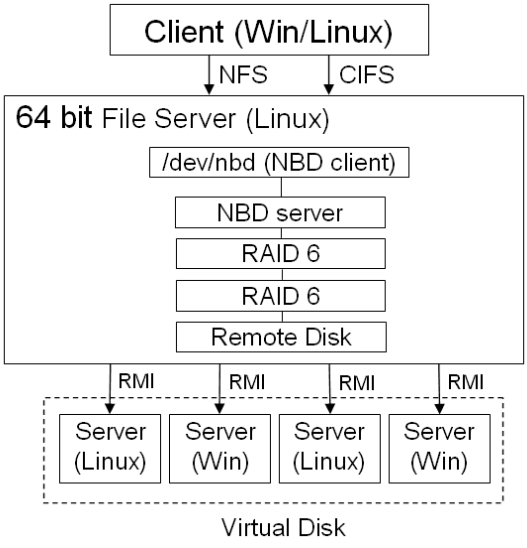


Fig. 2. System Overview

4 Repair

With RAID, even if failure occurs, the system can continue operating. RAID5 tolerates one disk failure, while RAID6 tolerates two. The state where redundancy is lost from RAID is called the degeneration mode. However, the performance at the time of failure may be lower than that normally. We evaluated the performance of the VLSD at the time of failure along the lines of [4]. As a result, we were able to confirm that the level was useable albeit with a decrease in speed.

As described in Section 2.1, the MTTF of RAID is dependent on both the MTTF and MTTR of a disk. Since our large-scale storage is based on hierarchical RAID, the MTTF is similarly expressed by the MTTF and MTTR of the disk.

MTTR is defined as the time from the disk crash until the system can be used normally again. RAID does not stop in the degenerate mode, but it will stop if an additional disk crashes.

The method of restoring the disk depends on the cause of the failure. If the cause of failure is not failure of the disk (for example, when a PC is shutdown), the disk can be recovered easily when the PC is rebooted. The time to reboot a PC is shorter than the time generally needed to restore a disk .This feature is especially noteworthy in the case of a mass disk. However, after the restart, it is necessary to apply an out of order difference for the period.

Next, when the cause of failure is failure of the disk itself, a hot spare is assigned immediately and restoration begins. The hot spare can be either a global or a partial hot spare. A global hot spare manages availability of the entire system. For example, in the trial system consisting of 500 PCs, only 484 (=22*22) PCs are actually used. Therefore, 16 global hot spares are available. A partial hot spare manages availability of the same PC or the same RAID. Although a global hot spare has large capacity, it

is slower because of network delays. The partial hot spare has faster access, although capacity is smaller.

Restoration is performed as follows. After swapping the broken disk for a spare disk, only the contents of the disk that failed within RAID are read and simply rewritten. Thus, data can be restored from a trouble-free disk and the data can be written on the spare disk. In VLSD, a RAID class reverts by carrying out sequential access of the block corresponding to the failed disk.

The disk being restored has reading and writing access at the time of data restoration. There are three separate cases, namely read/write access to the block that has been restored, read/write access to the block being restored and read/write access to the non-restored block.

Read/write can be permitted when there is read/write access to the restored block of the spare disk. This is because it is an area where the recovery of the data has been completed on the restored block. It is necessary to delay this until read/write has been restored when there is read/write on the block being restored. It is considered block failure without read/write when there is read/write on the non-restored block. Data can then be restored using the parity data.

The method of restoring data is different depending on the RAID level. For disk restoration with RAID1, only a disk copy is done. In the restoration of RAID4 and RAID5, data can be restored by calculating the exclusive-OR of the stripe block data of all disks other than the broken disk. We consider separately a one disk failure and a two disk failure in the restoration of RAID6. In a one disk failure, the same process as in RAID5 is carried out at the time of failure of a P block or a data block. If a Q block fails, a Q block is generated. The restoration at the time of a two disk failure is divided into four cases as described in Section 2. A P Block and Q Block failure should just generate a P Block and Q Block at the time of failure. When a P block and a data block fail, the P block is restored after restoring the data block from a trouble-free data block and Q block. When a Q block and a data block fail, the Q block is restored after restoring the data block from a trouble-free data block and P block. When two data blocks fail, the two data blocks can be restored by solving the simultaneous equations, derived from calculating the exclusive-OR of the P blocks and a trouble-free data block, and the GF multiplication from the Q blocks and a trouble-free data block.

This algorithm can be parallelized easily. If a block which is different by two or more threads is restored, the restoration process can be shortened. The degree by which the process is shortened is dependent on the execution environment and mounting.

5 Evaluation

Here, recovery time is measured and the reliability of the whole RAID is evaluated.

5.1 Experiment Environment

The experimental environment is as follows:

CPU: AMD Athlone(tm) 64 X2 Dual Core 3800+
RAM: 2GB
OS: Windows XP Professional x64

5.2 MTTF of HDD

The target large-scale storage, 320GB of highly cost-effective HDD is used. Table 1 shows the MTTF of a typical product. The averages are 850,000 hours. Therefore, there is the possibility of failure at a rate of one every 850,000/484=1756 h or every 73 days when 484 average HDDs are used.

Table 1. MTTF of HDD products

Maker	Product	MTTF[10^3h]
Seagate	ST3320620AS	700
IBM	HDT725032VLA360	1000
Average		850

5.3 Evaluation of Restoration Time

The results of the evaluation of recovery time for RAID5 and RAID6 are presented in this section.

Fig. 3 shows the restoration time for RAID5 and RAID6 constructed with four virtual disks.

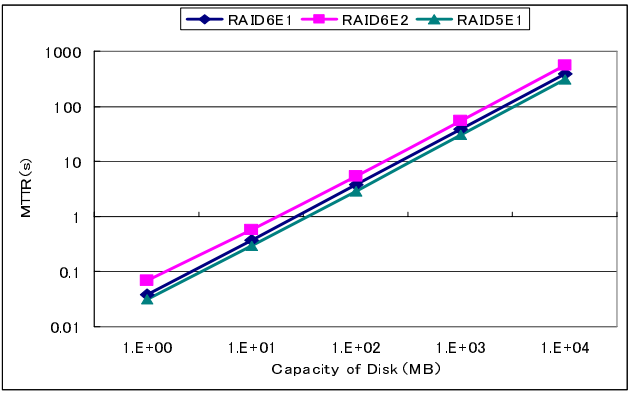


Fig. 3. System Overview

RAID5E1 and RAID6E1 are RAID5 and RAID6, respectively with one disk failure and RAID6E2 is RAID6 with two failed disks. The capacity of disks in RAID6E1, RAID6E2 and RAID5E1 is proportional to the MTTR as shown in the graph. The order of the recovery time is RAID6E2 > RAID6E1 > RAID5E1 in the ratio 3.1: 1.7: 1.

5.4 MTTR by Parallel Restoration

The restoration operation can be done in parallel. When restoring with N parallel processors, the Block number is classified according to mod N or 1/N.

Table 2. Evaluation of parallel restoration

N	1	2	4
RAID5E1[h]	8.97	7.54	8.44
RAID6E1[h]	15.35	14.93	19.48
RAID6E2[h]	35.68	23.82	32.19

Even if N is larger than necessary, the outcome was not expected. The results in Table 2 were obtained with N=1, 2, and 4.

In this experiment, the time for a dual core with 2 threads is optimal. The reason for this is that the I/O serves as a bottleneck for the CPU.

5.5 Evaluation of Reliability

The MTTF of the entire hierarchy RAID is calculated using measured MTTF and MTTR. Hierarchy RAID consists of a combination of RAID1, 5 and 6.

Table 3. MTTF and capacity efficiency of two level RAID

RAID level	MTTF[h]	Capacity Efficiency[%]
RAID11	4.25×10^{2463}	0.2
RAID15	1.7×10^{219}	4.3
RAID16	2.2×10^{328}	4.1
RAID51	4.19×10^{182}	4.3
RAID55	5×10^{15}	91.1
RAID56	1.1×10^{23}	86.8
RAID61	4.78×10^{272}	4.1
RAID65	7.8×10^{23}	86.8
RAID66	2.2×10^{35}	82.6

In Table 3, the total number of disks is equal to 484. Layers 1 and 2 each contain 22 disks. RAID3 and 4 are omitted because they are the same as RAID5. Moreover, in RAID1 it was assumed that the content was copied N times.

RAID1 should not be used, because the capacity efficiency decreases significantly. To achieve the largest capacity and highest MTTF, RAID66 is the best.

6 Summary

This paper described the restoration function and performance of VLSD. VLSD is a toolkit which builds mass storage using available PCs. According to this evaluation, without additional load, a single 180GB disk can be restored in 23.82 hours. The MTBF of the 70TB experimental system with RAID66 and consisting of 484 PCs is the worst at 2.2×10^{35} hours. This can be considered to be usable.

The evaluation was done using FileDisk of which the size is physical capacity. The size of FileDisk does not change at the time of execution and a large capacity is needed from the beginning. However, it is better to change the size if necessary during the actual operation. Such a changeable capacity disk is achieved with VariableDisk. However, because VariableDisk contains an unused area, the blocks are not consecutive. It is thus necessary to repeat this study for discontinuous blocks in the future.

Moreover, it is necessary to evaluate the allocation policy in the disk exchange. For example, if priority is given to performance at the time of execution, a local hot spare is preferable. If priority is given to decreasing the exchange time, a global hot spare is preferable. It is thus necessary to evaluate the performance of each.

In this study, the estimation used one machine. However, an actual system consists of two or more machines. Performance still needs to be investigated, when a network of two or more machines is used.

References

1. Patterson, D.A., Gibson, G., Katz, R.H.: A Case for Redundant Arrays of Inexpensive Disks (RAID). ACM SIGMOD, 109–116 (1988)
2. Chen, P.M., Lee, E.K., Gibson, G.A., Katz, R.H., Patterson, D.A.: RAID: High-Performance, Reliable Secondary Storage. ACM Computing Surveys 26(2), 145–185 (1994)
3. Chai, E., Uehara, M., Mori, H., Sato, N.: Virtual Large-Scale Disk System for PC-Room. In: Enokido, T., Barolli, L., Takizawa, M. (eds.) NBiS 2007. LNCS, vol. 4658, pp. 476–485. Springer, Heidelberg (2007)
4. Chai, E., Uehara, M., Mori, H.: Evaluating Performance and Fault Tolerance in a Virtual Large-Scale Disk, AINA. In: 22nd International Conference on Advanced Information Networking and Applications (AINA 2008), pp. 926–933 (2008)
5. Intelligent RAID6 Theory Overview and Implementation, `http://download.intel.com/design/storage/papers/30812202.pdf`

Constant-Width Zones Broadcast Algorithm in Mobile Ad-Hoc Networks

D. Liarokapis, A. Shahrabi, and C. Raeburn

School of Engineering and Computing
Glasgow Caledonian University
Glasgow G4 0BA, UK
{Dimitrios.Liarokapis,A.Shahrabi,C.Raeburn}@gcal.ac.uk

Abstract. Broadcast operation is perhaps the most fundamental services utilized frequently by other communication mechanisms in Mobile Ad-hoc Networks (MANETs).Supporting efficient broadcast operation is therefore very crucial for such networks. A novel distance-based broadcast algorithm, called Constant-Width Zones (CWZ), is proposed in this paper. CWZ can effectively alleviate the redundant rebroadcast (overlaying) problem by defining a constant upper limit for the width of all rebroadcast zones and, consequently, reducing the number of forwarding hosts. The results of our simulation-based performance study show that the proposed CWZ algorithm is able to rebroadcast packets more effectively in order to achieve higher reachability while reducing the latency especially for heavy host density networks.

1 Introduction

Mobile Ad Hoc Networks (MANETs) are becoming ever more popular, as they provide wireless connectivity between multiple users without relying on any fixed infrastructure [1]. Packets travel through the network hopping from node to node, in order to reach their final destination. As long as there is a path (hop by hop) from the sender to the receiver, communication is possible [2].

In MANETs, broadcast operation is perhaps the most fundamental services utilized frequently by other communication mechanisms such as routing protocols. Due to the dynamic nature of such networks, constant broadcast structures such as minimal spanning tree is no longer suitable to support broadcasting in MANETs. Simple Flooding (SF) is a basic approach to broadcasting without global information; in which a broadcast packet is forwarded exactly once by every node in the network. However, due to the broadcast nature of this environment, redundant transmissions in SF may cause the broadcast storm problem [3], in which redundant packets cause contention and collision.

Over the past few years, many algorithms have been proposed to solve this problem [4][5][6][7][8][9][10][11] with some works focusing on the reduction of rebroadcast yet neglecting the broadcasting latency; while other investigations have focused on reducing latency yet have paid little attention to rebroadcast reduction. In this paper, we propose CWZ (Constant-Width Zones) broadcast algorithm; CWZ can

M. Takizawa, L. Barolli, and T. Enokido (Eds.): NBiS 2008, LNCS 5186, pp. 159–168, 2008.
© Springer-Verlag Berlin Heidelberg 2008

effectively alleviate the redundant rebroadcast problem by defining a constant upper limit for the width of all rebroadcast zones and, consequently, reducing the number of forwarding hosts. The efficiency and effectiveness of CWZ is investigated through simulation comparing the results with those of simple flooding and distance-based broadcast algorithms.

The rest of the paper is organized as follows. In Section 2 we overview the related works. In Section 3 we explain the creation of broadcast zones and their effect in the distance-based broadcast scheme. Our proposed broadcast algorithm is introduced in Section 4. The performance study if presented in Section 5. Finally, we make concluding remarks in Section 6.

2 Related Works

The main concern of almost every broadcast technique is to set up the most efficient criteria to be followed by every node in order to decide whether to rebroadcast a packet or not [12], [13], [14], [15]. Most of the proposed solutions include compromises in favor of one factor or another [4], [5], [6], [7], [8], [9], [10], [11].

2.1 Existing Approaches

A rather comprehensive classification of broadcast algorithms has been presented in [12]. According to his work, broadcast protocols are divided into four families: Simple Flooding, Probability Based Methods, Area Based Methods and Neighbor Knowledge Methods. Simple Flooding (SF) is using a simple algorithm. The source node (broadcast initiator) sends the packet to all its neighboring nodes (nodes inside the transmission radius), which in turn every node rebroadcasts the packet to all of its neighboring nodes. Redundant rebroadcast, contention and collisions are the three main drawbacks of simple flooding and the main reason for the appearance of the broadcast storm problem described by Ni [3].

In Probability Based Methods, the rebroadcasting decision is made by every receiver of the broadcast message using a probabilistic model. Nodes in Area Based Methods family protocol require specific information about the topology of the network, in order to make the rebroadcasting decision. The distance-based Scheme calculates the distance between sender and receiver and using a threshold to determine whether to rebroadcast the packet [3].

Finally, in Neighbor Based Methods, nodes periodically exchange *Hello* messages with their neighboring nodes. Using the information from *Hello* messages, nodes have a clear picture of the network within their transmission radius. Different methods use this information in different ways in order to make the rebroadcasting decision [12].

2.2 The Problem with the Distance-Based Scheme

In this section, we explain the exponentially growing widths of zones in distance-based broadcast algorithm and how this can exacerbate the redundant rebroadcast problem.

2.2.1 Broadcast Zones

Let us consider a flat (2 dimensional) network topology, where a node initiates a broadcast. Figure 1 shows part of that network, focusing on the initiator's transmission radius. The black node initiates and sends a broadcast message. The outer circle represents the transmission range of that node. This is considered to be the first "wave" of the broadcast process. All the nodes covered by the transmission range will receive the message, but only some of the nodes inside the grey zone will retransmit the broadcast message according to the distance-based algorithm. Only the grey nodes represented in Figure 2 (placed on or just inside the inner circle) are eligible to retransmit the message. The extra area that will be covered is shown by the outer circle. It should be noted that the outer boundary of the covered area is not circular. This is a simplification to decrease the complexity of the figures.

In Figure 3, all the inside arcs represent the transmission ranges of individual nodes. This is actually the shape of the second wave's reaching point. The outer thick black arc is what we assume to be the second reaching point. Figure 3 also shows that our assumed reaching point produces a coverage area that is greater than the actual area, but the increase is in fact very small. Using the same method we accept that all the following waves will be homocentric circles with their centre being the initiator of the broadcast, as shown in Figure 4.

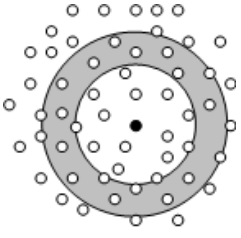


Fig. 1. Initiation of a broadcast and its first wave

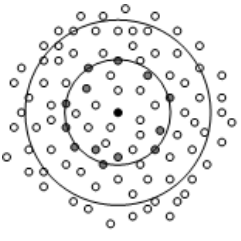


Fig. 2. Second wave

Fig. 3. Outer boundary of broadcast zone

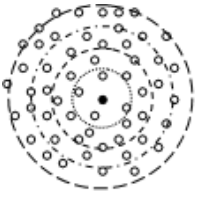


Fig. 4. Reaching points of the first 4 broadcast waves

2.2.2 Specifying the Problem

The distance-based scheme (DB) makes decision on which of the nodes that receive the broadcast message is considered for retransmission after the first wave. Because this approach uses a distance threshold, all the nodes that are covered by the zone formed from the initiator's transmission range (outer circle) and the distance threshold circle (inner circle) are eligible to retransmit (these are the grey nodes) as shown in

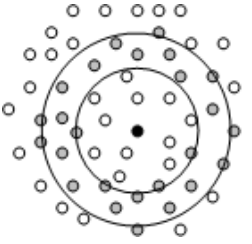


Fig. 5. DB scheme, first broadcast zone

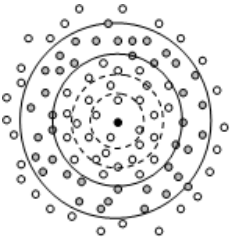


Fig. 6. DB scheme, first and second broadcast zone

Figure 5. The width of the zone (Z_1) is $W_1 = R_{TR} - R_{THR}$, where R_{TR} is the transmission range and R_{THR} is the distance threshold.

The worst case scenario dictates that some or all of the grey nodes placed closer to the inner circle (called *C* nodes) will rebroadcast. Some or all of the grey nodes placed closer to the outer circle (called *F* nodes) will also rebroadcast.

After all nodes, have made their re-broadcasting decision, the formation of the second zone (second wave) can be observed. This zone describes the new set of nodes that are eligible to rebroadcast. This zone is formed from two circles. The outer circle is formed from the transmission ranges of the F nodes. The inner circle is formed from the distance threshold of the *C* nodes.

Figure 6 presents the formation of both zones (first and second wave). The width of the new zone (Z_2) is $W_2 = 2R_{TR} - 2R_{THR} = 2(R_{TR} - R_{THR}) = 2W_1$. There is also one more zone ($Z_{1,2}$) to be considered, that is the zone formed as a consequence of Z_1 and Z_2. $Z_{1,2}$ has a width of $W_{1,2} = 2R_{THR} - R_{TR}$. It is obvious that when the distance threshold is chosen to be less than half of the transmission range, Z_1 and Z_2 will partially cover each other. We will show that whatever the distance threshold is chosen to be, there will be some point where two sequential zones will cover each other when the worst case scenario is considered.

The zone that covers the nodes eligible to rebroadcast after *x* waves, is formed by two circles. The outer circle's radius is increased by R_{TR} for each wave and the inner circle's radius is increased by R_{THR}. Therefore the distance between the current wave's zone and the previous wave's zone is given by the following equation:

$$D = (x+1)R_{THR} - xR_{TR},\ x = 1, 2, 3, \ldots \tag{1}$$

The variable *x* represents the previous wave's sequence number. For the second wave ($x = 1$), the distance between Z_1 and Z_2 is $D = 2R_{THR} - R_T$. For the third wave ($x = 2$), $D = 3R_{THR} - 2R_T$. In order for two sequential zones not to cover each other the following must always be true:

$$D \geq 0 \Leftrightarrow (x+1)R_{THR} - xR_{TR} \geq 0 \Leftrightarrow (x+1)R_{THR} \geq xR_{TR} \Leftrightarrow R_{THR}/R_{TR} \geq x/(x+1) \tag{2}$$

The following is also true:

$$\tfrac{1}{2} \leq x/(x+1) < 1 \text{ and } 0 < R_{THR}/R_{TR} < 1 \tag{3}$$

Because of this, and given the fact that the distance threshold R_{THR} is constant for a specific network, we conclude that there is always a value for *x* that makes (2) to be false.

3 CWZ: A Novel Approach

A major goal when trying to create a new broadcast protocol is to minimize the number of nodes that are eligible to rebroadcast and at the same time maximize the coverage area of the network. We have previously shown how the width of the broadcast zones formed from each wave is increasing. This caused the variable, nodes eligible to rebroadcast per square unit, to increase as well. What we actually require, is for this variable to be constant and one way to achieve that is to have a constant or an upper limit of width for every broadcast zone. This is what our approach actually achieves. Next, we will describe the detailed operation of our CWZ approach.

Let us consider the scenario of a large and highly dense mobile ad hoc network. All the nodes share the same transmission range and they are all initially configured with the same distance threshold R_{THR} (R_{THR} is only used from the broadcast initiator). One of the nodes initiates and sends a broadcast message. When a node receives the message, it executes the distance-based algorithm. Finally, if the node decides to rebroadcast, then it calculates a new distance threshold which replaces the old one inside the message. How do the nodes calculate the new threshold?

The nodes calculate the new threshold by using the following formula:

$$R_{THR(NEW)} = R_T + R_{THR} - d \tag{4}$$

Where,

$$R_{THR} \leq d \leq R_T \text{ and } 0 \leq R_{THR} \leq R_T \tag{5}$$

d is the distance between sender and receiver, R_{THR} is the previous distance threshold, R_T is the transmission range and $R_{THR(NEW)}$ is the new distance threshold. Figure 7 shows how the value of the new threshold changes according to the distance.

In Figure 8 node B receives the broadcast message from node A (broadcast initiator). Node B runs the distance-based algorithm. If it finally decides to rebroadcast, it then calculates the new threshold using equation (4) to be $R_{THR(NEW)} = R_T$. The new threshold equals with the transmission range. Only node E which is placed on c_4 is then eligible to rebroadcast. Random node N then calculates that $R_{THR(NEW)} = R_{THR} + R_T - d$. Only the nodes placed between c_4 and c_5 are then eligible to rebroadcast. Note that c_5 is the transmission range of N. Node C then calculates

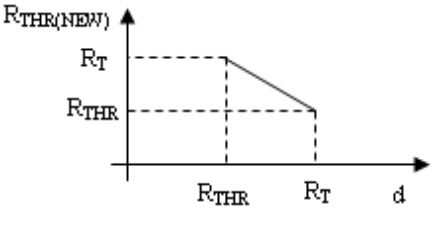


Fig. 7. The new threshold as a function of d

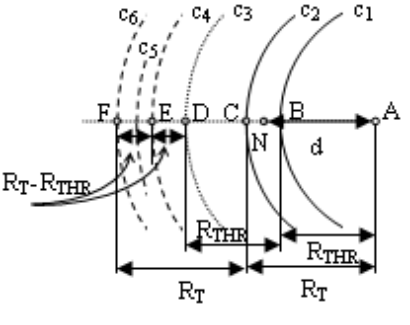


Fig. 8. CWZ – first and second wave

that $R_{THR(NEW)} = R_{THR}$. Only the nodes placed between c_4 and c_6 are then eligible to rebroadcast. Note that c_6 is the transmission range of C.

Whichever node rebroadcasts the message (B, C or N) the inner circle of the next broadcast zone is always c_4. The final set of nodes S that are eligible to rebroadcast is $S = S_1 \cup S_2 \cup \ldots \cup S_n$ where S_n is the set of nodes eligible to rebroadcast generated by a random node between c_1 and c_2. So S is the set of all the nodes placed between c_4 and c_6 which form the second broadcast zone. It is obvious that the widths of the two zones are the same and equal with $W = R_T - R_{THR}$.

The same sequence of events takes place for all the other waves as well. Again, the width of the newly formed broadcast zone is $W = R_T - R_{THR}$.

4 Performance Evaluation

We implemented Simple Flooding (FL), distance-based Scheme (DB) and CWZ using the network simulator NS2.30. We have used the NS2 code for FL and DB provided by [9][12].

4.1 Simulation Set-Up and Parameters

Node mobility is simulated using mobility files that are generated by the NS2 mobility generation feature *setdest*. The network area is of fixed size 500x500m. The mobility files are created with a zero mobility pause time.

We follow the *TRIALS* technique, where a *TRIAL* is a specific set of different network parameters. We are using a combination of number of nodes, speed and broadcast generation rate for each trial [9], [12]. Each scenario is restricted to the transmission of broadcast traffic only. We divide our simulations into two groups of trials in order to record the effects of different broadcast generation rates on the same combination of number of nodes and speed. The nodes are using a transmission range of 100m and a distance threshold of 50m. In order to avoid anomalies, we run three simulations for every trial using 3 different mobility files. The final results are created as an average of the three simulations. Each simulation has duration of 100secs. The trials that we used for our experiments are presented in Table 1.

We consider the following performance metrics:

- Reachability – the percentage of nodes that successfully receive the broadcast message.
- Speed – the reverse metric of delay, we use speed instead of delay because simple flooding appears to be extremely slow in comparison with DB and CWZ and the graphs generated cannot demonstrate the difference between the last two.
- Quality – a network metric that combines reachability and speed in a composite result as a percentage.

The definition of Quality as a metric is as follows:

$$Q(\%) = (A_1 x M_1 + A_2 x M_2 + \ldots + A_n x M_n) x 100 \quad (6)$$

Where $A_1 + A_2 + \ldots + A_n = 1$ and M_1, M_2, …, M_n are the normalized values of regular metrics like Reachability, Delay, Speed, Retransmission Attempts e.t.c.

Table 1. Network Trials Group 1

Number of nodes	50	100	150	200
Speed (m/s)	5	10	15	20
BGR (pack/s)	1	5	10	15
	TRIAL 1	**TRIAL 2**	**TRIAL 3**	**TRIAL 4**

Table 2. Network Trials Group 2

Number of nodes	50	100	150	200
Speed (m/s)	5	10	15	20
BGR (pack/s)	5	10	15	20
	TRIAL 1	**TRIAL 2**	**TRIAL 3**	**TRIAL 4**

Quality is a metric that produces a general result for the comparison of different protocols and schemes for the same scenario. The researcher will decide on the *A* values, depending on which metric he considers to be more important. In our experiments we use $M_1 = Reachability$, $M_2 = Speed$ and $A_1 = A_2 = 0.5$, meaning that we consider reachability and speed to be of equal importance.

4.2 Simulation Results

In both group 1 and group 2 of our simulations we use 4 different trials. In each trial we increase the values of all three network parameters as shown in Table 1. The difference between the two groups is that in the second one we use a higher broadcast generation rate for each trial in comparison with the first group. Our intention is to demonstrate that FL, DB and CWZ react in a similar way overall, independent of the broadcast generation rate.In Figure 9, FL appears to have a higher reachability than DB and CWZ up to trial 3. When the network becomes extremely dense, as in trial 4 with 200 nodes and high speed and broadcast rate, FL collapses. This is the point where we encounter the *broadcast storm problem* described by Ni. FL and CWZ follow the same pattern in terms of reachability, with CWZ being slightly better, as they make use of a similar algorithm.

Figure 10 shows again the collapse of FL in trial 4. CWZ appears to be better than FL and DB overall. The difference in speed between DB and CWZ increases steadily from trial 1 to trial 4, reaching a top value of approximately 33% in the latter.

Figure 11 reflects the overall performance of the three schemes. Figure 9 showed that FL has a higher percentage of reachability through trials 1, 2 and 3 but according to Figure 10 is the slowest. The question in regards to the first three trials is about the overall performance of FL in comparison with the other two methods. If we consider reachability and speed to be of the same importance (given that $A_1 = A_2 = 0.5$), Figure 11 shows that FL provides poorer performance than DB and CWZ, with the latter being better in all trials.

In the second group of trials we use higher broadcast generation rates in order to compare the three schemes in more extreme conditions.

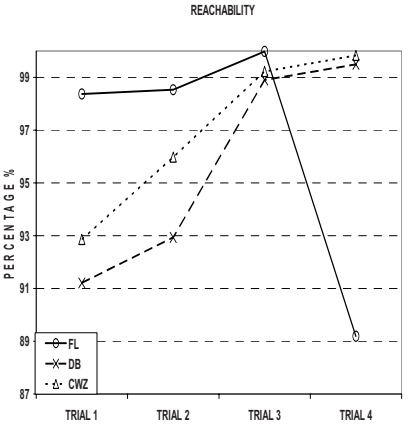


Fig. 9. Group 1 – Reachability

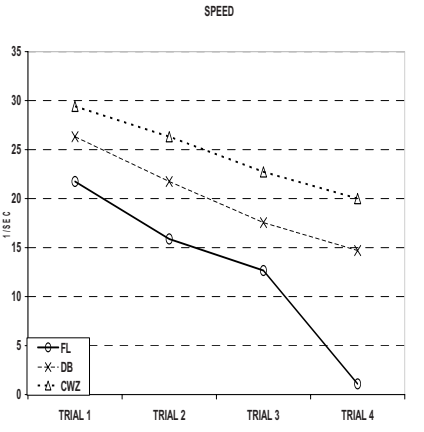


Fig. 10. Group 1 – Speed

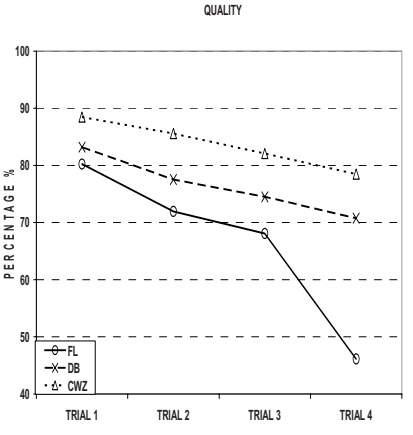


Fig. 11. Group 1 – Quality

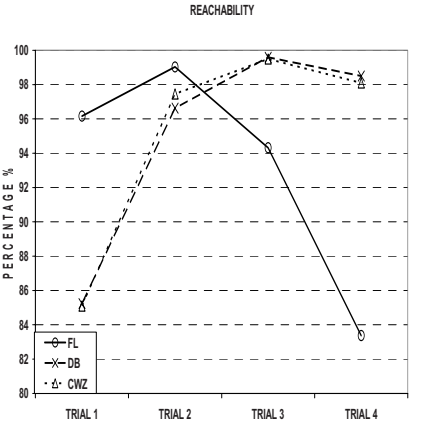


Fig. 12. Group 2 – Reachability

Figure 12 shows that the higher broadcast generation rates that are used in this group of trials have an early affect on FL in terms of reachability, which appears to start collapsing after the second trial despite the fact that it performs better than DB and CWZ up to that point. The distance-based Scheme and Constant Width Zone follow the same pattern throughout the experiment outperforming Flooding for trials 3 and 4.

Figure 13 reflects the average speed of the broadcast process in a more extreme environment (higher broadcast generation rate). DB and CWZ appear not to be affected considerably from that environment alteration. FL is outperformed in all trials in terms of speed, following the collapse of reachability after trial 2.

Again, Figure 14 reflects the overall performance of the three schemes with $A_1 = A_2 = 0.5$. The collapse of FL for trials 3 and 4 considering both reachability and speed

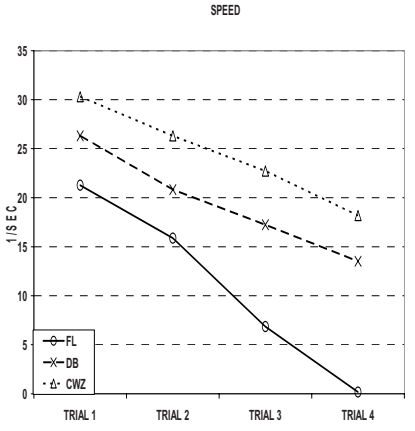


Fig. 13. Group 2 – Speed

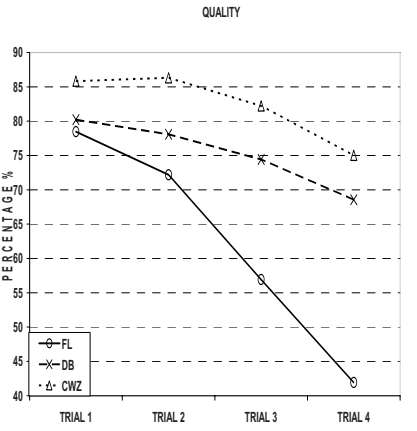


Fig. 14. Group 2 – Quality

is presented here in a single graph. Quality for DB is lower than CWZ as both schemes provided the same level of reachability with the latter being clearly faster for all trials.

5 Conclusions

In this paper, we have shown how the use of the distance-based scheme leads to growing of the width of broadcast zones and its effects on the number of forwarding hosts eligible to rebroadcast a particular broadcast message. We have also proposed a novel broadcast algorithm, called Constant-Width Zones (CWZ). CWZ can effectively alleviate the redundant rebroadcast problem by defining a constant upper limit for the width of all rebroadcast zones and, consequently, reducing the number of forwarding hosts.

Through simulation-based study, the performance of CWZ has been compared against those of simple flooding and distance-based algorithms using the network simulator NS2.30 under different operational conditions. It has been demonstrated that the proposed CWZ algorithm is able to rebroadcast packets more effectively in order to achieve higher reachability while reducing the latency especially for heavy host density networks.

References

[1] Cavin, D., Sasson, Y., Schiper, A.: On the Accuracy of MANET Simulators. In: POMC (2002)

[2] Viswanath, K., Obraczka, K.: Modeling the performance of flooding in wireless multi-hop ad hoc networks. Computer Communications 29, 949–956 (2006)

[3] Ni, S., Tseng, Y., Chen, Y., Sheu, J.: The broadcast storm problem in a mobile ad hoc network. In: MOBICOM (1999)

[4] Leng, S., Zhang, L., Wu Yu, L., Heng Tan, C.: An efficient broadcast relay scheme for MANETs. Computer Communications 28, 467–476 (2004)
[5] Zhu, C., Lee, M.J., Saadawi, T.: A Border-aware broadcast scheme for Wireless Ad Hoc Networks. IEEE Explore (2004)
[6] Qayyum, A., Viennot, L., Laouiti, A.: Multipoint relaying for flooding broadcast messages in mobile wireless networks. Computer Society (2002)
[7] Hsu, C., Chen, C., Wang, H.: DISCOUNT: A Hybrid Probability-Based Broadcast Scheme for Wireless Ad Hoc Networks. IEEE, Los Alamitos (2005)
[8] Purtoosi, R., Taheri, H., Mohammadi, A., Foroozan, F.: Improving broadcast performance by traffic isolation in wireless ad hoc networks. Int. J. Commun. Syst. 19, 1029–1043 (2006)
[9] Barritt, B.J., Malakooti, B., Guo, Z.: Intelligent Multiple-Criteria Broadcasting in Mobile Ad-hoc Networks. IEEE LCN / P2MNet (2006)
[10] Bauer, N., Colagrosso, M., Camp, T.: Efficient implementations of all-to-all broadcasting in mobile ad hoc networks. Pervasive and Mobile Computing 1, 311–342 (2005)
[11] Khelil, A., Marron, P.J., Becker, C., Rothermel, K.: Hypergossiping: A Generalized Broadcast Strategy for Mobile Ad Hoc Networks. Ad Hoc Networks archive 5(5) (2007)
[12] Williams, B., Camp, T.: Comparison of broadcast techniques for mobile ad hoc networks. In: MOBIHOC (2002)
[13] Zhang, H., Jiang, Z.: Modeling and Performance Analysis of ad hoc broadcasting schemes. Performance Evaluation 63, 1196–1215 (2006)
[14] Williams, B., Mehta, D., Camp, T., Navidi, W.: Predictive Modeling of Network Wide Broadcasting Protocols for Mobile Ad Hoc Networks. IEEE Transactions on Mobile Computing, TMC (2003)
[15] Lee, F.B.S., Seet, B.-C., Liu, G., Wong, K.-J., Lee, K.-K., Zhu, L., Huang, S.-Y.: Performance of New Broadcast Forwarding Criteria in MANET. In: Kahng, H.-K., Goto, S. (eds.) ICOIN 2004. LNCS, vol. 3090, pp. 34–43. Springer, Heidelberg (2004)

Orientation-Aware Indoor Localization Path Loss Prediction Model for Wireless Sensor Networks

Marc Lihan, Takeshi Tsuchiya, and Keiichi Koyanagi

Waseda University, Graduate School of Information, Production, and Systems
2-7 Hibikino, Wakamatsu-ku, Kitakyushu-shi, Fukuoka, Japan 808-0135
lihan@fuji.waseda.jp, tsuchiya@suou.waseda.jp, keiichi.koyanagi@waseda.jp
http://www.waseda.jp/ips/english/index.html

Abstract. There have been a large amount of research and interest in the area of ubiquitous and indoor location aware computing in the past decade. Among several proposed algorithms, fingerprint algorithm stands as one of the most accurate systems for localization. However, there is a lack of theoretical basis and understanding on the orientation of the user. This paper presents a model for orientation-aware indoor location tracking system using a Zigbee based protocol wireless sensor called Sun's SPOT (Small Programmable Object Technology). Our experiment shows better accuracy in location tracking when orientation and attenuation factors are considered for the path loss prediction model than the traditional path loss model. Orientation-aware fingerprint algorithm is also examined in our experiment to have a basis of comparison on an empirical algorithm.

Keywords: indoor localization, RSS, path loss model, WSN, fingerprint algorithm, ZigBee, tracking system.

1 Introduction

Location tracking system for indoor areas utilizing wireless technology is becoming an eager interest of research community in recent years. To provide location aware service, obtaining the position of a user accurately is important. In areas where global positioning system (GPS) does not work well, such as inside a building and in a campus area, several local positioning services have been proposed. Considering the physical and technological limitations, recent studies show that location fingerprint algorithm remains one of most feasible solution for indoor localization. Wireless geolocation techniques have shown promising result and modern applications, not only in commercial and military potential but also provides add-ons to today's pervasive wireless technology. Few examples are inventory tracking, patient monitoring, security, and asset tracking.

Received signal strength (RSS) is a measurement of the power present in a received radio signal. RSS-based localization is an attractive research topic where many researchers have proposed techniques by using existing infrastructure to

M. Takizawa, L. Barolli, and T. Enokido (Eds.): NBiS 2008, LNCS 5186, pp. 169–178, 2008.
© Springer-Verlag Berlin Heidelberg 2008

deploy a positioning system such as IEEE 802.11b, wireless LAN [1], [7], [8]. However when the additional number of nodes in the localization system multiply, additional hardware and setup will be costly. The rapidly increasing popularity of wireless sensor network has extensively studied and proposed several RSS-based algorithms. Low power, low data rate, radio frequency (RF) based nodes are ideal for RSS-based localization since only minimal data is needed. Wireless techniques such as infrared, Bluetooth, Wireless LAN (WLAN), ultra-wideband, and ZigBee can be used as infrastructure. Among them, we chose ZigBee because it is suited for high level communication protocols using small, low-power digital radio based on IEEE 802.15.4 standard for wireless personal area networks (WPANs). In addition, localization system requires low data rate, long battery life and secure networking which are the features of ZigBee.

Generally, in an indoor location tracking system there are two types of nodes: reference nodes and blind node. Reference nodes are nodes that are aware of their location and are strategically based in the area of navigation. On the other hand, the blind node is held by a user, or connected to an object, that moves around the area. When the blind node broadcasts packets to the reference nodes within the area, the reference nodes receives the signal characteristic indicator in the form of packets and measures the RSS in decibels. Not only the location but also the orientation of the moving person will be considered. In the paper, we focus on improving the path loss prediction model for localization system by considering the orientation of the user to determine his/her location using RSS of ZigBee networks. Theoretical model for simulating radio propagation is the path loss prediction model [2]. The model provides flexibility in accommodating different situations while taking into consideration the large scale path loss.

This paper is organized as follows. Section 2 deals with related wireless indoor localization techniques. In Section 3, a model for the orientation-aware location system is proposed and algorithms to implement in the experiment will be defined. Section 4, the attenuation factors for different directions are derived based on experimental data. Conclusion and future works are discussed in Section 5.

2 Related Work

Over the past few years many solutions have been proposed for RF-only localization in wireless ad-hoc and sensor networks. Existing indoor localization systems can be divided into systems that rely on off-the-shelf products and systems that use specialized hardware. The first comprise of mostly using IEEE 802.11 compliant network hardware (WiFi) [1], [7]-[11], [14]. The RADAR system was one of the earlier WLAN-based location estimation systems using fingerprinting approach and it became a common measurement for algorithm effectiveness[7],[14]. The system considered orientation of the user by recording the direction at the time the measurement is made. Studies have shown that RSS varies due to obstruction of human body however have not come up with a solution to mitigate the effects. COMPASS applied digital compasses to detect the orientations of the users to deal with the blocking effects [10]. Base on the orientation of the user, preselected subset of offline data is used by a probabilistic positioning algorithm

to determine the position of the user. Kaemarungsi and Krishnamurthy [1] designed a model for indoor positioning system based on location fingerprint algorithm. However the model assumed that the variations due to orientation had been averaged out when the RSS is recorded for all positions.

The second group of localization systems uses specialized hardware like Active Badge and Cricket system. Active Badge utilizes badges that are capable of emitting infrared signals mounted on the ceiling to detect signals and interpret the position of the users by determining the room where the user is located [12]. The successor of this system is called Active Bat, where it used ultrasound time-of-flight information and multilateration to determine the position of users more precisely. Similarly, Cricket system used ultrasound and RF receivers to locate specially equipped mobile devices [13]. Recent research by Cho, et al, proposed a maximum likelihood location estimation algorithm applied in ZigBee networks based on clustered tree topology for supporting home environments [6].

3 Research Methodology

In this section, we define a 2-dimensional grid model. We will estimate the location of the blind node by measuring the RSS. Describing the orientation will be followed based on the polar coordinate system of defined ranges. The two main algorithms in this research are fingerprint and path loss model. Fingerprint uses actual online data to estimate the location of a node. While, path loss model gives RSS measurement based on log-normal random variables [9].

3.1 System Model Setup

Consider an indoor positioning system with installed wireless sensor network on a single floor inside a building. Fig. 1 illustrates an indoor positioning system model that contains L points along both x and y axes, hence having a $L \times L = L^2$ square grid with $M = (L+1)^2$ markers a meter apart.

There are N reference nodes in our experiment that are strategically located around the area. Signal measurements will be handled in the form $(x, y, o, rssi)$,

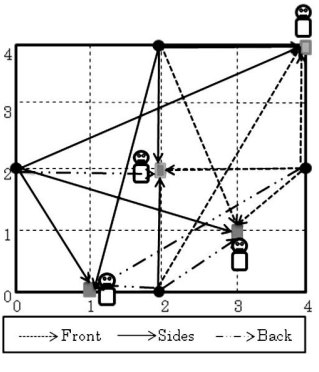


Fig. 1. Model for orientation aware location system with four sample scenarios

where $i \in \{1, \cdots, N\}$. For example, Fig. 1 shows the orientation from different points of a 4×4 grid $(1,0), (2,2), (3,1)$, and $(4,4)$. Hence, this shows a total of 5 front, 7 sides, and 4 back orientations.

3.2 Categorizing Orientation

Although indoor location systems have been investigated for quite some time, there is a lack of theoretical analysis in estimation of the blind node's orientation without using a specialized hardware like a digital compass. In real world scenarios, a blind node is accompanied with objects or human beings although studies have mentioned minimal details on what it is attached to. This is important because the object attached to the blind node may significantly change the signal strength. A blind node can be attached to an operating machine inside a hospital or visitors in a museum. In our experiment, node will be worn and located on the chest of a person, very similar to an ID badge of office workers or students. The user handling the sensor will be the same throughout the experiment.

Polar coordinate system will be implemented to measure the orientation of the nodes. The radial coordinate denotes the distance from the central point (blind node), to a reference node. The angular coordinate θ denotes the counterclockwise angle required to reach the point from the polar axis, commonly known as the positive x-axis in Cartesian plane. Adding more categorized orientations leads to a diminishing marginal utility hence improvement is minimal [7], [10]. Measurements will be distinguished in two dimensional space categorized and defined in Table 1. Left and right orientations are considered to be in one category and are less prioritize.

Table 1. Range in radians of the respective orientations

	Range in radians
Front(F)	$[\frac{1}{4}\pi, \frac{3}{4}\pi]$
Side(S)	$(\frac{3}{4}\pi, \frac{5}{4}\pi) and (\frac{7}{4}\pi, \frac{1}{4}\pi)$
Back(B)	$[\frac{5}{4}\pi, \frac{7}{4}\pi]$

3.3 Fingerprint Algorithm

There are two ways to determine the distance between the nodes using RSS information. First is to map the path loss of the received signal to the distance traveled by the signal from the reference node to the blind node. With at least 3 reference nodes, we can locate the mobile node using triangulation. Without using a database, the estimation will be related to communication and computation. A simple obstruction such as a person blocking in between the sensors will cause the RSS to change significantly. The other way to determine distance is to use a database for RSS values (e.g., fingerprint algorithm such as [1], [7], [9], [10]). Fingerprint technique is relatively simple to deploy compared to time-of-arrival and angle-of-arrival [1]. It also outperforms other well-known algorithms such as triangulation, and smallest M-vertex polygon [2]. Fingerprinting-based algorithms use two phases: offline phase and online phase.

Offline Phase. During this phase, the signal strength distributions collected from the reference nodes at predefined markers in operation area are stored together with their physical coordinate in a database. The RSS are measured with enough statistics, depending on dependent on the location coordinates and orientation, to create a database of predetermined signal values on every marker. Offline data that forms the fingerprints consist of the true means of all the RSS at a particular location from N reference nodes. It will be represented as vector $R_{jk} = \{r_1, r_2, \cdots, r_N\}$ for the j-th marker in k-th orientation.

Online Phase. In the online phase the blind node will report a snapshot of a measured RSS values from different reference nodes within range. Then, the blind node will collect the values and send the data to the designated server. The server will calculate the estimated location and direction of the blind node using a series of algorithms. The fingerprints in a database (offline) and are compared with the measured snapshot of RSS values of the blind node. Online data are the actual RSS of the blind node from N reference nodes. It will be represented as vector $P_{jk} = \rho_1\rho_2, ..., \rho_N$ for the j-th marker in k-th orientation.

There are two basic algorithms in computing for the blind node's location [8]. The signal distance between the online and offline vectors is used to calculate the location of the blind node. Note that the signal distance is not the actual physical distance between the two positions. The common metric to compute for the signal distance is the Euclidean distance. The signal distance for the j-th marker in k-th orientation will be the square root of the summation of the difference between the vectors of offline data, R, and the vectors of online data, P, and will be presented as SD_{jk}.

$$SD_{jk} = [\sum_{i=0}^{N} (r_i - \rho_i)^2]^{\frac{1}{2}} \ . \tag{1}$$

The second part algorithm involves selecting a number of smallest signal distances. Commonly, 3 signal distances are chosen and the corresponding coordinates of the j-th marker serves as input for the algorithm. The location is estimated as the centroid of these coordinates. Averaging the 3 closest markers (based on signal distances) will be performed in our experiment. In addition, we include a simple biased function. For example, 50% of the result is based on the first marker, 30% for the second, and 20% for the third. Producing results that are closer to the best SD might improve the estimation. We evaluate such modifications to the algorithms and the results show that the accuracy improvement is negligible. Note that equation 1 considers the k-th orientation of the blind node however disregarded in the second algorithm because we are not concerned about estimating the orientation of the node.

3.4 Distance Dependent Path Loss Prediction Model

The model used in [1]-[3] indicates that the mean path loss increases exponentially with distance. The absolute mean path loss measure in decibels (dB) is represented as mean of PL [2].

$$\overline{PL}(d_{ij}) = PL(d_0) + 10 \times \alpha \times log_{10}(\frac{d_{ij}}{d_0}) \ . \tag{2}$$

In equation 2, $PL(d_0)$ is free space propagation from a reference distance of $d_0 = 1$m. In this case, according to Janssen and Prasad [3], for line-of-sight propagation (LOS) is 41.5dBm and for non-line-of-sight (NLOS) is 37.3bBm. While, α denotes the path loss exponent, which for indoor locations of 2.4GHz frequency, it is reported to be 2 for LOS propagation and 3.3 for NLOS propagation, indicating how fast path loss increases with distance. In other situations, path loss exponent can be between 1 and 6. Variable d indicates the distance between the reference and blind nodes. For the physical distance d_ij in meters, it is the j-th point on the grid from i-th reference node. The mean RSS can be predicted as the difference between the transmitted power, Tp and mean of PL.

$$\overline{RSS}(d_{ij}) = Tp - \overline{PL}(d_{ij}) \ . \tag{3}$$

The mean received signal strength can be found by equation 3, which is the difference between the transmitted power (Tp) and the mean path loss. Tp of the reference node ranges from -24dBm (decibels of measured power referenced to milliWatt) to 0dBm based on the IEEE 802.15.4 of carrier frequency 2.4GHz [5]. In our experiment, SPOT is configured to its highest possible Tp which is 0dBm.

Section 3.1 suggests categorizing the orientation of the blind node into three: front, sides, back. Hence, we apply this to equation 2 by including attenuation factors for orientations facing the sides and back. Attenuation factors are initially introduced for path loss models considering multi-floor environments [2]. A constant floor attenuation factor (FAF) which is a function of the number of floors and building type, may be added to the mean path loss equation. Path loss effects of soft partitions and concrete walls between the reference and blind nodes have also been considered [2]. Similarly, we propose to include attenuation factors for orientations, sides and back.

4 Experiment and Analysis

Wireless sensor that is used for this research is produced by Sun Microsystems, called Sun SPOT [4]. It has a ZigBee protocol based on IEEE 802.15.4, which is designed for low-data-rate, low-power-consumption, and low-cost applications. IEEE 802.15.4 standard defines the physical layer and medium access control (MAC) sublayer specifications for low-rate wireless personal area network (LR-WPAN), which supports simple devices that consume minimal power. Not only opens the door to a large number of new applications, but also brings added value to various existing applications. Commonly used signal characteristic for WLAN's is the RSS, which is easily available and has been used in several researches including [1], [7], [8]. Similarly, Sun SPOT which is based on IEEE 802.15.4 standard also provides application for returning RSS [4].

4.1 Experimental Setup

Since we are developing an initial model for orientation-aware location tracking system, experiments are performed in a small scale area shown in Fig. 1. In our test site, there will be N=4 reference nodes located in $(0,2),(2,4),(4,2)$, and $(2,0)$ and will be called reference node A, B, C, and D respectively. For each marker and direction we collected at least 6 samples each, in total having 2,016 signal measurements. For most of our analysis, we use the mean rather than the original, raw data set.

4.2 Determining Attenuation Factors

For each of the discrete path loss measurements for sides and back orientations, we calculated the difference between the actual path loss, $-[R] = \{-r_1 - r_2, \cdots, -r_N\}$ and the path loss, $PL(d)$, that would occur due to free space propagation for a reference and blind nodes at the same distance. The minimum mean square was used to find the best fit for the attenuation factors of sides and back, represented as AF_S and AF_B, respectively. Referring to the left portion of Fig. 2, it was not a surprise that the front orientation had the minimal percentage error, while the back orientation had the most percentage error due to human body's obstruction. Minimizing the error for these directions inclusion of attenuation factors is analyzed to be $AF_S = 2.2$dBm and $AF_B = 6.7$dBm. Fig. 2 shows a detailed graph of the percentage error for each reference node (*subscripted with 1*) and when attenuation factors were considered (*subscripted with 2*).

Fig. 3 shows the averaged percentage error for with (*right*) and without (*left*) attenuation factors. Percentage error for sides and back are 11.43% and 13.7%, respectively. Also, Fig. 3 shows that the sides orientation improved slightly on the other hand, back orientation increased accuracy by 29.6%. The distribution of errors between measured and predicted path loss is log-normal with a standard deviation of 4.11dBm.

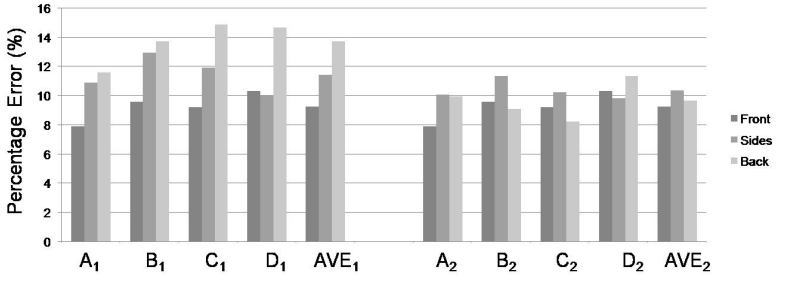


Fig. 2. Graph of the percentage error for the reference nodes without (*left*) and with (*right*) attenuation factors

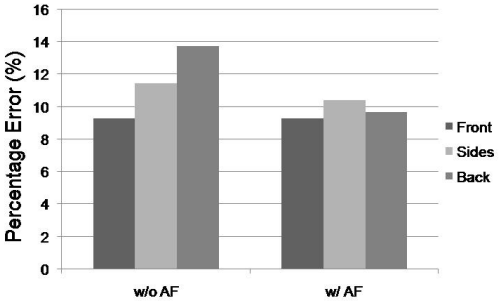


Fig. 3. Effect of attenuation factors to accuracy

4.3 Experimental Results

The theoretical and empirical localization algorithms are namely, path loss model based on equations 2 and 3, and fingerprint algorithm based on equation 1 and Euclidean distance formula. Traditional path loss prediction model and path loss with orientation attenuation factors are also compared in the experiment, labeled *PL* and *PL-AF* respectively. On the otherhand, fingerprint algorithm, Euclidean distances gathered the smallest SD between the reference node and the blind node, labeled as *Euclidean* in Fig. 4. Also based on Euclidean algorithm, *3-best* and *bias* functions (refer to second heading of Section 3.3) returns the 3 smallest SD's and then derive the estimated location. In addition, for each algorithm, the average, 70 percentile, 90 percentile, and median are also shown.

Fig. 4 shows the accuracy of these algorithms. As expected the traditional algorithms labeled *Euclidean* and *PL*, gave a higher distance error than modified algorithms. *PL* has the worst accuracy because of the variation of RSS with distance due to obstructions and multipath fading effects. The *3-best* or averaging the top ranked locations has the best 90% performance compared to all other schemes. Implementing path loss together with the derived attenuation factors shows an improvement in estimating the theoretical location.

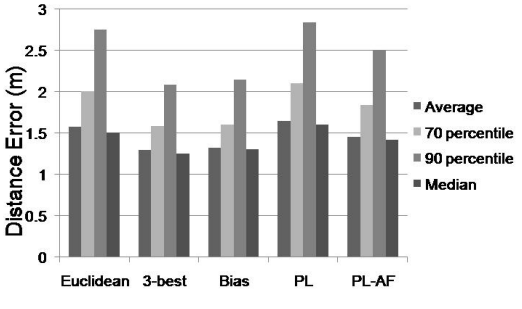


Fig. 4. Performance evaluation of different algorithms

5 Conclusion and Future Work

This paper has provided a framework for systematically analyzing orientation-aware indoor localization system using Sun's SPOTs. We have address issues relating to orientation of the blind node but more signal measurements need to be collected in larger environment. We only have preliminary results on the attenuation factors that showed improvement on the accuracy of path loss prediction model. Base on our research, fingerprint algorithm still stands as the better algorithm even for orientation-aware location tracking system.

As future work, including orientation in algorithms increases the computation time and database space. We will focus on the tradeoff between the above mentioned and accuracy. We will also perform experiments in a medium scale office area to gather a more signal measurements and more *realistic* data.

References

1. Kaemarungi, K., Krishnamurthy, P.: Modeling of Indoor Positioning Systems Based on Location Fingerprinting. In: INFOCOM 2004, vol. 2, pp. 1023–1022 (2004)
2. Seidel, S.Y., Rappaport, T.S.: 914 MHz Path Loss Prediction Models for Indoor Wireless Communications in Multi-floored Buildings. IEEE Transaction on Antennas and Propagation 40(2), 207–217 (1992)
3. Janssen, G.J.M., Prasad, R.: Propagation Measurements in an Indoor Radio Environment at 2.4 GHz, 4.75 GHz and 11.5 GHz. In: IEEE 42nd Vehicular Technology Conference, Denver, CO, USA, vol. 2, pp. 617–620 (1992)
4. Sun Microsystems, Project Sun SPOT, http://www.sunspotworld.com
5. CC2420 2.4 GHz IEEE 802.15.4/Zigbee-ready RF Transceiver. Chipcon Products from Texas Instruments Incorporated, http://ti.com/lit/gpn/cc2420
6. Cho, H., Kang, M., Park, J., Park, B., Kim, H.: Performance Analysis of Location Estimation Algorithm in Zigbee Networks using Received Signal Strength. In: 21st International Conference on Advanced Information Networking and Applications Workshops, AINAW 2007, pp. 302–306 (2007)
7. Bahl, P., Padmanabhan, V.N.: RADAR: An In-Building RF-based User Location and Tracking System. In: Proceedings of 19th Annual Joint Conference of the IEEE Computer and Communications Societies. INFOCOM 2000, vol. 2, pp. 775–784 (2000)
8. Prasithsangaree, P., Krishnamurthy, P., Chrysanthis, P.: On Indoor Position Location with Wireless LANS. In: The 13th IEEE International Symposium on Personal, Indoor Mobile Radio Communications, vol. 2, pp. 720–724 (2002)
9. Madigan, D., Elnahrawy, E., Martin, R.: Bayesian Indoor Positioning Systems. In: Proceedings of IEEE 24th Annual Joint Conference of the IEEE Computer and Communications Societies. INFOCOM 2005, Miami, Florida, vol. 2, pp. 1217–1227 (2005)
10. King, T., Kopf, S., Haenselmann, T., Lubberger, C., Effelsberg, W.: COMPASS: A Probabilistic Indoor Positioning System Based on 802.11 and Digital Compasses. In: Proceedings of the First ACM International Workshop on Wireless Network Testbeds, Experimental evaluation and CHaracterization (WiNTECH 2006), Los Angeles, CA, USA, September 2006, ACM, New York (2006)

11. Castro, P., Chiu, P., Kremenek, T., Muntz, R.R.: A Probabilistic Room Location Service for Wireless Networked Environments. In: Abowd, G.D., Brumitt, B., Shafer, S. (eds.) UbiComp 2001. LNCS, vol. 2201, pp. 18–34. Springer, Heidelberg (2001)
12. Want, R., Hopper, A., Falcao, V., Gibbons, J.: The Active Badge Location System. ACM Transactions on Information Systems (TOIS) 10(1), 91–102 (1992)
13. Priyantha, N.B., Chakraborty, A., Balakrishnan, H.: The Cricket location-support system. In: Proceedings of the 6th annual international conference on Mobile Computing and networking, International Conference on Mobile Computing and Networking, pp. 32–43. ACM, New York (2000)
14. Bahl, P., Padmanabhan, V.N.: Enhancements of the RADAR User Location and Tracking System. Technical Report MSR-TR-2000-12, Microsoft Research, Microsoft Corporation One Microsoft Way Redmond, WA 98052 (2000)

S-Web: An Efficient and Self-organizing Wireless Sensor Network Model

Hanh Le[1,*], Doan Hoang[2], and Ravi Poliah[3]

[1] Computer Science Department, University of Cape Town, South Africa
hanh@cs.uct.ac.za
[2] Computer Systems Department, University of Technology Sydney, Australia
dhoang@it.uts.edu.au
[3] Information & Communication Technology, Parliament of the Republic of South Africa
rpoliah@parliament.gov.za

Abstract. One of the major weaknesses of existing centralized algorithms in Wireless Sensor Network (WSN) routing is that sensor nodes are required to report their status (e.g. remaining energy level) to the central node (i.e. the Base Station –BS) and receive configuration information from the BS (e.g. next hop routing). As a result, the WSNs suffer from high communication and/or computation overheads at sensor nodes to accommodate topology changes in the WSN. This paper proposes a novel scheme, called a Sensor Web (or S-Web), that organizes sensors into clusters based on their geographical location without requiring the sensors to have a Global Positioning System or actively locate themselves. The S-Web enables nodes to route data packets while consuming low energy in a decentralized manner. The model is self-organizing and distributed without the need of global network knowledge.

Keywords: Wireless Sensor Network, Self-organizing, Decentralization, Energy Consumption, Sensor Web.

1 Introduction

Wireless Sensor Networks (WSNs) have emerged in the last few years providing a rich set of environmental information with a wide variety of useful applications. Sensors are small devices with limited storage, processing capability, and battery power. WSNs often involve a large number of sensors. Due to power limitation, sensors need to use their power wisely and sparingly to prolong their lives. They often employ communication protocols that allow them to sleep when they are idle in order to preserve their energy. Sensors may also be mobile and often cannot participate in WSNs for a long period. As a consequence the topology of a WSN is dynamic and requires frequent reconfiguring, leading to high communication and computation overheads.

A simple way to send data to a single sink is via broadcasting. The problem with this model is that it is not energy efficient as a sink may be far from the source. As a

[*] South African Patent Application Number ZA2008/02427.

M. Takizawa, L. Barolli, and T. Enokido (Eds.): NBiS 2008, LNCS 5186, pp. 179–188, 2008.
© Springer-Verlag Berlin Heidelberg 2008

result, all the relay sensors would excessively consume their resources leading to quick power exhaustion.

The main constraint in WSNs is that wireless sensor nodes have very limited battery power thereby limiting their transmission range. A base station is often more powerful with a continuous power supply as it has to perform some data pre-processing, aggregate data from multiple sensors, and transmit data to processing stations. To send data from sensors to the Base Station, multi-hop routing is often used in WSNs to reduce the power consumption in the sensor nodes and prolong the WSN lifespan. The most common routing solution is to form a tree [1-3] rooted at the sink. However a single sink/route means a lack of routing redundancy hence a single point of failure. If there is more than one sink in a system then the WSN needs to deploy a certain routing algorithm at the sensors to determine to which routes to forward the information [4, 5].

In this paper we propose a novel scheme, called a sensor web or (S-Web), to group sensors based on their geographic location relative to the base station (BS) for efficient transfer of data and hence prolong the lifespan of the whole WSN. S-Web possesses several innovative features. Firstly, S-Web clusters sensors in a distributed manner with minimal communication and computation overhead. Secondly, S-Web allows data to be transmitted between any sensors without going through the BS. The sensors use only local network knowledge without the need of global knowledge from the BS. Thirdly, S-Web seeks to produce short routes between sensors and the BS without trying to form a minimal spanning tree from sensors to the BS which would generate an additional level of overhead.

The rest of the paper is organized as follows. Section 2 presents related work. S-Web system design is described in Section 3 and Performance Evaluation is in Section 4. Finally Section 5 concludes our paper.

2 Related Work

Several approaches have been investigated for data gathering and transferring from sensors to their Base Station (BS). A simple approach is for each node to transmit its data directly to the BS. This "Direct" communication scheme consumes a large amount of energy especially when sensor nodes are far from the BS. This leads to the sensors exhausting their battery quickly.

LEACH [6] is another model of routing data to the BS. LEACH partitions sensor nodes into a number of clusters, each with a cluster head. Sensors in the same cluster take turns to be the head of the cluster and send data to their cluster head. Cluster heads then send their aggregated and compressed data packets to the BS to reduce the overall energy consumption.

In multi-hop routing model, sensor nodes form a tree structure (such as SHORT [1], tree [2]) to multi-hop fused data to the BS. Each routing hop is short so the energy consumed per hop is low. However the tradeoff of this model is that the delay from the source of the information to the BS is high ($> \log_2 N$ where N is the number of sensor nodes). This could make the scheme unsuitable for delay-sensitive applications.

Most of the previous work relies on a central point (i.e. the BS) to calculate the routing tree to optimize the data communication cost. However this comes at the price of higher overhead in gathering WSN information to the BS and configuring the WSN (e.g. designate the root, select neighboring nodes).

While existing schemes (e.g. [7], [8], [1]) adopt centralized algorithms at the BS to manage the network topology and to calculate the routing path, S-Web relies very little on the BS for these purposes. S-Web allows sensors to self-organize into clusters and use local information to route data. This helps to reduce energy consumption and traffic to the BS which can be far from the sensing field.

3 System Design

In this paper we propose a novel scheme, called a sensor web or (S-Web), to group sensors based on their geographic location relative to the base station (BS). This geographical location is determined by two factors: distance and angle. Distance from a sensor node to the BS is determined by signal strength received at the sensor node. Angle is an order of a scanning sweep by the BS at a specific time.

Initially, the BS will send beacon signals for every α degree angle at a time. Sensors that receive the beacons at time slot i will measure their signal strength to determine their relative distances to the BS. Let T be a predefined distance (which is inversely proportional to the received signal strength). All sensors which receive beacon signals at angle order β_i ($=i*\alpha$) with signal strength of δ_j*T (within sector j) will be in the same group/cluster, denoted as (β_i, δ_j).

Note that different from previous work, sensor nodes in S-Web neither require the capability to measure angles of arrived signal nor the need of a GPS receiver [8] to locate themselves. Rather the BS sends information of angle order β_i in the beacons for nodes in the angle $i*\alpha$, and (β_{i+1}) for nodes in angle $(i+1)*\alpha$ and so on. With this information and the measured signal strength from the BS, S-Web nodes can place themselves in a proper cluster.

Figure 1 illustrates a sensor network of 100 nodes distributed in the area of [100X100]. The root of the coordination (0;0) is at the top left corner. The BS is located at a position of (50;150). Scanning angle α is 10 degree and threshold T is 35.

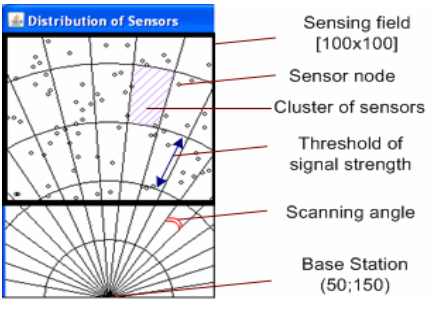


Fig. 1. An Illustration of S-Web

3.1 S-Web Clustering

In S-Web, nodes in a cluster will choose a cluster head which has the highest remaining energy to act as a router for the cluster. This task is done locally between nodes in the same cluster. The role of heads is rotated amongst nodes in the same cluster to balance the load and prolong the lifespan of the cluster [6].

We can apply previous methods such as: LEACH [6] in defining the topology structure of nodes in a cluster to achieve better output for S-Web. However in this paper, we focus mainly on the performance of the S-Web so we simply let all non-head nodes send data to their head as a routing relay node in order to communicate with nodes outside of the cluster.

Table 1 is an instance of cluster distribution where each cell in the table **(β, δ) nodeID** is an angle order β and distance order δ from cluster head nodeID to the BS.

Table 1. An example of cluster distribution

	(24,1) 27	(25,1) 5	(26,1) 65	(27,1) 16	(28,1) 62		
(23,2) 56	**(24,2) 1**	(25,2) 64	(26,2) 59	(27,2) 88	(28,2) 14	(29,2) 68	(30,2) 83
	(24,3) 10	(25,3) 55	(26,3) 50	(27,3) 58	**(28,3) 3**	(29,3) 82	
	(24,4) 79	(25,4) 28	(26,4) 78	(27,4) 21	(28,4) 94	(29,4) 49	

3.2 Routing Table

Each head node has a routing table `RT` which points to head nodes of surrounding clusters. Sensor myNode (β_{myNode}, δ_{myNode}) has a list of surrounding neighboring heads s, which has cluster information (β_s, δ_s). Following is a procedure to insert a neighboring sensor to myNode routing table `RT` of size [3x3].

```
Procedure Insert(Sensor myNode, Sensor s) {
  //calculate the column index: col
  angle_dist = βs - βmyNode;
  if (angle_dist < 0)col = 2;//right sector (bigger angle)
  else if (angle_dist == 0) col = 1; //same angle
  else col = 0;//left sector (smaller angle)
  //calculate the row index: row
  signal_dist = δs- δmyNode;
  if (signal_dist < 0)row =2;//bigger signal order
  else if (signal_dist == 0) row = 1;//same signal order
  else row = 0;//smaller signal order (stronger signal)
  RT[col][row] = s;
}
end Insert
```

Continuing with the example in Table 1, Node 1 and Node 3 would have routing tables as in Table 2.

Table 2. Examples of routing tables of node 1 and node 3

Routing table of node 1		
null	27	5
56	**1**	64
null	10	55

Routing table of node 3		
88	14	68
58	**3**	82
21	94	49

3.3 Next Hop Routing

When a sensor needs to transmit data to the BS or to other nodes, if it is non-head, it will transmit the data to its head. The head node will then forward the data packet to its neighbour which is closer to the destination than itself. Closeness is determined by the difference of angle order and signal strength order between the neighbour and destination in comparison to the one between the current node and destination. Following is the Pseudo-code of finding a next hop neighbour. In which:

The destination node is `dest`
The present node is `myNode`
`RT` is a routing table of `myNode`
β_x is an angle order of node `x`
δ_x is an order of signal strength range of node `x`

```
Procedure findNextHop(Sensor myNode; Sensor dest) {
   //find next hop in the routing table of current
   //node: myNode to the destination sensor dest
   next = null; //next hop
   min_dist = | β_dest - β_myNode |+ | δ_dest - δ_myNode |;
   //equal to the number of hops from myNode to dest
   if (min_dist == 0)
      next = myNode;//myNode is the head of dest cluster
   else {
      //find a sensor in the routing table that closest
      //to dest in terms of angle β and signal δ
      s ∈ RT which has either(|β_s-β_dest|.< |β_dest-β_myNode|) or
      (|δ_s-δ_dest| < |δ_dest-δ_myNode|)
      next = s that has minimum (|β_s-β_dest| + |δ_s-δ_dest|)
      //next has the lowest number of hops to dest
   }
end findNextHop
```

Routing data from a node to the BS is done differently because the BS has the same angle with every node. Therefore a node `myNode` will forward a data packet to its neighboring cluster head *X* which has signal order δ_x smaller than the current node δ_{myNode}. The neighbor node *X* should be in cluster (β_{myNode}, δ-*i*) where *i* = 1, 2, ... If a head does not have a neighbor in the direction of the BS, it can transmit data directly to the BS. However if the head is far from the BS, its energy would be used up quickly.

Depending on the distance from the BS to the sensing field, cluster heads which are closest to the BS (lowest δ_i) can link together to take turns to send fused data to the BS. This could save more energy and would further prolong the WSN lifetime. However we do not implement this scenario in this paper.

Different from previous centralized routing schemes (e.g. PEGASIS, PEDAP-PA, SHORT), in which the next hop neighbor has to be calculated at the BS, our proposed scheme, S-Web, allows nodes to determine next hop routing using their local network knowledge. In centralized schemes, gathering WSN information (such as remaining energy at sensors) requires nodes to report their status information to the BS as well as to receive the configuration information from the BS (e.g. next hop neighbors). However the impact of this communication overhead on the WSN lifespan has not been measured. Our decentralized way of routing means energy saving. At the same time, S-Web nodes can send data to any destination rather than just to the BS.

3.4 Cluster Size

The S-Web structure divides the sensing field into clusters with different sizes. The area of cluster (β_i, δ_i) is A_i.

$$A_i = \frac{\alpha * \pi}{360} * \left[\delta_{i+1}^2 - \delta_i^2\right] = \frac{\alpha * \pi * T^2}{360} * (2 * i + 1) \quad where \quad i = 0, 1, \ldots \tag{1}$$

In which α is the scanning angle and T is a predefined distance. T is inversely proportional to received signal strength of beacons from the BS (where $\delta_i = i*T$). The size of a cluster therefore, depends on the position of the cluster to the BS as well as the density of sensor nodes.

In S-Web, clusters which are further away from the BS have bigger areas than the one closer to the BS. If sensors are uniformly distributed, the cluster further away would have a larger number of sensors. Since further-away clusters need more energy to send data packets to the BS, with a larger number of nodes in the same cluster S-Web would make the overall lifetime of these clusters longer compared to the same size clustering scheme [6].

To limit the maximum distance from a sensor to its head in large/far clusters, these clusters can be divided into smaller clusters by i) further grouping nodes into two half angle order $\beta_i/2$; and/or ii) splitting the original cluster based on the order of signal strength δ_i into $\delta_i/2$, $\delta_i/3$, $\delta_i/4$ and so on. In this paper, as the furthest sensor distance to the BS is less than 160m, we do not apply these techniques.

4 Performance Evaluation

To evaluate the proposed S-Web scheme, we simulated 100 sensors randomly distributed in the area [100 X 100] m^2 field and the BS is located at the position of (50;150). All nodes have the same initial energy of 1 Joule. We implemented "Direct" and "SHORT" algorithms for comparison purposes. We use the same radio model presented in [6], in which a radio to run the transmitter or receiver circuitry consumes E_{elec} and ε_{amp} for the transmitter amplifier.

$$E_{elec} = 50nJ/bit \quad \text{and} \quad \varepsilon_{amp} = 100\,pJ/bit/m^2 \tag{2}$$

Formulas to calculate transmitting energy E_{Tx} at a distance of d and receiving energy E_{Rx} for a k-bit packet as follows:

$$E_{Tx}(k,d) = E_{elec} * k + \varepsilon_{amp} * k * d^2 \quad \text{and} \quad E_{Rx}(k) = E_{elec} * k \tag{3}$$

A data packet here has k = 2000 bits unless otherwise specified. We assume that the radios have the capability to control power and to expend the minimum required energy to reach the intended destination [7]. We also assume that sensors do not have data to send all the time. In this paper we simulate data that is sent on demand without data fusion (e.g. for time sensitive data application).

4.1 Number of Clusters

Figure 2 shows the number of clusters/groups of sensors against different values of scanning angles α and distance threshold T. T is reversely proportional to received signal strength. As expected, when α and T are too small, all nodes are on their own. As a result, there are a large number of clusters but a small number of nodes in each cluster. When α and T are large the whole area is covered by a small number of large clusters.

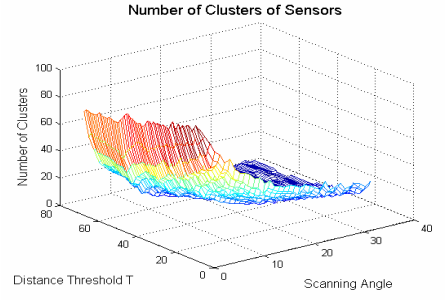


Fig. 2. The Number of Clusters vs. Scanning Angles α and Threshold of Signal Strength T

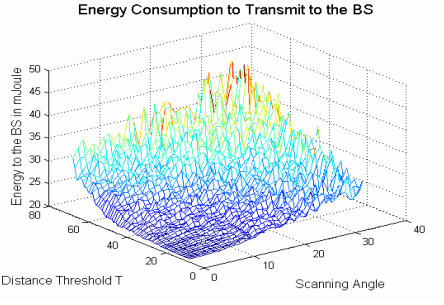


Fig. 3. Energy Consumption for Sensors to transmit to the BS without data fusion

4.2 Energy Consumption

In this paper we define an energy parameter ε which is the sum of intra-cluster energy ε_{intra} and inter-cluster energy ε_{inter} to connect all sensors. It is calculated by the formula below.

$$\varepsilon = \varepsilon_{intra} + \varepsilon_{inter} = \sum_{i=1}^{M} \sum_{j=0}^{ClusterSiz\ e} e_{ij} + \sum_{i,j=1}^{M} E_{ij} \tag{4}$$

In which, ε_{intra} is the sum of energy (e_{ij}) consumed to send and receive packets from non-head nodes to their head in the same cluster. M is the number of clusters. ε_{inter} is the sum of energy consumed to send and receive data packets between neighboring heads (E_{ij}) and from the nearest BS cluster heads to the BS.

Figure 3 depicts the energy ε with different values of α and T. When α and T are small, S-Web has the structure closer to the geometry of the WSN layout. This leads

to lower energy consumption ε. However, data from non-head sensors can be fused at the cluster head before sending to a neighbor head to be forwarded to the BS. This would reduce the energy when α and *T* are high compared to figures in Fig. 3.

4.3 Maximum Delay Units

Figure 4 and 5 are the average of maximum delay units/hops to transfer data to the BS. When sensors within a cluster are CDMA enabled, they can simultaneously send data to their head. As a result, bigger clusters (high α and *T*) have lower numbers of hops and suffer less delay (in Fig. 4). However for non-CDMA sensors, the delay is higher as cluster heads have to schedule time slots for more non-head nodes to transmit their data (in Fig. 5).

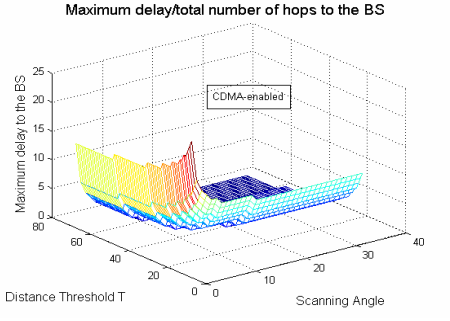


Fig. 4. Maximum delay units to transmit data to the BS with CDMA-enabled sensors

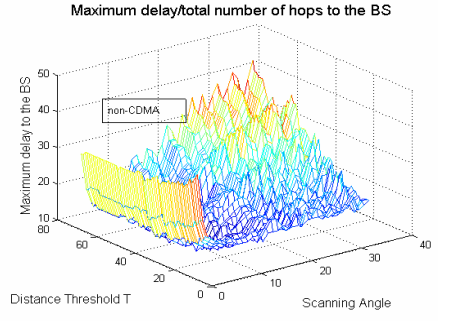


Fig. 5. Maximum delay units to transmit data to the BS for no-CDMA sensors

Previous work either tries to minimize the energy consumption at the expense of delay or optimize the [energy x delay] metric when nodes are first deployed and/or configured. It is more dynamic in S-Web where one can adjust α and *T* to optimize energy or delay while sensors are already in operation or even on the move. Note that S-Web can be applied to mobile WSNs because a new location of a mobile sensor can be determined by receiving beacon signal from the BS at the next beacon period.

We have simulated S-Web for CDMA and non-CDMA sensors. However, the best metric results also depend on the pattern of data gathering. For the rest of this paper we choose α = 10 degree and *T*= 30m in order to keep the total number of clusters smaller than 30 percent of the total number of nodes making intra-cluster energy ε_{intra} approximately equal to inter-cluster energy ε_{inter}.

4.4 Communication Efficiency

The following result is the average of the number of hops and consumed energy per message over 2000 messages.

Table 3 is the performance result of communication between sensors and the BS. S-Web consumes least energy whereas “Direct” consumes highest energy of 2619μJ per message. S-Web has a lower average number of hops and energy consumption per

Table 3. Communication Sensors and the BS

	Avg. Hops	Hops Dev	Energy (μJ)	Energy dev (J)
Direct	1.00	0.00	2619.71	1.20
S-Web	2.87	1.15	1477.71	0.39
SHORT	4.25	1.41	1787.05	0.48

Table 4. Communication between Pairs of Sensors

	Avg. Hops	Hops Dev	Energy (μJ)	Energy dev (mJ)
Direct	2.00	0.00	2679.19	1.25
S-Web	3.63	1.42	1342.80	0.57
SHORT	5.29	1.43	1902.82	0.50

message than SHORT does. "Direct" has the least hop count because data is sent directly to the BS. However if non-CDMA sensors have the same transmission frequency and send data to the same destination, collisions would occur. This might be resolved by the TDMA technique, but leading to longer delays than the shown number of hops.

Table 4 is the communication between random pairs of sensors. As can be seen from Table 4, communication between sensors in S-Web consumes the lowest average energy per message 1342μJ/message. It is even lower than the one in Table 3 to transmit to the BS. This is because sensors in S-Web can communicate with each other directly without having to go to the BS, which is located far from the sensing field.

4.5 Number of Rounds

To evaluate the WSN lifespan, we simulate random sensors that have data to send and use a round as a measure unit. A round is defined as when 1000 messages reach their destinations. Fig. 6 and 7 is the number of rounds against the percentage number of

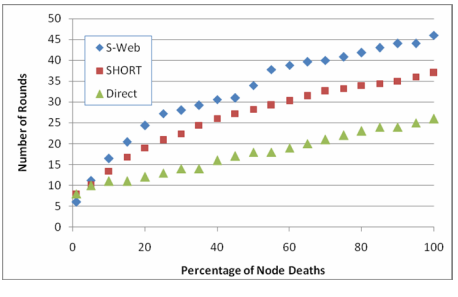


Fig. 6. Number of Communication Rounds between sensors and the BS

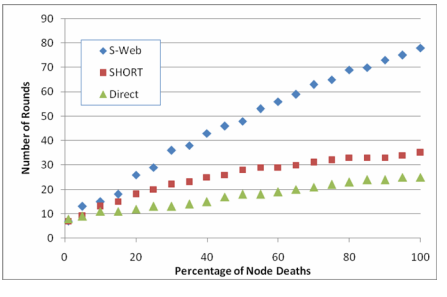


Fig. 7. Number of Communication Rounds between a random pairs of sensors

node deaths when sensors send data to the BS or to other sensors respectively. The result is an average of 5 different runs. Generally S-Web has the longest lifespan as shown in Fig. 6 and 7. However S-Web outperforms much better compared to "Direct" and SHORT especially when destinations/sinks are sensors (Fig. 7). This is because in S-Web, sensors communicate with each other directly without having to go through the BS. This proves significant as S-Web can accommodate multiple sinks (probably multiple BSs) in the system and outperforms other schemes.

5 Conclusions

We propose a Sensor Web scheme, called S-Web, which groups sensor nodes based on the scanning angle and signal strength from the Base Station (BS). The S-Web allows sensors to be clustered in a distributed manner with minimal communication and computation overheads. The size and the number of clusters can easily be adjusted to cover the sensing field with a desired level of delay and energy performance. The BS only has to sweep its beacon over the field at α degree intervals for a sensor to determine its cluster without exercising complicated locating procedure or wasting its energy on a GPS receiver to locate itself. Even though S-Web does not aim to form a minimal spanning tree from a sensor to the BS, it produces short routes between sensors and the BS. The distinctive feature of S-Web is that, using only local network knowledge, data can be transferred between any sensors without going through the BS. This prevents bottlenecks at sensors near the BS while making the WSN more scalable with a longer lifespan.

References

[1] Yang, Y., Wu, H.-H., Chen, H.-H.: SHORT: Shortest Hop Routing Tree for Wireless Sensor Networks. In: ICC 2006, Istanbul (2006)

[2] Zhenjiang, Z., Yun, L.: An Energy-Efficient Redundant Nodes Tree Mechanism for Wireless Sensor Networks. In: International Conference on Systems and Networks Communications (2006)

[3] Thepvilojanapong, N., Tobe, Y., Sezaki, K.: On the Construction of Efficient Data Gathering Tree in Wireless Sensor Networks

[4] Islam, O., Hussain, S.: An Intelligent Multi-hop Routing for Wireless Sensor Networks. In: WI-IAI 2006 workshop. IEEE/WIC/ACM (2006)

[5] Khanna, R., Liu, H., Chen, H.-H.: Self-Organization of Sensor Networks Using Genetic Algorithms. In: ICC 2006 (2006)

[6] Heinzelman, W.R., Chandrakansan, A., Balakrishnan, H.: Energy efficient communication protocol for wireless microsensor networks. In: The 33rd Hawaii International Conference on System Science, Hawaii, USA (2000)

[7] Lindsey, S., Raghavendra, C., Sivalingam, K.: Data gathering algorithms in sensor networks using energy metrics. In: IEEE 2002 Transactions on Parallel & Distributed Systems (2002)

[8] Tan, H.Ö., Körpeoğlu, I.: Power efficient data gathering and aggregation in wireless sensor networks, presented at ACM SIGMOD (2003)

Agent Based Analytical Model for Energy Consumption among Border Nodes in Wireless Sensor Networks

Haroon Malik[1], Elhadi Shakshuki[2], and Tarek Sheltami[3]

[1] School of Computing
Queens University
Kingston, Ontario, Canada K7L 3N6
malik@cs.queensu.ca
[2] Jodrey School of Computer Science
Acadia University
Wolfville, Nova Scotia, Canada B4P 2R6
Elhadi.Shakshuki@acadiau.ca
[3] Computer Engineering Department
King Fahd University of Petroleum and Minerals
Dammam, Saudi Arabia
tarek@kfupm.edu.sa

Abstract. This paper presents an agent-based analytical model to minimize the energy consumption for border nodes in Sensor-Medium Access Control (S-MAC), a cluster-based contention protocol. The S-MAC protocol is based on a unique feature; it conserves battery power at nodes by powering off nodes that are not actively transmitting or receiving packets. In doing so, nodes also turn off their radios. Inspired by the energy conservation mechanism of the S-MAC, the paper further augments the node life time in sensor networks by reducing the duty cycle of border nodes. These border nodes act as shared nodes between virtual clusters. Virtual clusters are formed on the basis of sleep/listen schedule of nodes. Towards this end, not only our proposed approach allows border nodes to join cluster where they experience minimum energy drain. Indeed, it stops the formation of abrupt communication holes due to hastily burning of border nodes. Thereby, prolonging network life and preventing disgraceful segmentations of wireless sensor network. To validate the proposed approach, a java based custom simulator is implemented. The results demonstrate, quantitatively, the trade of between energy consumption and node placement to prolong the life time of border nodes in S-MAC.

1 Introduction

Wireless Sensor Networks (WSN) enables pervasive, ubiquitous and seamless communication with the physical world. A few common applications are military, security, habitat monitoring, industrial automation and agriculture [1]. WSN comprises numerous sensor devices commonly known as motes, to monitor the physical entities such as temperature, light, motion, objects and humidity [11].

M. Takizawa, L. Barolli, and T. Enokido (Eds.): NBiS 2008, LNCS 5186, pp. 189–201, 2008.
© Springer-Verlag Berlin Heidelberg 2008

The characteristics of WSNs differ from traditional wireless networks in several ways [1]. Firstly, sensor networks consist of a number of nodes and have high network density that competes for the same channel. Secondly, nodes in sensor networks are battery powered, and it is often very difficult to replenish their batteries. Thirdly, nodes are often deployed in an ad hoc fashion rather than with careful pre-planning; thus, they are self-organized to form communication network. Fourthly, sensor networks are adaptable to local failures. Fifthly, broadcasting is the main mode of communication in sensor networks and this may cause channel contentions. Finally, most traffic in the network is triggered by sensing events, and it can be extreme at times. These characteristics of WSNs suggest that traditional MAC protocols are not suitable for wireless sensor networks without modifications.

This paper presents our approach to the expansion of Sensor-Medium Access (S-MAC) protocol [3]. It is a cluster based contention protocol to minimize energy consumption among Border Nodes (BN). The S-MAC protocol is based on unique feature; it conserves battery power by powering off nodes that are not actively transmitting or receiving packets. In doing so, nodes also turn off their radios. The manner in which nodes power themselves off does not influence any delay or throughput characteristics of the protocol. Inspired by the energy conservation mechanism of the S-MAC, we unmitigated our efforts to augment the node life time in sensor network. In WSN, border node acts as shared node between two or more virtual clusters. Virtual clusters are formed on the basis of sleep-listen schedule of nodes as shown in Fig. 1. Border nodes follow the schedule of one or more virtual clusters to which they belong as shown in Fig. 2, hence consume more energy due to extended duty load. In our proposed approach, border nodes are allowed to intelligently switch between virtual clusters where they experience minimum energy drain. Towards this end, we construct an analytical model which is the focus of this paper to verify the energy consumption and performance of our proposed multi-agent system at each node. It is further validated using our custom based java simulation. This system includes two types of agents, namely: stationary agent and mobile agent. The stationary agent is a static agent that has the capability of monitoring the changes of its energy consumption for a predefined period of time. The mobile agent is able to travel and interact with stationary agents at neighbouring clusters and benefit from their acquired knowledge.

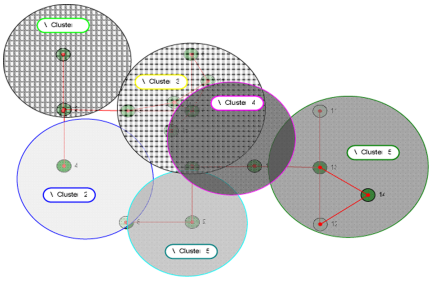


Fig. 1. Border node

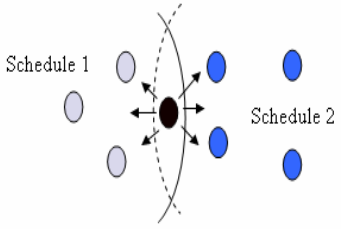


Fig. 2. Virtual cluster

2 The S-MAC

The operation of wireless network depends, to a large extent, on the effectiveness of the low-level Medium Access Control (MAC) layer. MAC in WSN aims to ensure that no two nodes are interfering with each other's transmissions. S-MAC is a slotted-based MAC protocol specifically designed for wireless sensor networks. S-MAC is built on contention-based protocols similar to IEEE 802.11 [3]. This protocol retains the flexibility of contention-based protocols while improving energy efficiency in multi-hop networks. It implements an approach to reduce energy consumption from all major factors, including idle listening, collision, overhearing and control overhead.

S-MAC Philosophy

The S-MAC contention-based protocol addresses not only the transmission interfering issues, but also intends to minimize the protocol-overhead, overhearing and idle-listening. Its principle is based on locally managed synchronizations and periodic sleep-listen schedules. Neighbouring nodes form virtual clusters based on sleep schedules. If two neighbouring nodes reside in two different virtual clusters, they should be in duty at listen periods of both clusters. Nodes synchronize their schedules by exchanging Synchronization (SYNC) packets. A node broadcasts the SYNC packet that contains its planned sleep time. The time allotted for sending the SYNC packet is called synchronization period.

Communication Structure

In S-MAC, the communication between nodes is typically achieved by exchanging packets that start with Carrier Sense (CS) to avoid collision. This is followed by Ready to Send and Clear to Send (RTS/CTS) packets which are unicasted to win the communication media, as shown in Fig. 3. Upon success, nodes start transmission of the desired data. Since all immediate nodes have their own sleep schedules, periodic sleep may result in high latency especially in multi-hop routing algorithm. The latency caused by periodic sleeping is called sleep delay.

An adaptive listening technique is proposed to improve the sleep delay, and thus the overall latency [3]. In their approach, a node which overhears its neighbour's transmissions listens for a short time at the end of the transmission. Hence, if a node is the next-hop node, its neighbour could pass data immediately. The end of the transmissions is known by the duration field of RTS/CTS packets. The energy waste caused by idle listening is reduced by sleep schedules. Broadcast data packets do not use RTS/CTS, which increases the probability of collision.

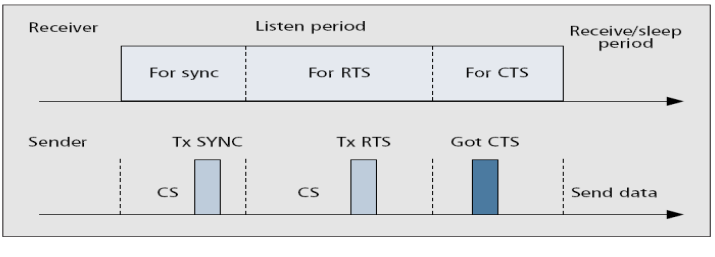


Fig. 3. S-MAC frame structure

S-MAC Dilemma

One of the major problems of S-MAC is the possibility of a node to follow two or more different schedules. It goes to listen-state frequently in order to relay packets from one virtual cluster to another. Consequently, it drains more battery power.

3 Related Work

Many researchers have proposed different ideas and solutions to minimize the energy consumption of border nodes in S-MAC. For example, the work presented by Ye and Heidemann [3], the border node adapts the first received schedule. According to their approach, a node can still communicate with neighbours in other clusters using its scheduled table containing its neighbouring nodes' schedules. In their approach, the cost of transmitting more than once to different clusters is overlooked, i.e. a border node has to transmit the corresponding SYNC packet to each virtual cluster whose schedule is being followed. In some cases, this is randomly selected schedule which is sent out to neighbouring nodes as SYNC packet. It is also possible to follow a SYNC packet received from a neighbour node, first received schedule, immediately after dispatching its own SYNC packet.

The work proposed by Lee et al. [4], a border node selects one schedule that belongs to the higher synchronizer node. This is due to the intuitive reason; the higher synchronizer node has been on network longer and possibly may have other followers. The border node creates and unicastes a unifying packet, which contains the highest synchronizer' schedule to be used as a target schedule. As soon as the neighbour node in the respective virtual cluster receives this unifying packet, it becomes the border node. This neighbour will attempt to achieve its non-border status by further transmitting the unifying packet to its immediate neighbours. This process repeats for each neighbour receiving the unifying packet, until there are no more border nodes left in the cluster. Lee et al. [4] has attempted to merge the virtual clusters having border nodes to form one unified large cluster over a single schedule. Formation of large virtual clusters lead to degradation of performance at the cost of missed events, i.e. sleeping of all the nodes over large spanned cluster at one time.

4 Proposed Approach

Inspired by the energy conservation mechanism of S-MAC, we sorted our efforts to enhance the node life time in sensor network. Border nodes act as shared nodes between virtual clusters. Virtual clusters are formed on the basis of sleep/listen schedule of nodes. Towards this end, this paper proposes an agent system that allows nodes to join cluster where they experience minimum energy drain. This system includes two types of agents, including: stationary and mobile agents.

The main objective of these agents is to help border nodes to move to virtual clusters where they experience minimum energy drain. Each node is equipped with Stationary Agent (SA) that closely examines the energy consumption in sensor node, and a Mobile Agent (MA) that frequently visits the border node's neighbours, with the ability to clone itself. The mobile agent analyzes its neighbour node's energy

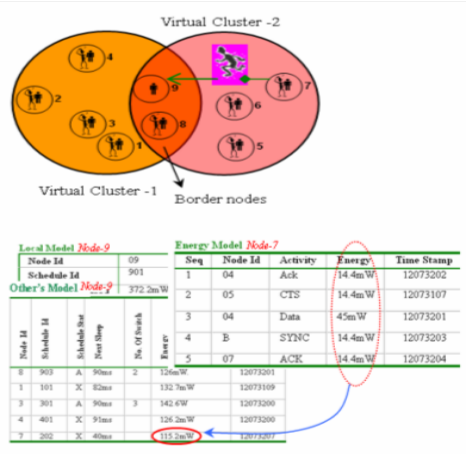


Fig. 4. Energy model update

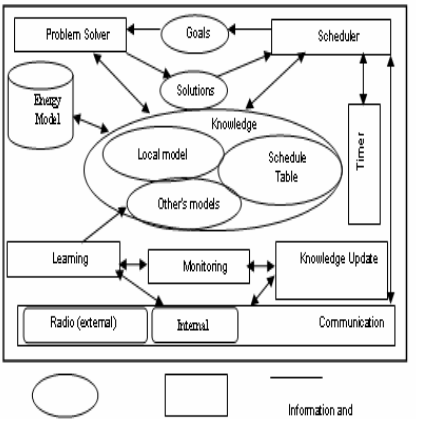


Fig. 5. Agent's architecture

model (see Fig. 4) in a predefined period, and then reports back to border node's stationary agent (i.e. the source agent). The stationary agent compares between the border node local energy model and others' energy model reported by MA to predict the energy efficient cluster. The virtual clusters may span over different geographical regions. The border node then can shift itself to energy efficient virtual cluster and become non-border node for amount of time until it finds more energy efficient cluster. The mobile agent frequently visits the host neighbour to query its energy model and reports to the stationary agent with the updated energy model. We only provide a high level overview of agent's energy models and schedule table. Full details on building the nodes energy model and maintaining cluster schedule can be found in [10].

4.1 Agent's Architecture

The architecture of the proposed agents is based on the agent model described in [10]. Each agent posses the basic structure, as shown in Fig. 5. This similar structure shared by both MA and SA helps to reduce energy and protocol overhead during MA transfers to its neighbouring nodes. It allows only the states of MA along with its local model to be transferred using as little as 250 bytes, as shown in Table 1. Once MA reaches to neighbouring nodes, it profits from inheriting the similar structure of host agent to continue its job. The agent architecture consists of knowledge components and executable components. The knowledge components contain the information about the WSN environment such as the number of clusters, neighbour nodes, goals that need to be satisfied and possible solutions generated to satisfy a goal. The learning component provides the agent with the capability to learn; it uses the monitored observations stored in its knowledge to predicate the energy efficient cluster. The scheduler component provides the agent with a time agenda to start and stop certain activities such as monitoring and mobility. The communication component allows the agent to exchange messages with other agents and with event occurring in a node. The two proposed SA and MA agents are the subject of the following two subsections.

4.2 SA and MA Process Flow

The agents' flow control to reduce the energy consumption of the border nodes is described in Fig. 6. Each sensor node is equipped with stationary and mobile agent. Once the node is deployed, the sensing region for the SA comes into effect. The SA will check to see if there exists any listen/sleep schedule. It does so, by looking at the value associated with SYNC variable in its knowledge. If no schedule is found the stationary agent initializes the set of tables stored in its knowledge. Meanwhile, the wireless sensor node starts listening for possible SYNC packet from its neighbouring nodes. It will listen for SYNC packet for the full frame length for the first time it joins the network. If no SYNC packet is received for the full frame length of time it will randomly choose its own schedule, i.e. the time to go to sleep. It will then broadcast the duration after which it will go to sleep in SYNC packet. Once SYNC packet is sent, the node will start its sleep timer corresponding to the duration value broadcasted in SYNC packet. At this point, the node status is changed to Synchronizer. If the node receives a SYNC packet from any of its neighbours before its sleep timer expires, the node adds that schedule to its schedule table. At this point, the node has two different schedules to follow. First, its own schedule that it broadcasted. Second, the received schedule from it neighbour. A node can not drop of its broadcasted schedule due to the reason that any new nodes joining the network may have synchronized with it.

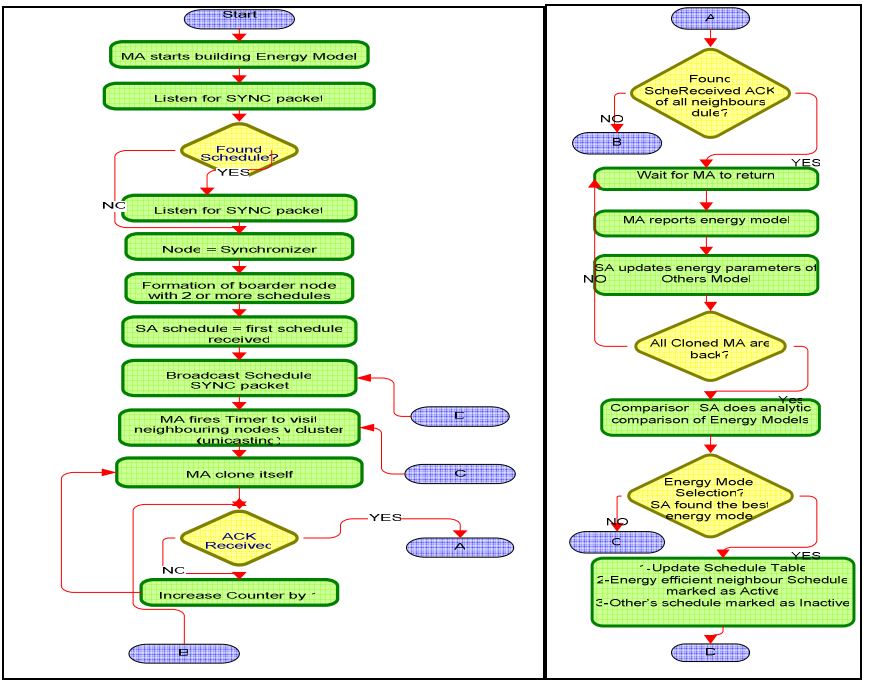


Fig. 6. Flow of control

The SA marks the first schedule to be active in its knowledge that was broadcasted by itself. The neighbouring nodes frequently exchange their identification (node_id) and their schedule identification (schedule_id) by a simple "hello" packet broadcast. A node goes to sleep once its sleep timer is expired. It also sets up its wake timer before it goes to sleep. Once a node wakes up, it broadcast its SYNC packet. The SA then initializes and starts the MA "visit interval" timer. The duration of the visit interval time is both user and application dependent. A realistic value set for the visit interval i.e. 1800 sec for agricultural or weather monitoring applications will not impact the saturation of WSN bandwidth.

The value associated with the visit interval defines the visit frequency of MA to visit neighbouring nodes. Once "visit interval" timer is fired, the SA sets up the "clone counter" for MA. The value of the "clone counter" represents the number of neighbours in a table stored in its knowledge called "Other's Model". When MA is dispatched to the neighbour, the SA waits for the acknowledge packet from that neighbour node. Once the acknowledgment is received, it decreases the clone counter by 1 and MA is dispatched to the next neighbour. As soon as all MAs are dispatched, the SA initializes the return counter. At the host neighbouring nodes, the MA profit from the knowledge acquired by the host SA. The MA queries the host node "Local Model" for its energy consumption. The "Local Model" is also a table in SA knowledge. After acquiring the energy consumption model of host node, SA makes its way back to source node via unicast. Upon arrival of MA back to source node, the SA decrements the return counter for the corresponding host node.

SA then updates the "Other's Model" from the information reported by MA in its knowledge. Once all the MAs are back, the SA makes an analytical comparison between reported data in other's model and its local model. If the reported data for other schedule outperforms the energy efficiency of the node, it will mark that schedule active in "Other's Model" and will override it with the current schedule in "Local Model". This node will start following the energy efficient schedule. MA will continue visiting the neighbouring nodes and reporting data back to SA. At any point of time when SA finds other schedule out performing its own, it will again switch its schedule.

4.3 Radio Communication Model

The communication model utilized in this paper is similar to that presented in [5]. The model employees Embedded Sensor Board (ESB), that is a prototype wireless sensor network device developed at Freie University Berlin [6]. The ESB consists of a Texas Instruments MSP430 low-powered microcontroller with 2k RAM and 60k flash ROM, a TR1001 radio transceiver, a 32k serial EEPROM, an RS232 port, a JTAG port, a beeper, and a number of sensors (passive IR, active IR sender/receiver, vibration/tilt, microphone, temperature). In this model, the energy E_r spent while in a given radio state r to transmit or receive a message is calculated using Equation (1).

$$E_r = P_r \times T_r \tag{1}$$

Where, P_r is the consumed power during time T_r in a give radio state r, and it is calculated using Equation (2).

$$P_r = V_r \times I_r. \quad (2)$$

Where, V_r is the voltage applied and I_r is the current induced at a give radio state r. For a shorter distance transmission such as within source, the energy consumed by a transmit amplifier is proportional to d^2 [7] where d is the distance between nodes. The energy consumed ET_{ij} by an agent to carry a message of length l bit from a node i to a node j is given by Equation (3).

$$ET_{ij} = lEe + lEs\, d_{ij}^2 \quad (3)$$

Moreover, the energy consumed ER to receive the l bits message is given by Equation (4):

$$ER = lEe + lE_{BF}. \quad (4)$$

In our quantitative analyzes, we used the following variables and constants as described in Table 1.The following subsections explain how all the values of all variables are derived and assigned, as well as the constants for the radio model.

Table 1. Radio model

Parameters	Units
Radio bandwidth	19,200 bits/s.
Control packet length	10 bytes
Size of aggregated data	242 bytes
MAC header length	8 bytes
Sensed packet size	50 bytes
Sensed packet interval	1 s
Size of mobile agent	250 bytes
Size of static agent	900 bytes
Energy consumed in the electronic circuit to transmit or receive signal	0.10125mWs
Energy consumed for transmitting 1 data packet	13.5mWs
Basic energy consumption PCL	12.0mA
Additional energy consumed for sending PTX	12.0mA
Additional energy consumed for receiving PRX	4.5mA
Total energy needed in sleep mode PSL	0.008mA

4.4 Energy Consumed by Electrical Circuits (E_e)

The MSP430 is running at a clock rate of 8MHz which means that a single cycle takes 1/8×106 seconds. If we assume that the average instruction takes 3 clock cycles and that we need 5,000 instructions for the measurement, for data processing and for preparing a packet for transmission over the network, we would be able to calculate the time T spent by electronic circuit for inter-node processing as follows:

$$T_{processing} = (5000 \times 3)/(8 \times 106) = 15/(8 \times 103) = 0.001875 \text{ seconds.}$$

Now, E_r can be calculated in mille watt seconds, as follow:

$$4.5v \times 12mA \times 0.001875 = 0.10125mWS.$$

4.5 Energy Consumed for Data Transmission (E_t)

To calculate the energy consumed for data transmission, we assumed that data resulting against interest, as presented in our previous work [8], by the sensor node consists of 50 bytes. Meanwhile, sending data takes place at a speed of 19,200 bits/s. A node has to forward data coming from other sensor nodes that will eventually reach to sink. We also assumed that we have to forward an average of 242 bytes from the neighbouring nodes. This assumption is very optimistic given a set of few nodes. Additional overhead is incurred of 8 bytes of MAC header. To send one bit, it takes 1/9200 seconds. Thus, the following shows how the time required to send the whole packet can be calculated as:

$$\frac{(50 + 242 + 8) \times 8}{19200} = 0.125 \ s$$

If sending packet requires the processor to be in full operational mode, i.e. 12mA for a node in basic mode (processor and sensors switched on), and 12mA additional energy consumed for sending, we need (12+12) mA in total. Therefore, the time required and the energy in seconds can be calculated as follows:

$$0.125 \ s \times (12 + 12) mAs = 3 \, mAs$$
$$4.5V \times 3 \, mAs = 13.5 \, mWs \ .$$

4.6 Energy Consumed in Sleep Mode

In order to calculate the energy consumed in sleep mode, we need to calculate the sum of the time spent by the electronic circuit (i.e. 0.001875 seconds), and the time took the sensor node to transmit the packet of size 300 bytes. Thus, the sum is calculated as follows:

$$0.001875 + 0.125 = 0.127s$$

This shows that the node can sleep for about one second minus 0.127 which is equal to 0.873seconds. The ESB needs about 0,008mA in sleep mode according to the manufacturer [9], which totalling in 0.007mAs for the 873ms of idle time within every second.

5 Implementation and Simulation Results

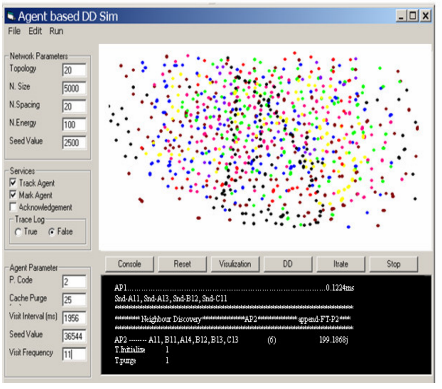
All graphical user interfaces, agents and visualizations are implemented as a simulator in Java. The user interacts with the agents using graphical user interfaces to submit queries, get results and tune some parameters. A network visualization interface is implemented to assist the user in validating energy efficiency of the proposed approach. The agents' components are implemented as objects. For example, the problem solver is implemented as a rule-based system, with a set of rules and some supporting classes, such as Agent-class, Network-class and Interest-class.

The main interface of the simulator is shown in Fig. 7-a. This interface is divided into input and output sections. The input section allows the user to input the desired network parameters, using text-fields. Using these text-fields, parameters can be adjusted dynamically by inputting new data such as the required topology for the simulation. The node population for deployment, i.e. node spacing is the minimum distance required between adjacent nodes regardless of how they are paced either randomly distributed or carefully placed based on topology parameter. The value of node spacing also marks the radio range on the node. Node energy represents the battery energy for every sensor deployed in the area. Seed value ensures that similar topology can be configured for each simulation run, for verification of consistent results. The lower left section is used to tune the agent parameters. Based on the demand of application the visit interval of MA can be fixed. P. Code refers to the size of MA agent. Cache purge is the amount of time SA will hold energy model table in its knowledge.

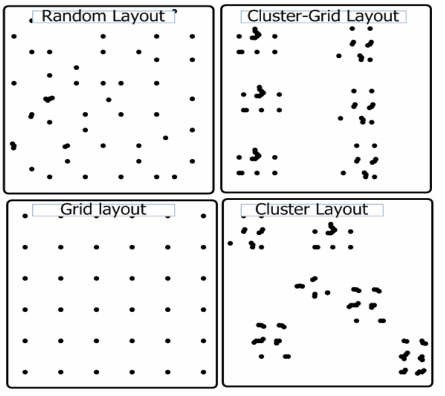
5.1 Discussion of Results

The experiments were conducted on a network size of 100 nodes, where energy parameter was kept constant through out the simulation. Node topology used by our experiments is shown in Fig. 7-b. We used random rectangle areas to simulate the increase and decrease of sensing. All the nodes in a rectangle area were selected and exposed to variable load. We performed our analysis on the energy consumption of the nodes in different layouts. The graphs shown in Fig. 7-c, 7-d, 7-e and 7-f reflect our analysis. The x-coordinate of the graph represents the time line. The y-coordinate represents the power consumed by the network. Fig. 3-d shows the average of 10 simulation results for Grid layout. Fig. 7-e also shows that our approach outperforms S-MAC. Although S-MAC performed slightly better during the initial simulation time, this is due to the time required for the SA to build its knowledge. From the obtained results of our experiments, cluster-grid topology was found to be most suitable for our approach. Our proposed approach performed worst when nodes are placed in cluster layout The variation in energy consumed with respect to MA's visit interval in cluster topology is demonstrated in Fig. 7-g. This figure shows that as the value of visit interval decreases will result into a decrease in energy consumption. Smaller the visit interval, frequent MA can visit neighbours and faster SA can built others model. Consequently, BN can make early and frequent decision to adapt to better neighbour. It should be noted that the visit interval is application dependent Fig. 3-g shows the effect of large visit interval on sensor nodes energy when arranged in grid topology. The x-axis shows the node density and time interval is shown along y-axis. We can see the smooth liner decay in nodes energy as that of abrupt decay in Fig. 7-g. This is due to the fact that there is no sudden connectivity holes created into network when nodes are carefully placed in planned manner. The energy decay of node is linearly proportional to the mobile agents visit interval.

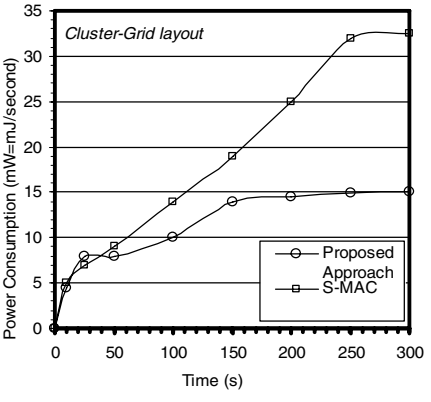
This is due to the fact that border nodes located in the region facing heavy loads are exposed for longer time before they can make a switch to cluster where there is less energy drain, i.e. are not exposed to extreme sensing. The last experiment was performed to compare the network health between our proposed approach and S-MAC. This experiment was conducted for 30 nodes over 10 simulation cycles.



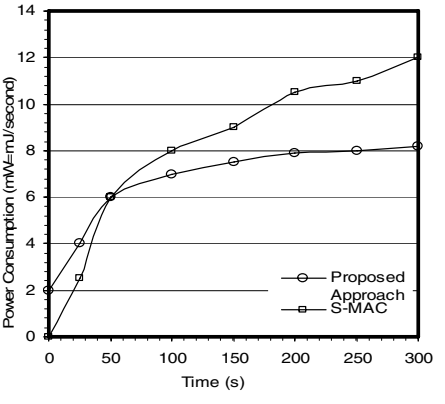
(a) Simulator main interface



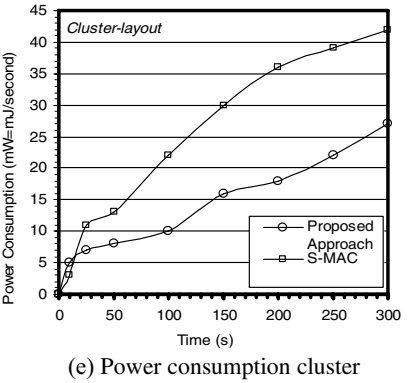
(b) Various deployment layouts



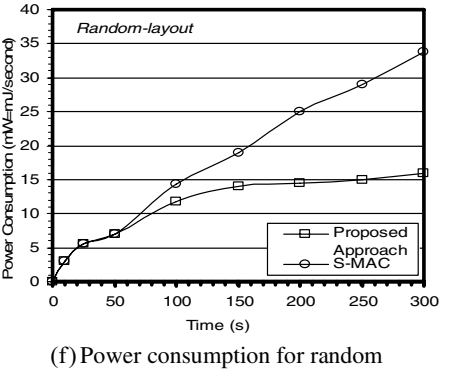
(c) Power consumption for C-grid



(d) Power consumption for grid

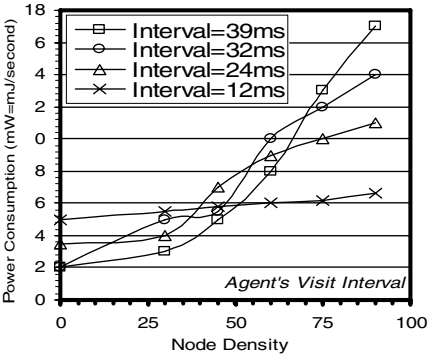


(e) Power consumption cluster

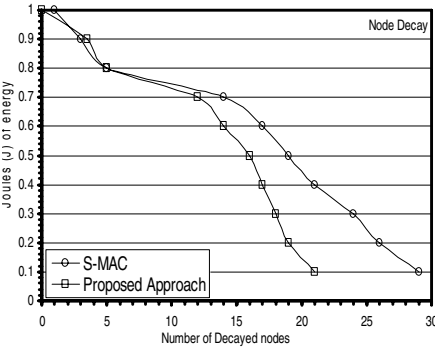


(f) Power consumption for random

Fig. 7. Graphical user interfaces



(g) Visit interval and energy



(h) Node decay

Fig. 7. (*continued*)

Each cycle ran for 120 seconds. Total network energy allocated to network was one joule and was evenly distributed among all nodes. Fig. 3-h shows that on average 17 nodes are expired using the proposed approach, where S-MAC burned 26 nodes.

6 Conclusion and Future Work

This paper identified the energy consumption issue in S-MAC a contention based medium access control protocol and proposed a mechanism to address this issue for wireless sensor networks. The mechanism compromises of multi-agent system to support border node in joining most energy efficient cluster among all virtual clusters it is member of. To demonstrate the effeteness of the proposed approach, we constructed an analytical model for the energy consumption of border nodes. The results produced by our custom java simulator were validated against our model and showed that our proposed approach performed better than S-MAC in terms of energy consumption. In future, we plan to install agent platform in sensor motes. We yet have to study the effect and cost of installing agent platforms. There exist several commercial agent platforms such as FIPA, IBM Aglet, and Voyager [12]. However, they have not yet been deployed and tested for wireless sensor network.

References

[1] Akyildiz, I.F., Su, W., Sankarasubramaniam, Y., Cayirci, E.: A survey on sensor networks. IEEE Communications Magazine 40(8), 102–116 (2002)

[2] Sohrabi, K., Pottie, G.: Performance of a novel self organization protocol for wireless ad hoc sensor networks. In: Proceedings of the IEEE 50th Vehicular Technology Conference, pp. 1222–1226 (1999)

[3] Ye, W., Heidemann, J.: Medium Access Control in Wireless Sensor Networks. In: Raghavendra, C.S., Sivalingam, K., Znati, T. (eds.) Wireless Sensor Networks, pp. 73–92. Kluwer Academic Publishers, Dordrecht (2004)

[4] Lee, W., Lee, D., Lee, H.: Lifetime extension of border nodes in SMAC-based wireless sensor networks by unifying multiple sleep schedules among adjacent virtual clusters. In: PE-WASUN 2005, pp. 267–268 (2005)
[5] Thomas: An FDL'ed Textbook on Sensor Networks (GNU FDL) (2006)
[6] Adams, D., http://www.sics.se/~adam/contiki/contiki-2.0-doc/a00435.html
[7] Heinzelman, W.R., Chandrakasan, A., Balakrishnan, H.: Energy efficient communication protocol for wireless microsensor networks. In: Proceedings of the IEEE Hawaii International Conference on System Sciences (January 2000)
[8] Shakshuki, E., Malik, H., Xing, X.: Agent based Routing for Wireless Sensor Network. In: The 2007 International Conference on Intelligent Computing (ICIC 2007). IEEE Computational Intelligence Society, Qingdao, China (2007)
[9] Marcel, B., Reiner, K., Clemens, M.: A Modular Platform for Sophisticated Real-time Wireless Sensor Networking, TR 399 (2005), http://www5.informatik.uni-wuerzburg.de/publications/techreports/
[10] Shakshuki, E., Malik, H.: Agent Based Approach to Minimize Energy Consumption for Border Nodes in Wireless Sensor Network. In: AINA Proceedings of the 21st International Conference on Advanced Networking and Applications, Niagara Falls, Canada, pp. 134–141 (2007)
[11] Culler, D., Estrin, D., Srivastava, M.: Overview of sensor networks. IEEE Computer 37(8) (August 2004)
[12] Bauer1, M., Elst1, L., Fischer1, K., Freisleben, B., Funk, P., Paul1, G., Thomas, S., Vogler, H.: AgentSurvey: Assessing the state of the Art of Industrial Applications Using Agent Technology and AI. In: The Fifteenth Canadian Conference on Artificial Intelligence (AI), Calgery, Alberta (2002)

A Self-organising Network Based on Lightweight Agents

John Debenham and Ante Prodan

Faculty of Information Technology,
University of Technology, Sydney, Australia
{debenham,aprodan}@it.uts.edu.au

Abstract. A lightweight multiagent system is deployed at each node in a communications network with the aim of self-organising the network as usage alters. The distributed, light-weight, co-operative multiagent system guarantees scalability of the approach. As the solution is distributed it is unsuitable to achieve any global optimisation goal — it simply seeks to continually improve network performance as demands change. Algorithms are described for adjusting the communication channels and for adjusting the network links. Experiments show that the method is robust and delivers good performance.

1 Introduction

The work discussed is based on previous work in the area of mesh networking and in particular in distributed algorithms at Columbia University, Microsoft Research, University of Maryland and Georgia Institute of Technology. In particular: [1], [2], [3] and [4]. The system described is fully distributed across the network with each node autonomously acting on the basis of signals that it observes. The actions at each node are determined by its proactive and reactive reasoning — in this sense the solution is agent-like although it does not exhibit all of the characteristics of an intelligent multiagent system. The term *lightweight agent* refers to this level of functionality. There are three principal inputs to this work that we assume are available to the proposed methods:

- A load model. Given any contiguous set of nodes in a mesh, the *load model* specifies the actual or desired level of traffic flowing into, or out of, nodes in that set.
- A load balancing algorithm. Given any contiguous set of nodes in a mesh and the load model for that set, the *load balancing algorithm* determines how the traffic is allocated to links in the mesh so as to reach its desired destination.
- An interference model. Given any contiguous set of nodes in a mesh, the *interference model* stipulates the interference level that each node in the mesh gives to the other nodes in the mesh given a known level of background interference due to transmission devices that are external to the mesh.

The work described below makes no restrictions on these three inputs other than that they are available to every node in the mesh. The load model, and so too the load balancing algorithm, will only be of value to a method for self-organisation if together

M. Takizawa, L. Barolli, and T. Enokido (Eds.): NBiS 2008, LNCS 5186, pp. 202–211, 2008.
© Springer-Verlag Berlin Heidelberg 2008

they enable future load to be predicted with some certainty. We assume that the load is predictable.

Below we introduce some terms, concepts and notation. Section 2 describes the role of the load balancing algorithm that our methods take as a given input. The measurement of interference cost is discussed in Section 3. Methods for the adjusting the channels in a multi-radio mesh networks for predictable load are described in Section 4, and for adjusting the links in Section 5. Future plans are described in Section 6.

The discrete time intervals mentioned below, e.g. t, $t+1$, are sufficiently spaced to permit what has to be done to be done.

A *node* is a set of radio interfaces (or "antennae") where each *interface* is associated with a particular *channel*, together with a controller that (intelligently we hope) assigns the channel on each interface. Interfaces that are part of the same node are assumed to be 'close' topologically, but this is not important. We assume for simplicity that each interface has its own, independent MAC layer.

A *link* is a pair of interfaces where each interface is assigned the same channel. The idea is that two interfaces communicate through a shared link. That is, if an interface is part of a link its state will be "listening and transmitting", otherwise its state will be "listening only".

Notation: nodes are denoted by Latin letters: a, b, c,..., the interfaces for node a are denoted by: $a[i]$ for $i = 1, \ldots$, and links are denoted by Greek letters: α, β, γ,.... The interfaces communicate using an illocutionary communication language that is defined informally (for the time being) with illocutions being encapsulated in quotation marks: "$\cdot$".

For any node n, S_n is the set of nodes in node n's interference range. Likewise, for any link α, S_α is the set of links that contain nodes n's interference range $\forall n \in \alpha$.

Given a node a, define $V_a = \cup_{n \in S_a} S_n$.

Γ_x^t is channel used by x to communicate at time t where x may be either an interface or a link.

$f(\cdot, \cdot)$ is an *interference cost function* that is defined between two interfaces or two links. It estimates the cost of interference to one interface caused by transmission from the other interface. This function relies on estimates of the interference level and the level of load (i.e.: traffic volume). So this function requires an *interference model* and a *load model*. This function is described in Section 3.

An interface is either 'locked' or 'unlocked'. A locked interface is either locked because it has committed to lock itself for a period of time on request from another interface, or it is 'self-locked' because it has recently instigated one of the self-organisation procedures in Section 4.

The abbreviation SNIR means "signal to noise plus interference ratio".

2 The Load Balancing Algorithm

We assume that if the external demands on a set of nodes S are known and that there is a *load balancing algorithm* — that may or may not be intelligent — that determines

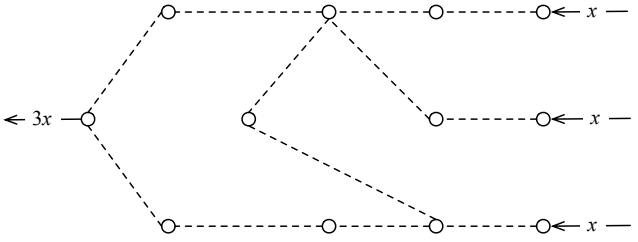


Fig. 1. The load balancing algorithm determines the allocation of load

how the load is routed through S. Figure 2 shows a set of twelve nodes connected by a mesh that is shown as dashed lines. The load on the mesh is shown by the four solid arrows. We assume that the load balancing algorithm will determine how the load is allocated to the links in the mesh.

3 Measuring Interference Cost

Suppose that during some time interval Δt two interfaces a and b are transmitting and receiving on channels Γ_a and Γ_b. During Δt, the *interference limit* that interface x imposes on interface y, $\tau_{y|x}$, is a ratio being the loss of traffic volume that interface y could receive if interface x were to transmit persistently divided by the volume of traffic that interface y could receive if interface x was silent:

$$\tau_{y|x} = \frac{(m_y \mid \text{ interface } x \text{ silent}) - (m_y \mid \text{ interface } x \text{ persistent})}{m_y \mid \text{ interface } x \text{ silent}}$$

where m_y is the mean SNIR observed by interface y whilst listening on channel Γ_y, where as many measurements are made as is expedient in the calculation of this mean[1]. The *interference load* of each interface, v_a and v_b, is measured as a proportion, or percentage, of some time interval during which that interface is transmitting. Then the *observed interference* caused by interface b transmitting on channel Γ_b as experienced by interface a listening on channel Γ_a is: $\tau_{a|b} \times v_b$, and the *observed interference cost* to interface a is[2]:

$$f(a \mid b) \triangleq \tau_{a|b} \times v_b \times (1 - v_a)$$

and so to interface b:

$$f(b \mid a) = \tau_{b|a} \times v_a \times (1 - v_b)$$

Now consider the interference between one interface a and two other interfaces c and d. Following the argument above, the *observed interference* caused by interfaces c

[1] For $\tau_{y|x}$ to have the desired meaning, m_y should be a measurement of *link throughput*. However, link throughput and SNIR are approximately proportional — see [5].

[2] We assume here that whether or not interfaces a and b are transmitting are independent random events [6]. Then the probability that a is transmitting at any moment is v_a, and the probability that b is transmitting and a is listening at any moment is: $(1 - v_a) \times v_b$.

and d as experienced by interface a is[3]: $\tau_{a|c} \times v_c + \tau_{a|d} \times v_d - \tau_{a|\{c,d\}} \times v_c \times v_d$. The observed interference cost to interface a is:

$$f(a\,|\{c,d\}) = (1 - v_a) \times \left(\tau_{a|c} \times v_c + \tau_{a|d} \times v_d - \tau_{a|\{c,d\}} \times v_c \times v_d\right)$$

Suppose that v_β is the proportion of Δt for which either interface c or interface d is transmitting. Then for some κ_β, $0 \leq \kappa_\beta \leq 1$: $v_c = \kappa_\beta \times v_\beta$, and $v_d = (1 - \kappa_\beta) \times v_\beta$. Thus:

$$f(a \mid \beta) = (1 - v_a) \times v_\beta \times \left(\tau_{a|c} \times \kappa_\beta + \tau_{a|d} \times (1 - \kappa_\beta)\right)$$

Now suppose that interfaces a and b are linked, and that v_α is the proportion of Δt for which either interface a or interface b is transmitting. Then for some κ_α, $0 \leq \kappa_\alpha \leq 1$: $v_a = \kappa_\alpha \times v_\alpha$, $v_b = (1 - \kappa_\alpha) \times v_\alpha$. Then as a will only receive interference when it is listening to b transmitting:

$$f(a \mid \beta) = v_b \times v_\beta \times \left(\tau_{a|c} \times \kappa_\beta + \tau_{a|d} \times (1 - \kappa_\beta)\right)$$

and so:

$$\begin{aligned} f(\alpha \mid \beta) =& (1 - \kappa_\alpha) \times v_\alpha \times v_\beta \times \left(\tau_{a|c} \times \kappa_\beta + \tau_{a|d} \times (1 - \kappa_\beta)\right) \\ &+ \kappa_\alpha \times v_\alpha \times v_\beta \times \left(\tau_{b|c} \times \kappa_\beta + \tau_{b|d} \times (1 - \kappa_\beta)\right) \end{aligned} \tag{1}$$

Note that v_α, v_β, κ_α and κ_β are provided by the load model, and the $\tau_{x|y}$ are provided by the interference model.

4 Adjusting the Channels

Our solution is based on the distinction in multiagent systems between proactive and reactive reasoning. Proactive reasoning is concerned with planning to reach some goal. Reactive reasoning is concerned with dealing with unexpected changes in the agent's environment. So in the context of self-organising networks we distinguish between:

- A *reactive logic* that deals with problems as they occur. The aim of our reactive module is simply to restore communication to a workable level that may be substantially sub-optimal.
- A *proactive logic* that, when sections of the network are temporarily stable, attempts to adjust the settings on the network to improve performance.

The reactive logic provides an "immediate fix" to serious problems. The proactive logic, that involves deliberation and co-operation of nearby nodes, is a much slower process.

Informally the proactive logic uses the following procedure:

- *Elect* a node a that will manage the process
- *Choose* a link α from a to another node — precisely a trigger criterion (see below) permits node a to attempt to improve the performance of one of its links $\alpha \ni a$ with a certain priority level.
- *Measure* the interference
- *Change* the channel setting if appropriate

[3] That is, the interference caused by either interface c or interface d.

The following is a development of the ideas in [1].

choose node a **at** time $t-2$;
set $V_a = \cup_{n \in S_a} S_n$;
$\forall x \in V_a$ **transmit** "**propose organise**$[a, x, p]$";
unless $\exists x \in V_a$ **receive** "**overrule organise**$[a, x, q]$" in
$[t-2, t-1]$ where $q > p$ **do** {
$\forall x \in V_a$ **transmit** "**propose lock**$[a, x, t, t+1]$";
if $\forall x \in V_a$ **receive** "**accept lock**$[a, x, t, t+1]$" in $[t-1, t]$
then {
unless $\exists x \in V_a$ **receive** "**reject lock**$[a, x, t, t+1]$"
do {improve a;}
}
}
where: improve $a = \{$
choose link $\alpha \ni a$ on channel Γ_α^t;
set $B \leftarrow \sum_{\beta \in S_\alpha} f(\alpha \mid \beta) + \sum_{\beta \in S_\alpha} f(\beta \mid \alpha)$;
if (feasible) **re-route** α's traffic;
for $\Gamma_\alpha = 1, \ldots, K, \Gamma_\alpha \neq \Gamma_\alpha^t$ **do**{
if $\sum_{\beta \in S_\alpha} f(\alpha \mid \beta) + \sum_{\beta \in S_\alpha} f(\beta \mid \alpha) < B \times \epsilon$ **then**{
$\Gamma_\alpha^{t+1} \leftarrow \Gamma_\alpha$;
selflock node a **in** $[t+1, t+k]$;
break;
};
};
$\forall x \in V_a$ **transmit** "α's interference test signals";
apply load balancing algorithm to S_a;
}

The statement **selflock** is to prevent a from having to activate the method too frequently. The constant $\epsilon < 1$ requires that the improvement be 'significant' both for node a and for the set of nodes S_a. The stability of this procedure follows from the fact that it produces a net improvement of the interference cost within S_a. If a change of channel is effected then there will be no resulting change in interference outside S_a.

Interference model. We assume that each node, a, knows the channel of every node in V_a. We assume that each node is capable of measuring the strength of signals from every node in V_a. So if each node had access to all of this information from the point of view of every node in V_a, and, perhaps the level of background noise around V_a then a can derive estimates for the $\tau_{x|y}$ factors for all x and y in V_a. In particular, a will be able to estimate all these factors to evaluate Equation 1 as required by the above algorithm. *In addition*, the procedure above suggests that if node a is involved in changing its channel then at the end of this process — time permitting — it should transmit a 'beep-silence-beep-silence' message to enable every other node in V_a to observe the actual τ values. *Further*, it is reasonable to suggest that this transmission of test signals could be carried out periodically in any case when network load permits.

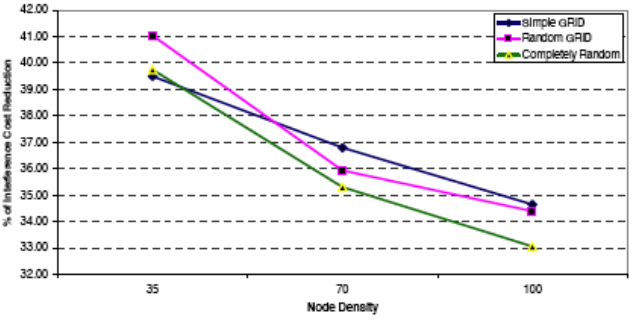


Fig. 2. Interference cost reduction as a function of node density

4.1 Results and Discussion

Impact of network (node) density on the performance. As the density of network increases (i.e. an increase in the number of routers located within the same area) the IC reduction relatively decreases. This trend is shown across all the topologies. The impact of node density on the algorithm is relatively consistent for all topologies at the same router densities. From Figure 2 it can also be observed that the range of the interference reduction across the topologies at router densities of 35 routers and 100 routers is 1.55 and 1.58, respectively.

Impact of typical topologies on the interference cost. Figure 3 shows the variation in the interference cost reduction as a function of network topology across different node densities. It can be deduced that the impact of the topologies on the performance of the algorithm (i.e. in terms of interference cost reduction) is insignificant. The mean of IC reduction calculated from the data obtained shows that the topology with the smallest average IC reduction is the completely random with a mean of 36.02 and topology with the most IC reduction is the random grid with a mean of 37.12. The difference in performance between best and worst case is just 1.1 which confirms that the performance of the algorithm is almost completely independent of the type of topology.

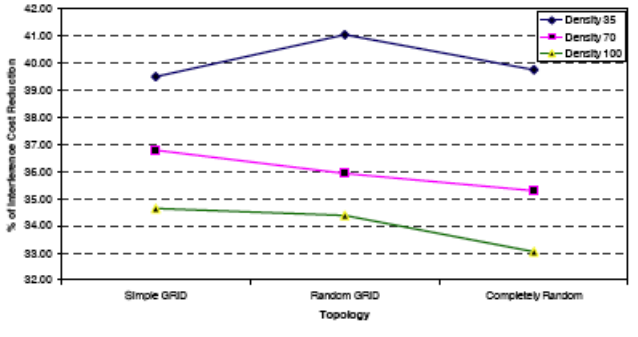


Fig. 3. Interference cost reduction as a function of topologies

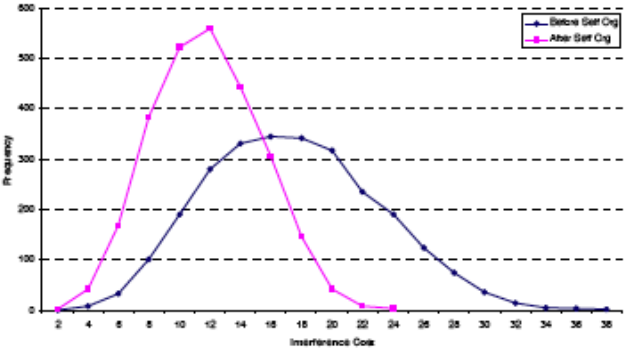


Fig. 4. Comparison of IC across the network before (blue) and after (red) selforganisation

Performance Comparison across the Network. In this study, we obtained interference cost (IC) in different regions of the MR-WMN for the same set of links before and after the self-organisation algorithm is invoked. Comparison of the results obtained is shown in Figure 4 where the Interference cost is on the X-axis. From Figure 4 we can see that there were no nodes (red dots) that caused more interference after the self-organisation than it had caused before (blue dots) the self-organisation was invoked.

5 Adjusting the Links

The algorithm for adjusting the links is precisely the same as the algorithm in Section 4 but with the following 'improve' methods.

Link adjustment with known traffic load. Suppose that node a has interference range S_a. Let M_a be the set of nodes in S_a excluding node a. Then use the method in Section 4 with the following 'improve' method:

improve $a = \{$
  for link $\alpha \ni a$, where $\alpha = [a, b]$
  suppose α is on channel Γ^t_α;
  set $B \leftarrow \sum_{\beta \in S_\alpha} f(\alpha \mid \beta) + \sum_{\beta \in S_\alpha} f(\beta \mid \alpha)$;
  if (feasible) **re-route** α's traffic;
  set $\gamma \leftarrow \alpha$;
  for $y \in M_a$ **do** $\{$
    for $\Gamma_{[a,y]} = 1, \ldots, K$, **do** $\{$
      if $\sum_{\beta \in S_a} f([a, y] \mid \beta) + \sum_{\beta \in S_a} f(\beta \mid [a, y]) < B \times \epsilon$
      then $\{$
        set $\gamma \leftarrow [a, y]$;
        selflock node a **in** $[t + 1, t + k]$;
        break;
      $\}$;
    $\}$;

};
$\forall x \in V_a$ **transmit** "γ's interference test signals";
apply load balancing algorithm to S_a;
}

Trigger for attempting to adjust a link with known traffic load. Consider a mesh with known traffic load such as that illustrated in Figure 2. Suppose that the load balancing algorithm has allocated load to links on the mesh, and let link (a, b)=$\arg\max_{x \in N_a^t} \rho(x)$. If replacing (a, b) with (a, x) would mean that there exists a cut through the mesh that traverses (a, x) and that all other links on that cut have a load $< \rho(a, b)$ then let node a initiate the link adjusting procedure. Likewise if replacing (a, b) with (y, b).

5.1 Reactive Logic

The relationship between the reactive and proactive logics is determined by:

if event [link α is broken] **then** {
activate [activate the *Reactive Method* for link α];
$\forall x \in \alpha$ **if state** [node x locked by "**accept lock**$[a, x, s, t]$"
then {**transmit** "**reject lock**$[a, x, s, t]$";}
}

where the *Reactive Method* is as follows; it simply fixes disasters as they occur possibly with a configuration that is less satisfactory than the prior. It has no implications for neighbouring interfaces, and so it presents no instability issues.

Reactive Method. Important assumption for the functioning of the reactive logic discussed here is that all interfaces capable of reactive reconfiguration use omnidirectional antennas. The benefits and shortcomings of the usage of different antennas are discussed in details in our previous report. Two interfaces connected through directional antenna behave similarly to a wired point to point link because they cannot connect to any other interface to which their antennas are not aligned. This does not represent an impediment for the proposed architecture since majority of nodes will be equipped with omnidirectional antenna.

For the implementation of reactive logic we propose usage of simple mechanisms that are derived from routing protocols recently developed for stationary multi-radio mesh networks [4]. In conjunction with an appropriate routing protocol these mechanisms should ensure high reactivity in minimising effect of link interruptions caused by various factors.

Link adjustment with unknown traffic load. Suppose that node a has interference range S_a. Let M_a be the set of nodes in S_a excluding node a. For nodes $x, y \in S_a$, let $c(x, y)$ denote the cost[4] of the least cost path that connects x and y. We assume that: $(\forall x, y) c(x, y) = c(y, x)$, and that if the least cost path between nodes u and v is a

[4] The precise meaning of this cost function does not matter. It could be simply the number of hops, or some more complex measure involving load and/or interference.

subset of the least cost path between x and y then $c(u,v) \leq c(x,y)$. Let N_a^t be the set of links in S_a at time t, and $N_a^t(\ominus[a,x], \oplus[a,y])$ denotes the network configuration with link $[a,x]$ replaced by $[a,y]$. Let $C(N_a^t)$ denote the cost of the path of greatest cost in S_a: $C(N_a^t) \triangleq \max_{x,y \in S_a} c(x,y)$. Choose the pair of nodes b and c by:

$$(b,c) = \arg \min_{(x,y)|[a,x] \in N_a^t, y \in M_a} C(N_a^t(\ominus[a,x], \oplus[a,y]))$$

and swap link $[a,b]$ for link $[a,c]$ if:

$$C(N_a^t(\ominus[a,b], \oplus[a,c])) < C(N_a^t) \times \epsilon$$

where $\epsilon < 1$ is a threshold constant [7].

5.2 Results and Discussion

This part of study firstly proposes the method for the link substitution that results with the reduction of the path length. Secondly, to provide the insight in algorithms effectiveness we produce over 3000 simulations. The simulation results are statistically processed and the outcomes for 3 different densities (35,70 and 100) are obtained.

Simulation parameters. We have used a Java based framework to carry out the simulations for the results shown and discussed in this section. The key attributes of the simulation were:

- Number of interfaces per router was randomly selected from 3 to 5.
- Default signal strength was 100 mW (20 dBm — Signal strength for each interface was randomly generated with +/- 25% variation.
- Network size had an area of $750m \times 500m$

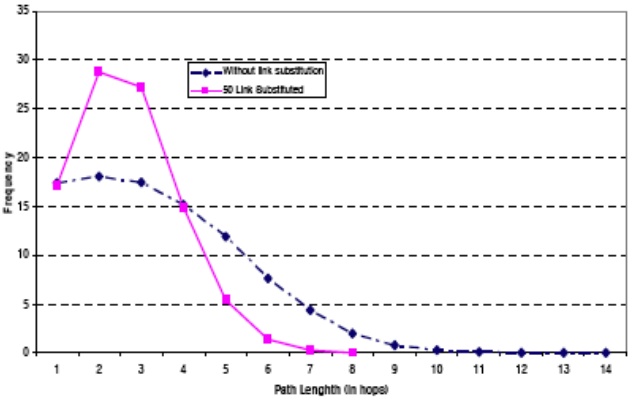


Fig. 5. Frequency distribution of the path length (in hops) without and with link substitution algorithm at 100 node network density and 10 additional links

In addition to the simulation parameters described above we limited the number of links to $n - 1$; where n is number of router (density) in a network. Consequently, the number of links created was 34,69 and 99 for the corresponding network densities. In addition to these link numbers we tested the effectiveness of the link substitution algorithm by creating additional 10 links when link substitution reached efficiency threshold. The number of the substituted link was limited in all simulation to (10, 20, 30, 40 and 50) and separate results are shown. We now compare path lengths with and without link substitution. From the Figure 5 we can observer that our method significantly reduces path length by eliminating longer paths (maximum path length is 8 with the link substitution and 14 without it). This method also increases the number of shortest path (in particular paths 2 and 3 hops long).

6 Conclusion and Future Work

Through the work described in this report we have examined motivation and developed an algorithm for the topological control of MR-WMN. The goal of this algorithm is to increase the number of shortest paths to the portal nodes without adversely effecting interference cost. In addition to interference cost reduction implementation of this algorithm on MR-WMN further improve the system capacity.

Our future work will be focused on the development of our Java framework that is multi threaded so each node is represented as an independent thread. We believe that this will enable us to develop algorithms for tuning the capacity of the network links according to fluctuations in demand by mobile users.

References

1. Ko, B.J., Misra, V., Padhye, J., Rubenstein, D.: Distributed Channel Assignment in Multi-Radio 802.11 Mesh Networks. Technical report, Columbia University (2006)
2. Mishra, A., Rozner, E., Banerjee, S., Arbaugh, W.: Exploiting partially overlapping channels in wireless networks: Turning a peril into an advantage. In: ACM/USENIX Internet Measurement Conference (2005)
3. Mishra, A., Shrivastava, V., Banerjee, S.: Partially Overlapped Channels Not Considered Harmful. In: SIGMetrics/Performance (2006)
4. Akyildiz, I.F., Wang, X., Wang, W.: Wireless mesh netowrks: a survey. Computer Networks, 445–487 (2005)
5. Vasudevan, S.: A Simulator for analyzing the throughput of IEEE 802.11b Wireless LAN Systems. Master's thesis, Virginia Polytechnic Institute and State University (2005)
6. Leith, D., Clifford, P.: A self-managed distributed channel selection algorithm for wlans. In: Proceedings of RAWNET, Boston, MA, USA, pp. 1–9 (2006)
7. Ramachandran, K., Belding, E., Almeroth, K., Buddhikot, M.: Interference-aware channel assignment in multi-radio wireless mesh networks. In: Proceedings of Infocom 2006, Barcelona, Spain, pp. 1–12 (2006)

A Fuzzy-Based Handover System for Wireless Cellular Networks: A Case Study for Handover Enforcement

Leonard Barolli[1], Arjan Durresi[2], Fatos Xhafa[3], and Akio Koyama[4]

[1] Department of Information and Communication Engineering
Fukuoka Institute of Technology (FIT)
3-30-1 Wajiro-Higashi, Higashi-Ku, Fukuoka 811-0295, Japan
barolli@fit.ac.jp
[2] Department of Computer and Information Science
Indiana University Purdue University Indianapolis
723 W. Michigan Street SL 280, Indianapolis, IN 46202, USA
durresi@cs.iupui.edu
[3] Department of Languages and Informatics Systems
Polytechnic University of Catalonia
Jordi Girona 1-3, 08034 Barcelona, Spain
fatos@lsi.upc.edu
[4] Department of Informatics, Yamagata University
4-3-16 Jonan, Yonezawa 992-8510, Yamagata, Japan
akoyama@yz.yamagata-u.ac.jp

Abstract. Presently, the wireless mobile networks and devices are becoming increasingly popular to provide users the access anytime and anywhere. The mobile systems are based on cellular approach and the area is covered by cells that overlap each other. Many handover algorithms are proposed in the literature. However, to make a better handover and keep the QoS in wireless networks is very difficult. In this paper, we propose a new handover system based on fuzzy logic. We consider the case of handover enforcement and verify that the proposed system makes the necessary handovers for all simulations.

1 Introduction

During the last few years wireless multimedia networks have been a very active research area [1,2]. The QoS support for future wireless networks is a very important problem. To guarantee the QoS, a good handover strategy is needed in order to balance the call blocking and call dropping for providing the required QoS [3,4]. In the future, the wireless networks will adopt a micro/pico cellular architecture. However, smaller cell size naturally increases the number of handoffs a Mobile Station (MS) is expected to make [5,6].

Many metrics have been used to support handover decisions, including Received Signal Strength (RSS), Signal to Interference Ratio (SIR), distance between the mobile and BS, traffic load, and mobile velocity, where RSS is the most commonly used one. The conventional handover decision compares the RSS from the serving BS with that from one of the target BSs, using a constant handover threshold value (handover margin).

M. Takizawa, L. Barolli, and T. Enokido (Eds.): NBiS 2008, LNCS 5186, pp. 212–222, 2008.
© Springer-Verlag Berlin Heidelberg 2008

However, the fluctuations of signal strength associated with shadow fading cause the ping-pong effect [7].

Recently, many investigations have addressed handover algorithms for cellular communication systems. However, it is essentially complex to make handover decision considering multiple criteria. Sometimes, the trade-off of some criteria should be considered. Therefore, heuristic approaches based on Neural Networks (NN), Genetic Algorithms (GA) and Fuzzy Logic (FL) can prove to be efficient for wireless networks [8,9,10,11]. In [10], a multi-criteria handover algorithm for next generation tactical communication systems is introduced. The handover metrics are: RSS from current and candidate base transceivers, ratio of used soft capacity to the total soft capacity of base transceivers, the relative directions and speeds of the base transceivers and the mobile node. In [11], a handover algorithm is proposed to support vertical handover between heterogeneous networks. This is achieved by incorporating the mobile IP principles in combination with FL concepts utilizing different handover parameters.

In this paper, in different from other works we use Random Walk (RW) model and FL to design a new handover system in order to make a good handover decision. The structure of this paper is as follows. In Section 2, we present the handover decision problem. In Section 3, we give a brief introduction of RW model. In Section 4, we introduce the proposed system. In Section 5, we discuss the simulation results. Finally, some conclusions are given in Section 6.

2 Handover Decision Problem

Handoffs which are consistently both accurate and timely can result in higher capacity and better overall link quality than what is available with today systems [12,13]. Now with increasing demands for more system capacity, there is a trend toward smaller cells, also known as microcells. Handoffs are more critical in systems with smaller cells, because for a given average user speed, handoff rates tend to be inversely proportional to cell size [5].

The main objectives of handover are link quality maintenance, interference reduction and keeping the number of handoffs low. Also, a handover algorithm should initiate a handoff if and only if the handoff is necessary. The accuracy of a handover algorithm is based on how the algorithm initiates the handover process. The timing of the handoff initiation is also important. There can be deleterious effects on link quality and interference if the initiation is too early or too late.

Because of large-scale and small-scale fades are frequently encountered in mobile environment, it is very difficult for handover algorithm to make an accurate and timely decision. Handover algorithms operating in real time have to make decisions without the luxury of repeated uncorrelated measurements or future signal strength information. It should be noted that some of handover criteria information can be inherently imprecise, or the precise information is difficult to obtain. For this reason, we propose a FL-based approach, which can operate with imprecision data and can model nonlinear functions with arbitrary complexity.

3 RW Model

We use the MC method for realizing RW model. We consider a 2-dimensional field. The initial position is considered as a origin point and we decided based on MC method the moving pattern for each walk. If we consider that n user movements, angle θ and distance d for each walk are generated by general or Gaussian distribution, when the movement changes in x and y directions are Δx and Δy, respectively, then we have the following relations.

$$\Delta x_n = d_n cos\theta_n, \quad \Delta y_n = d_n sin\theta_n \tag{1}$$

$$x_{n+1} = x_n + \Delta x_n, \quad y_{n+1} = y_n + \Delta y_n \tag{2}$$

The Base Station (BS) position can be expressed by Cartesian coordinates. By converting Cartesian coordinates to polar ones, we can calculate the angle θ.

We consider that in the cellular system each cell has a hexagonal shape and the BS is located in the center of the cell. The angle θ between Dipole Antenna (DA) and vector $\boldsymbol{r}$ is $D(\theta) = \sin\theta$. If we consider the transmission power as W, the antenna radiation intensity can be calculated as follows:

$$\boldsymbol{E} = \sqrt{45W}\sin\theta\frac{e^{-j\kappa r}}{r^n}\boldsymbol{u_0} \tag{3}$$

where, the DA gain is $G = 1.5$ and $\boldsymbol{u_0}$ is the unit vector that shows DA direction. In Fig. 1 $\boldsymbol{u_0}$ is in Z direction.

In Eq.(3), when $\theta = 90°$, the E value will be maximal in horizontal direction. However, in real situations, the direction of antenna is in not set $90°$ in order to cover better the cell area. If we consider the beam tilting angle and the distance, the E can be calculated by the following equation.

$$\boldsymbol{E} = \sqrt{45W}\sin(\theta - \phi)\frac{e^{-j\kappa r}}{r^n}\boldsymbol{u_0} \tag{4}$$

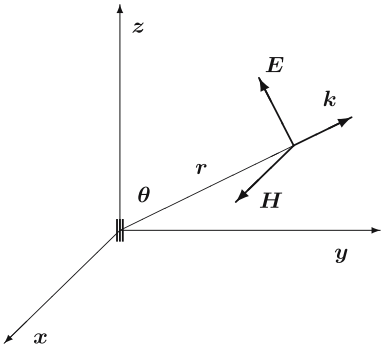


Fig. 1. Dipole antenna

4 Proposed System Model

The Fuzzy Logic Controller (FLC) is the main part of the proposed system and its basic elements are shown in Fig. 2. They are the fuzzifier, inference engine, Fuzzy Rule Base (FRB) and defuzzifier. As shown in Fig. 3, as membership functions we use triangular and trapezoidal membership functions because they are suitable for real-time operation [14].

In Fig. 3, x_0 in $f(.)$ is the center of triangular function; $x_0(x_1)$ in $g(.)$ is the left (right) edge of trapezoidal function; and $a_0(a_1)$ is the left (right) width of the triangular or trapezoidal function.

The proposed fuzzy model is shown in Fig. 4. In this system, the $Node\_B$ shows the wireless transmitter and receiver of BS, RNS indicates Radio Network System, POTLC stands for Post Test-Loop Controller and PRTLC for Pre Test-Loop Controller.

The input parameters for FLC are: Change of the Signal Strength of Present BS (*CSSP*), Signal Strength from the Neighbor BS (*SSN*), and the distance of MS from BS (*DMB*), while the output linguistic parameter is Handover Decision (*HD*).

The system operation is as follows. First, after receiving the control information from MS, the POTLC check the quality of the signal. If the signal strength is still good enough the handover is not carried out. If the signal strength is lower than a predefined value, then based on *CSSP*, *SSN* and *DMB*, the FLC decides whether the handover is necessary or not.

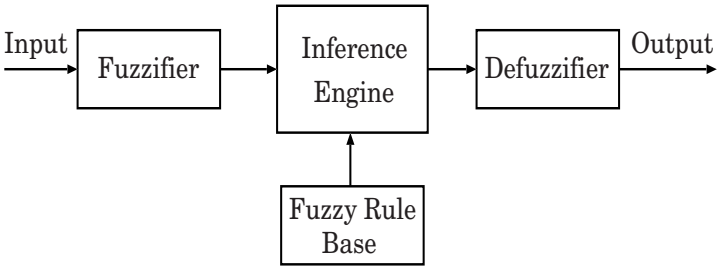


Fig. 2. FLC structure

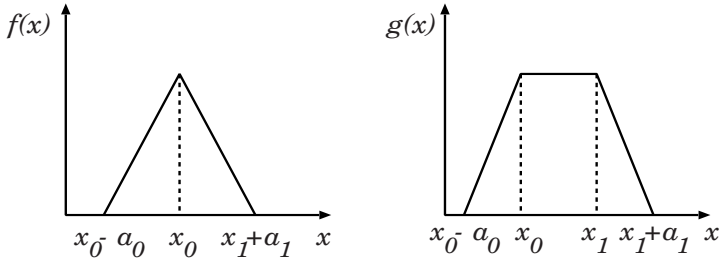


Fig. 3. Membership function shapes

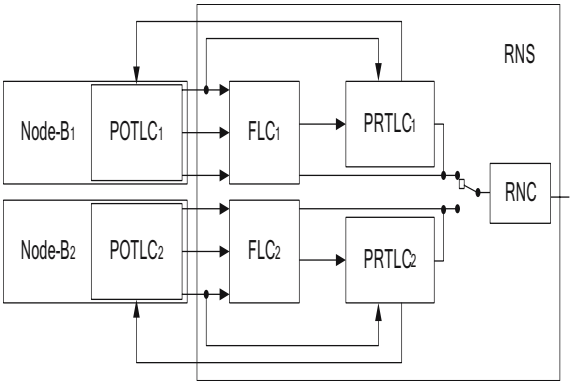


Fig. 4. System model

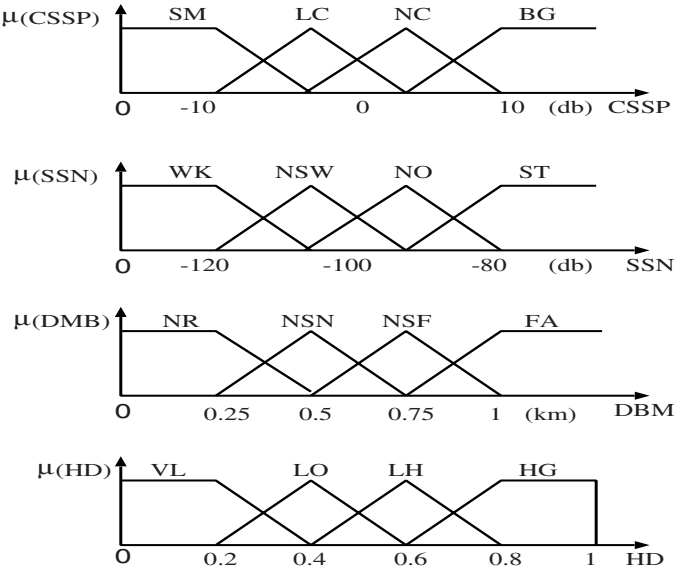


Fig. 5. Membership functions

If the handover is not necessary, the control is returned to the present BS. Otherwise another check of the signal strength is carried out in PRTLC and the present signal strength is compared with the previous signal strength. When the present signal strength is lower than the strength of the previous signal, the handover procedure is carried out.

The term sets of *CSSP*, *SSN* and *DMB* are defined respectively as:

$$\begin{aligned} T(CSSP) &= \{Small, Little\ Change, No\ Change, Big\} \\ &= \{SM, LC, NC, BG\}; \\ T(SSN) &= \{Weak, Not\ So\ Weak, Normal, Strong\} \end{aligned}$$

Table 1. FRB

Rules	CSSP	SSN	DMB	HD	Rules	CSSP	SSN	DMB	HD
1	SM	WK	NR	LO	33	NC	WK	NR	VL
2	SM	WK	NSN	LO	34	NC	WK	NSN	VL
3	SM	WK	NSF	LH	35	NC	WK	NSF	VL
4	SM	WK	FA	LH	36	NC	WK	FA	LO
5	SM	NSW	NR	LO	37	NC	NSW	NR	VL
6	SM	NSW	NSN	LO	38	NC	NSW	NSN	VL
7	SM	NSW	NSF	LH	39	NC	NSW	NSF	VL
8	SM	NSW	FA	LH	40	NC	NSW	FA	LO
9	SM	NO	NR	LH	41	NC	NO	NR	VL
10	SM	NO	NSN	HG	42	NC	NO	NSN	LO
11	SM	NO	NSF	HG	43	NC	NO	NSF	LO
12	SM	NO	FA	HG	44	NC	NO	FA	LH
13	SM	ST	NR	HG	45	NC	ST	NR	LH
14	SM	ST	NSN	HG	46	NC	ST	NSN	LH
15	SM	ST	NSF	HG	47	NC	ST	NSF	HG
16	SM	ST	FA	HG	48	NC	ST	FA	HG
17	LC	WK	NR	VL	49	BG	WK	NR	VL
18	LC	WK	NSN	VL	50	BG	WK	NSN	VL
19	LC	WK	NSF	LO	51	BG	WK	NSF	VL
20	LC	WK	FA	LO	52	BG	WK	FA	VL
21	LC	NSW	NR	LO	53	BG	NSW	NR	VL
22	LC	NSW	NSN	LO	54	BG	NSW	NSN	VL
23	LC	NSW	NSF	LO	55	BG	NSW	NSF	VL
24	LC	NSW	FA	LH	56	BG	NSW	FA	LO
25	LC	NO	NR	LH	57	BG	NO	NR	VL
26	LC	NO	NSN	LH	58	BG	NO	NSN	VL
27	LC	NO	NSF	HG	59	BG	NO	NSF	LO
28	LC	NO	FA	HG	60	BG	NO	FA	LO
29	LC	ST	NR	LH	61	BG	ST	NR	VL
30	LC	ST	NSN	HG	62	BG	ST	NSN	VL
31	LC	ST	NSF	HG	63	BG	ST	NSF	LO
32	LC	ST	FA	HG	64	BG	ST	FA	LO

$$= \{WK, NSW, NO, ST\};$$
$$T(DMB) = \{Near, Not\ So\ Near, Not\ So\ Far, Far\}$$
$$= \{NR, NSN, NSF, FA\}.$$

The output linguistic parameter $T(HD)$ is defined as $\{Very\ Low, Low, Little\ High, High\} = \{VL, LO, LH, HG\}$.

The membership functions of FLC are shown in Fig. 5. The FRB forms a fuzzy set of dimensions $|T(CSSP)| \times |T(SSN)| \times |T(DMB)|$, where $|T(x)|$ is the number of terms on $T(x)$. The FRB is shown in Table 1 and has 64 rules. The control rules have the following form: IF "conditions" THEN "control action".

5 Simulation Results

The cell shape is hexagonal and the coordinates of BSs are indicated as shown in Fig. 6. The BS is located in the center of the cell, the transmission antenna power is 10 W, and cell radius is 2 km. In Table 2 are shown the simulation parameters.

In Fig. 7 is showing the walking pattern for $iseed = 200$ and $nwalk = 10$, which is the case of the handover enforcement (the MS moves to other cells). In this case, the MS moves in the cells: (0,0)→(-1,2)→(-2,1)→(-1,2).

In Fig. 8, Fig. 9 and Fig. 10 are showing the received power from the BS(0,0), BS(2,-1), BS(1,-2), respectively. We have also the results for un-necessary handover, but for the sake of space will not show in this paper. As can be seen from Fig. 8, when the MS is going far from the BS the received power is decreased, while when the MS is approaching neighbor BS the received power from these BSs is increased (see Fig. 9 and Fig. 10).

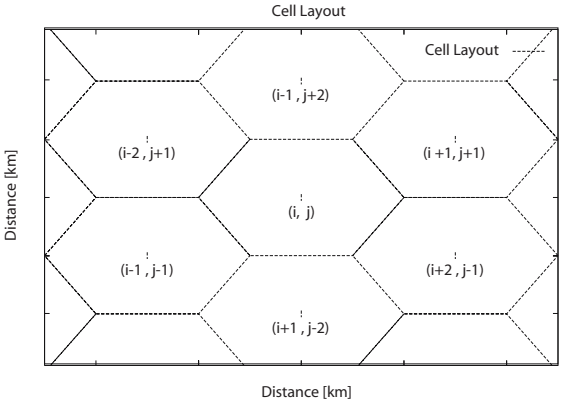


Fig. 6. Cell layout

Table 2. Simulation parameters

Distribution Law	Gaussian Distribution
Number of Walks	5, 10
Random Types	100, 200
Cell Radius	1km, 2km
Transmission Power	10W, 20W
Frequency	2000MHz
Transmission Antenna Beam Tilting	3°
Transmission Antenna Height	40m
Receiving Antenna Height	1.5m
Average Value for a Walk	0.6km
n	1.1

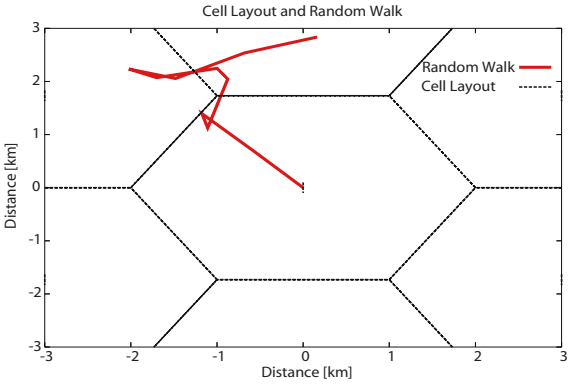


Fig. 7. RW pattern

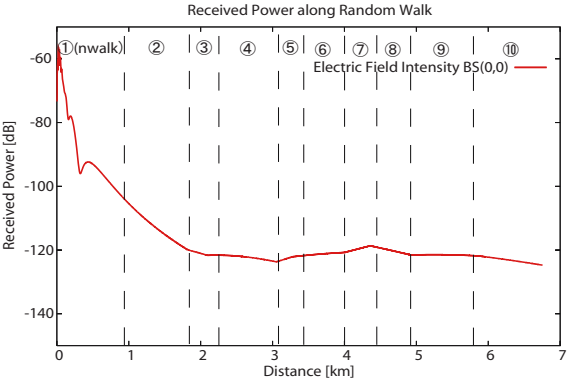


Fig. 8. Received power from BS(0,0)

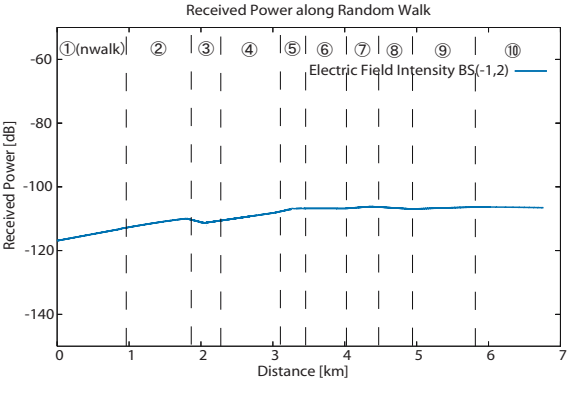


Fig. 9. Received power from BS(-1,2)

Table 3. Simulation results

Measurement Points	Point 1		Point 2		Point 3	
Speed 0 km/h						
CSSP BS	−2.0149	−3.4731	−2.1681	−3.7153	−7.1891	−7.9733
Neighbor BS	−105.55	−102.07	−103.52	−96.763	−103.85	−88.422
Distance	1.9597	2.4628	1.8367	2.3453	1.8021	3.0449
System Output Value	0.645	**0.745**	0.634	**0.740**	0.692	**0.730**
Speed 10 km/h						
CSSP BS	−2.0149	−3.4731	−2.1681	−3.7153	−7.1891	−7.9733
Neighbor BS	−107.55	−104.07	−105.52	−98.763	−105.85	−90.442
Distance	1.9597	2.4628	1.8367	2.3453	1.8021	3.0449
System Output Value	0.632	**0.780**	0.634	**0.710**	0.671	**0.730**
Speed 20 km/h						
CSSP BS	−2.0149	−3.4731	−2.1681	−3.7153	−7.1891	−7.9733
Neighbor BS	−109.55	−106.07	−107.52	−100.76	−107.85	−92.422
Distance	1.9597	2.4628	1.8367	2.3453	1.8021	3.0449
System Output Value	0.616	**0.777**	0.620	**0.726**	0.633	**0.730**
Speed 30 km/h						
CSSP BS	−2.0149	−3.4731	−2.1681	−3.7153	−7.1891	−7.9733
Neighbor BS	−111.55	−108.07	−109.52	−102.76	−109.85	−94.422
Distance	1.9597	2.4628	1.8367	2.3453	1.8021	3.0449
System Output Value	0.596	**0.743**	0.597	**0.756**	0.606	**0.730**
Speed 40 km/h						
CSSP BS	−2.0149	−3.4731	−2.1681	−3.7153	−7.1891	−7.9733
Neighbor BS	−113.55	−110.07	−111.52	−104.76	−111.85	−96.422
Distance	0.3536	0.4821	0.6824	0.9047	1.3158	1.4976
System Output Value	0.576	**0.715**	0.574	**0.794**	0.591	**0.728**
Speed 50 km/h						
CSSP BS	−2.0149	−3.4731	−2.1681	−3.7153	−7.1891	−7.9733
Neighbor BS	−115.55	−112.07	−113.52	−106.76	−113.85	−98.422
Distance	0.3536	0.4821	0.6824	0.9047	1.3158	1.4976
System Output Value	0.545	**0.703**	0.553	**0.713**	0.579	**0.703**

For evaluation of the proposed fuzzy-based handover system, we carried out the measurement for 3 points, where the MS is in the boundary of the 3 cells. In Fig. 11 are shown the measurement points. In this case, the handover is necessary because the MS is moving inside the neighbor cells.

In our system, we consider that the handover is carried out when the output value is bigger than 0.7. We assume that during the RW for each 10 km/h the signal strength is decreased 2 db. We carry out 10 times simulations and calculate the average values. The simulation results are shown in Table 3. The MS is moving inside the neighbor cells, so the handover should be carried out 3 times.

In the results of Table 3, the proposed system in all cases has done 3 handovers. This shows that the proposed system has a good handover decision.

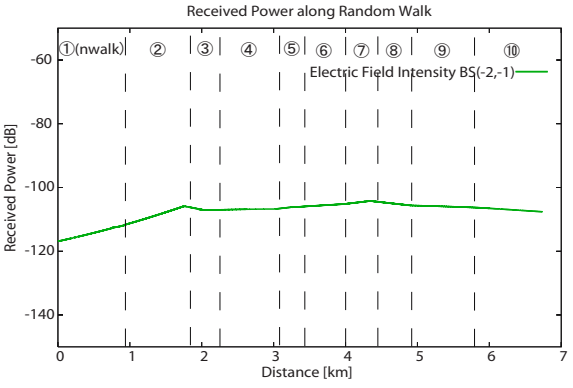


Fig. 10. Received power from BS(-2,1)

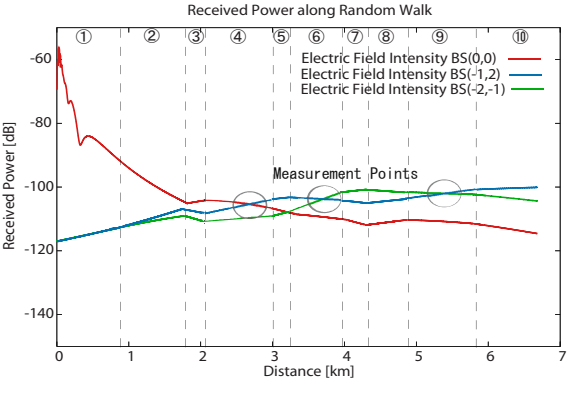


Fig. 11. Three measurement points

6 Conclusions

The mobile systems are based on cellular approach and the area is covered by cells that overlap each other. Many handover algorithms are proposed in the literature. However, to make a better handover and keep the QoS in wireless networks is very difficult. In this paper, we proposed a fuzzy based handover system and considered the case of handover enforcement. From the simulation results, we conclude that the proposed system has a good handover decision.

In the future, we would like to make more extensive simulations and compare the performance of the proposed system with other non-fuzzy-based handover algorithms.

References

1. Berezdivin, R., Breining, R., Topp, R.: Next-Generation Wireless Communication Concepts and Technologies. IEEE Communication Magazine 40(3), 108–116 (2002)
2. Guo, Y., Chaskar, H.: Class-Based Quality of Service over Air Interfaces in 4G Mobile Networks. IEEE Communication Magazine 40(3), 132–137 (2002)

3. Wang, W., Wang, X., Nilsson, A.A.: Energy-Efficient Bandwidth Allocation in Wireless Networks: Algorithms, Analysis, and Simulations. IEEE Transactions on Wireless Communications 5(5), 1103–1114 (2006)
4. Fang, Y., Zhang, Y.: Call Admission Control Schemes and Performance Analysis in Wireless Mobile Networks. IEEE Transactions on Vehicular Technology 51(2), 371–382 (2002)
5. Wong, K.D., Cox, D.C.: A Pattern Recognition System for Handoff Algorithms. IEEE J-SAC 18(7), 1301–1312 (2002)
6. Kovvuri, S., Pandey, V., Ghosal, D., Mukherjee, B., Sarkar, D.: A Call-Admission Control (CAC) Algorithm for Providing Guaranteed QoS in Cellular Networks. International Journal of Wireless Information Networks 10(2), 73–85 (2003)
7. Lin, H.P., Juang, R.T., Lin, D.B.: Validation of an Improved Location-Based Handover Algorithm Using GSM Measurement Data. IEEE Transactions on Mobile Computing 4(5), 530–536 (2005)
8. Fiengo, P., Giambene, G., Trentin, E.: Neural-based Downlink Scheduling Algorithm for Broadband Wireless Networks. Computer Communication 30(2), 207–218 (2007)
9. Barolli, L., Koyama, A., Suganuma, T., Shiratori, N.: GAMAN: A GA Based QoS Routing Method for Mobile Ad-hoc Networks. Journal of Interconnection Networks (JOIN) 4(3), 251–270 (2003)
10. Onel, T., Ersoy, C., Cayirci, E.: A Fuzzy Inference System for the Handoff Decision Algorithms in the Virtual Cell Layout Base Tactical Communications System. In: IEEE Military Communications Conference (MILCOM 2002), vol. 1, pp. 436–441 (2002)
11. Chan, P.M.L., Sheriff, R.E., Hu, Y.F., Conforto, P., Tocci, C.: Mobility Management Incorporating Fuzzy Logic for a Heterogeneous IP Environment. IEEE Communications Magazine 39(12), 42–51 (2001)
12. Mohanty, S., Akyildiz, I.F.: A Cross-Layer (Layer 2+3) Handoff Management Protocol for Next-Generation Wireless Systems. IEEE Transactions on Mobile Computing 5(10), 1347–1360 (2006)
13. Yu, F., Krishamurthy, Y.: Optimal Joint Session Admission Control in Integrated WLAN and CDMA Cellular Networks with Vertical Handoff. IEEE Transactions on Mobile Computing 6(1), 126–139 (2007)
14. Dubois, D., Prade, H., Yager, R. (eds.): Fuzzy Sets for Intelligent Systems. Morgan Kaufman Publishers, San Francisco (1993)

Fault Tolerance for Small-World Cellular Neural Networks

Kautsuyoshi Matsumoto, Minoru Uehara, and Hideki Mori

Dept. of Open Information Systems, Graduate School of Engineering, Toyo University
2100, Kujirai Kawagoe Saitama, Japan
{dz080001x,uehara,mori}@toyonet.toyo.ac.jp

Abstract. In this paper, we propose a mechanism for fault tolerance for Small-World Cellular Neural Networks (SWCNN). Small-world networks exist in the range between regular and random networks. SWCNN is a Cellular Neural Network (CNN) that has small-world network structure and has better performance than CNN. SWCNN needs to be fault tolerant because it has higher levels of error propagation than CNN.

1 Introduction

In recent years, sensor networks have been proposed for use in surveillance [1]. Such sensor networks have surveillance cameras that use CCDs. If these cameras were used for image processing, the load on the sensor network server would be lightened.

Research is underway on using Cellular Neural Networks (CNN) for image processing [2][3][4][5]. Small-World CNNs (SWCNNs), as proposed by [3][4][5], provide superior image processing in comparison to CNNs. A SWCNN consists of Small-World network structures, which have small network diameters and are highly clustered. SWCNNs can be easily used for processing in a wide range of applications but they do not tolerate faults that occur when some neurons fail.

2 Small-World Cellular Neural Networks

2.1 Small-World Network

An example of a Small-World network is the global inter-connectedness of people, where any person is connected to any other in the world through a chain of 6 people. Watts and Strogatz formulate Small-World networks in terms of graph models [6]. They introduce two types of coefficients:

- L : Characteristic path length
- C : Clustering coefficient

L is the number of edges in the shortest path between two nodes, averaged over all pairs. C_v for node v has k links, and $k_v(k_v - 1)/2$ edges can exist between them. C is the average of all C_v. A feature of Small-World networks is that C is large and L is small (Fig.1).

M. Takizawa, L. Barolli, and T. Enokido (Eds.): NBiS 2008, LNCS 5186, pp. 223–231, 2008.
© Springer-Verlag Berlin Heidelberg 2008

Fig. 1. Small-World Network

2.2 Small-World Cellular Neural Networks (SWCNN)

A CNN is a network, where the neuron elements are regularly arranged in an array, and are connected in a lattice [2]. A Small-World Cellular Neural Network (SWCNN) is a CNN that has Small-World network structure. CNN introduces a random connection rate p_c for which the CNN network structure becomes a Small-World network (Fig.2). The state (1) and output (2) equations are as follows:

$$\begin{aligned} x(t+1) = -x_{ij} + I + w_c M(ij;pq)\, y_{ij}(t) \\ + \sum_{kl \in N_r(ij)} A(ij;kl)\, y_{ij}(t) \\ + \sum_{kl \in N_r(ij)} B(ij;kl) u_{ij}(t) \end{aligned} \tag{1}$$

Each cell takes an input u_{ij}. A state variable x_{ij} is set from input u_{ij} and output y_{ij}. This template shows the coupling coefficient with the center and neighborhood cell.

$A(ij; kl)$ is an output template and $B(ij; kl)$ is an input template, as shown in equation (4). The output is determined by equation (2).

$$y(t) = (|\, x_{ij}(t)+1\,| + |\, x_{ij}(t)-1\,|) \tag{2}$$

The output template is a feedback template that depends on the output y_{ij}. The input template is a feedforward template that depends on input u_{ij}. $N_r(ij)$ gives the neighborhood cells around the central cell ij. Decisions are made by calculating the distance to that nearest neighbor. Fig.2 shows the neighborhood neuron cells when $r = 1$. I is a threshold. w_c is a coefficient to represent the degree of connections between two cells at random. $M(ij; pq)$ represents the connection between $cell_{ij}$ and

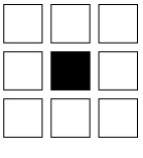


Fig. 2. Neighbor cells

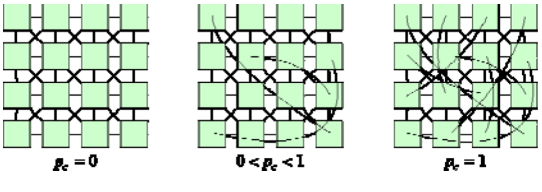


Fig. 3. Small-World Cellular Neural Networks

$cell_{pq}$. Rate p_c is the probability that $cell_{ij}$ and $cell_{pq}$ are connected (Fig.3). When $cell_{ij}$ and $cell_{pq}$ are connected $M(ij; pq) = 1$; otherwise, $M(ij; pq) = 0$.

SWCNN has a faster convergence speed than CNN in image processing [2][3][4][5]. The templates for Noise Reduction and Edge Detection in SWCNN are as follows.

Noise Reduction

The following templates are used to reduce noise in binary images.

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 1 \end{bmatrix} \quad B = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad I=0 \; w_c=1.0 \tag{3}$$

Given the image shown in Fig.4, image processing with a random connection rate $p_c = 1.0$ results in the image shown in Fig.5.

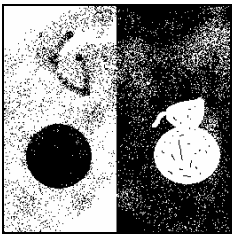


Fig. 4. Original image

Fig. 5. Processed image

Edge Detection

The following templates are used for edge detection in gray scale images.

$$A = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad B = \begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix} \quad I=0 \; w_c=1.0 \tag{4}$$

Given the image shown in Fig.6, image processing with a random connection rate $p_c = 1.0$ results in the image shown in Fig.7.

Fig. 6. Original image

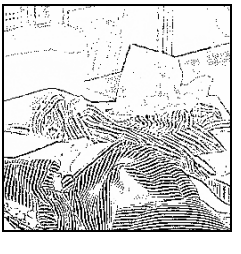


Fig. 7. Processed image ($p_c=1.0$)

3 Fault Tolerance for Small-World Cellular Neural Networks

3.1 Multiplex Redundancy

Duplex with a Comparator
Detection of duplicate redundancy is carried out with a comparator and duplicate hardware, as shown in Fig.8. Computations are performed in parallel and the results compared [7]. In the event of a disagreement, an error signal is generated. In its most basic form, this approach can detect faults but cannot tolerate them as there is no method for determining which of the two is faulty.

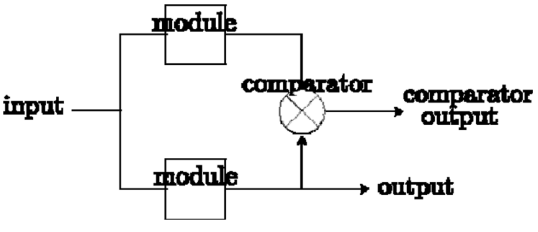


Fig. 8. Duplex with a Comparator

Triple Modular Redundancy (TMR)
Triple Modular Redundancy (TMR), as shown in Fig.9, selects one of the outputs by voting between three modules that execute the same three operations in parallel [7]. TMR gracefully tolerates the great majority of faults.

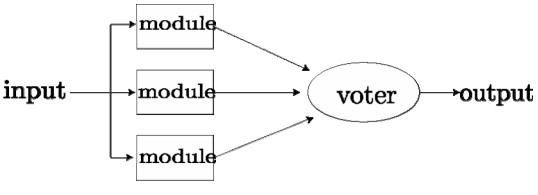


Fig. 9. Triple Modular Redundancy (TMR)

3.2 Fault Tolerance for Small-World Cellular Neural Networks

SWCNN using Comparison and Isolation
We propose three fault tolerance methods for SWCNN. The first uses Comparison and Isolation, as shown in Fig.10. This method duplicates a SWCNN, with one placed on top of the other. It compares the output of two neurons from the same position in each plane. If the neuron outputs do not match, those neurons are isolated from each plane in the SWCNN and are not used. The common input is used for the output reset.

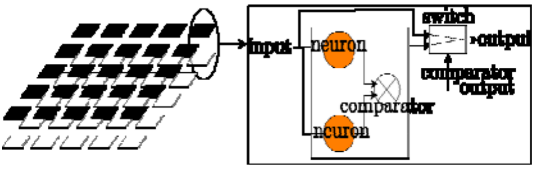


Fig. 10. SWCNN using Comparison and Isolation

SWCNN using TMR

The second approach to a fault tolerant architecture that we propose with SWCNN uses TMR, as shown in Fig.11. The TMR method stacks three of the same SWCNNs. Voting is performed between the three neurons located in the same vertical position. The output value is determined by voting between the three neurons.

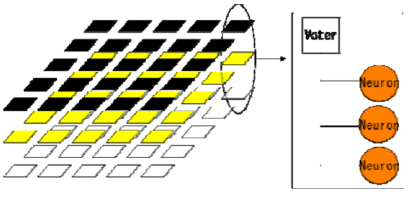


Fig. 11. SWCNN using TMR

SWCNN using TMR with a Reliability Counter

Finally, we propose a fault tolerant SWCNN using TMR with an Reliability Counter for intermittent faults, as shown in Fig.12. The Reliability Counter for each plane counts up the number of times that that neuron was disregarded. If the three neuron output values are all different, the output value is chosen from the neuron with the lowest value in the Reliability Counter.

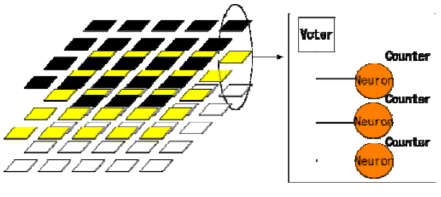


Fig. 12. SWCNN using TMR with Reliable Counter

4 Simulation Results

4.1 Simulation Method

In this work, we simulated our SWCNNs using templates for Noise Reduction and Edge Detection. We used two type of faults: stack faults and intermittent faults. A

stack fault results in a value of 0 or 1. An intermittent fault occurs in a neuron. It is random and is followed immediately by a correct value.

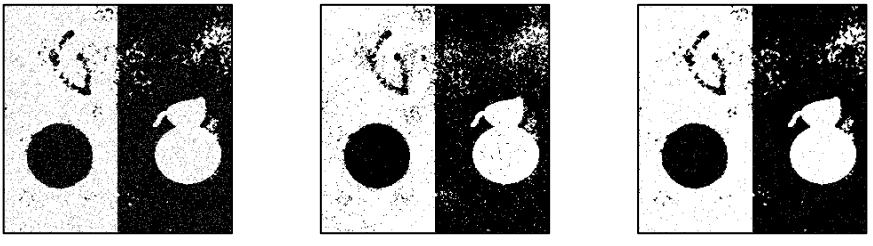
4.2 Evaluation

Noise Reduction

The output images shown in Fig.13-15 result from using our methods for Noise Reduction.

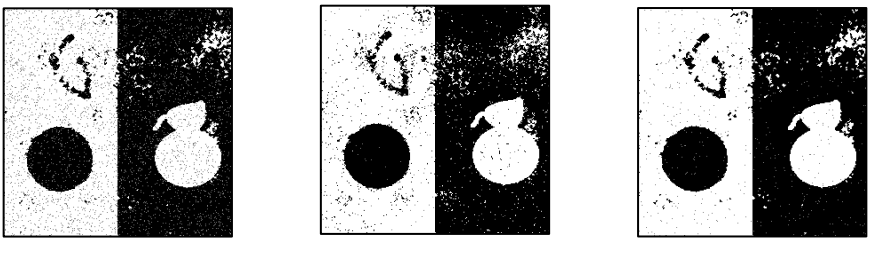
With stack faults, images resulting from the usual SWCNN process are noisy. Our proposed methods recover well from these faults. However, some noise cannot be removed with SWCNN using the Comparison and Isolation method because the output is ignored, the input is used as output, and the noise is maintained. On the other hand, SWCNN using TMR masks faults in both cases at a low failure rate.

With intermittent faults, output images are noisy with SWCNN and with SWCNN using Comparison and Isolation. In SWCNN using Comparison and Isolation, the processed image is the same as the original image wherever this method has detected and isolated failure neurons, given that the output values for isolated neurons are determined by the input values. On the other hand, faults in the output image are masked well both with SWCNN using TMR and with SWCNN using TMR with Reliability Counter.



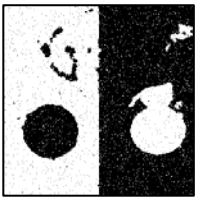
(a)SWCNN (b)SWCNN using Comparison and Isolation (c)SWCNN using TMR

Fig. 13. Output Images (fault at 0, fault rate 0.05)

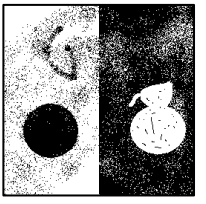


(a)SWCNN (b)SWCNN using Comparison and Isolation (c)SWCNN using TMR

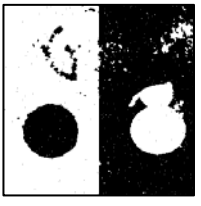
Fig. 14. Output Images (fault at 1, fault rate 0.05)



(a)SWCNN



(b) SWCNN using Comparison and Isolation



(c)SWCNN using TMR

(d)SWCNN using TMR with Reliability Counter

Fig. 15. Output Images (Intermittent fault, fault rate 0.05)

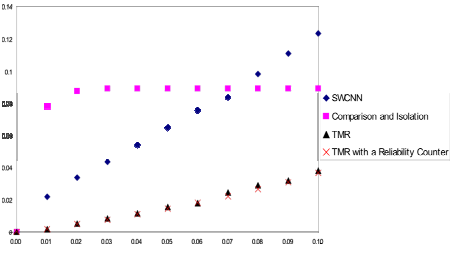


Fig. 16. Unmatched pixel rate

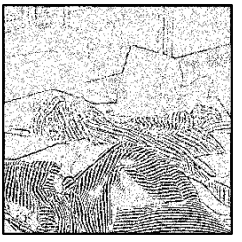
Fig.16 shows the unmatched pixel rate for four methods. Two of the proposed methods, SWCNN using TMR and SWCNN using TMR with Reliability Counter, give good results. However, the results for SWCNN using Comparison and Isolation are bad because the output image is the same as the original image.

Edge Detection

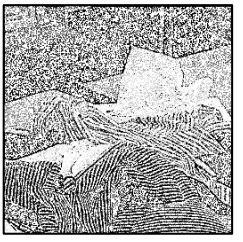
The following images show the results of using SWCNN and our proposed methods for Edge Detection (Fig.17-19).

With stack faults, the output images are noisy with SWCNN and with SWCNN using Comparison and Isolation. With SWCNN using Comparison and Isolation, the processed image is poor because the isolated failure neurons meant that input values were used instead. SWCNN using TMR (with and without a Reliability Counter) masks the faults well.

With intermittent faults, the output images are noisy both with SWCNN and with SWCNN using Comparison and Isolation. With SWCNN using Comparison and Isolation, the processed image is poor because the isolated failure neurons meant that



(a)SWCNN

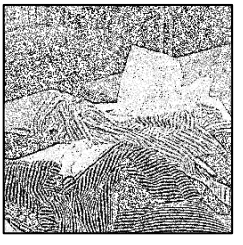


(b)SWCNN using Comparison and Isolation

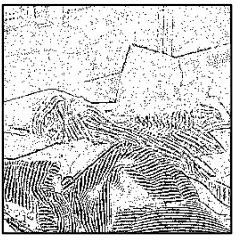
(c)SWCNN using TMR

Fig. 17. Output Images (stack fault at 0, fault rate 0.1)

(a)SWCNN



(b)SWCNN using Comparison and Isolation



(c)SWCNN using TMR

Fig. 18. Output Images (stack fault at 1, fault rate 0.1)

(a) SWCNN

(b) SWCNN using Comparison and Isolation

(c)SWCNN using TMR

(d)SWCNN using TMR with Reliability Counter

Fig. 19. Output Images (Intermittent fault, fault rate 0.1)

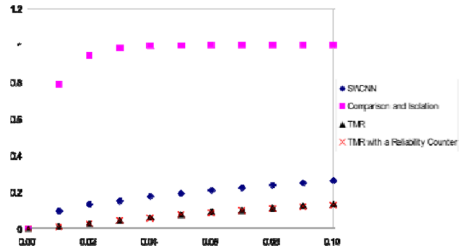


Fig. 20. Unmatched pixel rate

input values were used instead. SWCNN using TMR (with and without a Reliability Counter) masks the faults well.

The results are excellent for two of the proposed methods, SWCNN using TMR and SWCNN using TMR with Reliability Counter. However, the results are poor with SWCNN using Comparison and Isolation because the output image is the same as the original.

5 Conclusion

In this paper, we have proposed three methods for fault tolerance in SWCNNs. SWCNN using Comparison and Isolation, SWCNN using TMR, and SWCNN using TMR with Reliability Counter are described and evaluated. The proposed fault tolerant SWCNNs provide better fault tolerance than the original SWCNN. While SWCNN using Comparison and Isolation shows lower fault recovery, it has smaller hardware redundancy, and therefore lower complexity, compared to the other proposed methods. Our aim in future work is to obtain better fault recovery performance from SWCNN using Duplex.

References

1. Collins, R.T., et al.: Algorithm for Cooperative Multisensor Surveillance. Proc. IEEE 89(10), 1456–1477
2. Chua, L.O., et al.: Cellular Neural Networks: Theory. IEEE Transactions Circuit and Systems 35(10), 1257–1272
3. Tsuruta, K., et al.: Small-World Cellular Neural Networks for Image Processing Applications. In: European Conference on Circuit Theory and Design, vol. 1, pp. 225–228
4. Tsuruta, K., et al.: Two Types of Network Topologies of Small-World Cellular Neural Networks. In: RISP International Workshop on Nonlinear Circuit and Signal Processing, pp. 113–116 (2004)
5. Nakano, M., et al.: Analysis of Edge Detection Using Direction Preserving Small World Cellular Neural Networks. In: RISP International Workshop on Nonlinear Circuit and Signal Processing, pp. 145–148 (2007)
6. Watts, D.J., Strogatz, S.: Collective dynamics of 'small-world' networks. Nature 393, 440–442
7. Pradhan, D.K.: Fault-Tolerant Computer System Design. Prentice Hall PTR, Englewood Cliffs

A 4+1 Bit Month-Scale Regularity Mining Algorithm with One-Path and Distributed Server Constraints for Mobile Internet

Toshihiko Yamakami

[1] ACCESS, 2-8-16 Sarugaku-cho, Chiyoda-ku, Tokyo, Japan
Toshihiko.Yamakami@access-company.com
[2] Graduate School of Engineering, Kagawa University

Abstract. Mobile Internet becomes increasing visible in everyday life. As increased penetration leverages mobile application business opportunities, it is crucial to identify methodologies to fit mobile-specific demands. Regularity is one of the important measures to enclose *easy-come, easy-go* mobile users. It is known that a user with multiple visits in one day with a long interval has a larger revisiting possibility in the following month than the others. The author proposes a 4+1 bit method to incorporate this empirical law in order to cope with the two major mobile restrictions: distributed server environments and large data stream. The proposed method can be performed in a one-path manner with 32-bit word boundary-aware memory compaction. The experimental result shows the method is promising to identify revisiting users under mobile-specific constraints.

1 Introduction

"Mobile Internet" has become a multi-faceted term covering a wide range of functions and aspects as it has deeply penetrated everyday life. The penetration reveals a new aspect of human behavior, with a large amount of access log data. It demands a new measurement for evaluating user behaviors in a mobile-specific context. The mobile handset has a small-sized screen, therefore, it is crucial to increase the end-user loyalty, and to enclose them in mobile services. For subscription-based mobile customers, it is important to evaluate the long-term regularity of visits rather than the total number of visits. The author proposes a new method for evaluating regularity in mobile Internet services in order to cope with the two major mobile restrictions: distributed server environments and large data stream.

2 Related Works

The dynamics and volatility of mobile Internet services prevented long-term observational studies, even given this future forecast. The first large-scale mobile Internet analysis is done by Halvey. He reported a positive relationship to the day

M. Takizawa, L. Barolli, and T. Enokido (Eds.): NBiS 2008, LNCS 5186, pp. 232–241, 2008.
© Springer-Verlag Berlin Heidelberg 2008

of the week in the mobile clickstream [1]. The author conducted the regularity study on the mobile clickstreams and reported 80 It is still an active topic for researchers to study how many different types of regularity behaviors people show and how stable each behavior is over a long period of time.

Mining data streams is a field of increasing interest due to the importance of its applications and the dissemination of data stream generators. Research dealing with continuously generated massive amount of data has quickly caught the attention of researchers in recent days [2] [3]. Considering the fast growth of the mobile Internet, it is an important research topic to be covered.

Mobile clickstream analysis is an unexplored research field because there are still WML1.3-based mobile Internet sites used in many countries. A WML deck consists of multiple cards, where many user clicks are absorbed in the client and not available to the servers.

The author proposed an early version of the time slot method, to identify regular users with a long interval of sub-day web visits [4]. The method was coined on the conjecture that the users that come to a web service twice in one day tend to return the service in the following month [5]. From the empirical results, it appeared to be a promising method. The method identifies a regular user with an explicit division among active time slots. This leads to inflexibility in setting up time slot sizes.

The disadvantage of the proposed method is that an explicit division can be identified only after all access logs are analyzed. This is a considerable drawback when it is applied in a stream-mining manner. In stream mining, with the constraints of storage, it is desirable to identify the outcome in an on-the-fly manner. The preceding method did not match this requirement.

3 Requirements

There are following requirements:

Efficient Large-scale Mining Requirement. In order to cope with a large mobile user base, it is desirable to pack each user data into one-word (32-bit) memory in an efficient way. When it is assume to be process-able for one million users, it needs 20 bits to store the ID. In order to get one-percent accuracy, it needs 7 more bits. It allows only 5 bits for regularity mining work area.

One-path Constraint. With a large amount of data stream, it is realistic that an algorithm is performed in a one-path manner. In other words, each piece of click is processed only once during the stream processing. There is no centralized server to sort and store all the click logs.

Distributed Server Configuration Requirements. In order to make load balancing, it is common to use a distributed server configuration. In this distributed configuration, it is realistic that clickstream is distributed among multiple servers. It is realistic that some of the server logs have time lag when they arrive at an analysis system.

4 Method

4.1 TCW-Method

The author performed a preliminary study of commercial mobile Internet users using clickstream logs. The patterns obtained indicated that a user that returns to a Web site after a certain length of time has a greater possibility of returning to the same Web site in the following month.

In order to capture this rule in an efficient method, the author proposed a method called the *time slot count in a window method (TCW-method)* [4]. In the TCW-method, a window size is set to determine the revisiting user patterns. Usually, this window size is set to one day to capture sub-day-level user patterns. Then, the window is split into multiple time slots. The time slot size reflects service-specific characteristics. The clickstream per user is distributed into these time slots. Then, the number of slots containing clicks is counted. For example, if a user visits a web once every hour, it shows 24 visits in a day.

The processing flow is illustrated in Fig. 1 with a time slot count threshold value t_{th}.

From empirical observations, the author sets 2 as the default threshold value for the method.

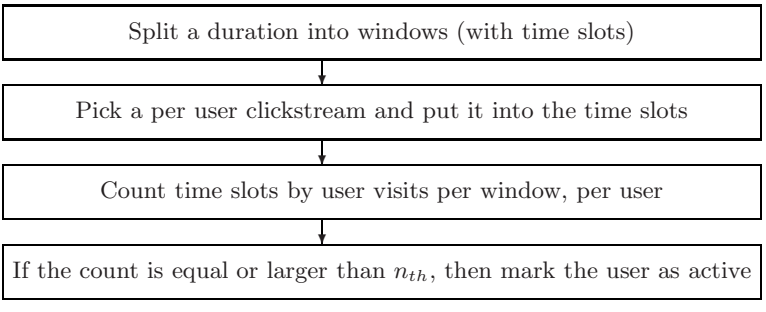


Fig. 1. Processing flow of TCW-method

4.2 Relaxed Realtime N+1 Bit Method

In order to cope with one-path, distributed server requirements, the author proposes a relaxed real-time N+1 bit method. For each user, N+1 bits are assigned to store Boolean values representing N+1 time slots whether a user visited the web site during the time slot. Bit-0 represents the most resent T period, Bit-n represents the duration between (n-1)*T and n*T. In this paper, this method is referred as *RRN1-method*. In this method, the regularity detection is done in the following steps:

1. Start with month beginning, set T as time slot size (e.g. 6 hours), set N+1 as the number of bits; each bit corresponds to a Boolean whether a user visits a web site during the time slot.

2. After each time duration T, mark the user as a regular user, when there are multiple bits set between bit-1 to bit-(N+1).
3. Discard the bit-(N+1), rename bit-0 to bit-(N) to bit-1 to bit-(N+1), and set false to bit-0.
4. When a log with time stamp for a user arrives, appropriate bit is set for a user.
5. Repeat this procedure until T after the end of month is processed.
6. When T after the end of month is detected, the processing for this month is completed.

This allows any delay of T period from any of the distributed servers. For example, when T is 6 hours, 6-hour-delay can be tolerable from any of the distributed servers. The random arrival of web logs with time stamps can be tolerable when the delay is less than T. This method is one-path, therefore, the logs can be discarded after the appropriate bit is set. At completion, all the users are marked either a regular user or a non-regular user.

This method is based on observation that a user who will revisit a mobile web after a long interval within a day has a higher possibility to revisit the web in the following month. The algorithm follows the time slot count in a window method (TCW-method), with 24-hour window size, with 4 time slots in a window (each time slot is 6-hours). When T is set to 6 hours, it needs 5 bit to track regularity: 4 bits for 24 hours and extra one bit to make one-path, relaxed real-time processing. In 32-bit word, 27 bits are used to store hashed value of user identifiers.

This 32-bit-fit algorithm works well in the standard 32-bit processor. With 64-bit processor, one 64-bit word can accommodate two users.

The memory structure is depicted in Fig. 2. *E of Word* denotes end of word pool.

In this structure, two-level hash mechanism is used to store the status word pool. Each status word contains last 24 hours+6 hours visiting history, where 1 bit is used for Boolean for visiting history in 6 hours. In a case where 27-bit user hash to store user identifiers, when there are a million users, 20-bit is needed to uniquely identify the entire users.

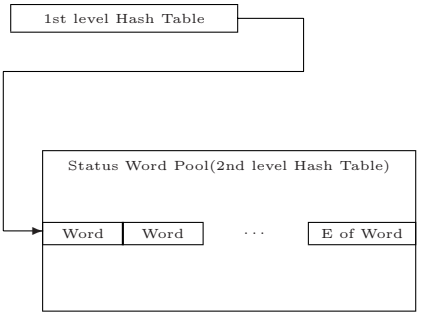


Fig. 2. A memory structure for RRN1-method

In order to effectively store the data, we can use first level hash to index the final 20-bit users. The size of the first level hash depends on trade-off between memory size and processing complexity. If the first level hash approaches 20-bit, it needs more index table size, with less amount of processing.

When first level hash uses 16-bit, it needs 2^{16} word for index memory. When 1 word is 4 bytes, it is 256K bytes. A million 32-bit words are 4M bytes, therefore, the index table overhead is tolerable. The size of status word pool is a design issue. For example, when the status word pool is too large, it wastes memory. When the status word pool is too small, the overhead to store the pool increases as well as the processing overhead.

5 Case Study

5.1 Revisit Ratio

The author uses a revisit ratio to evaluate the classification of regular users. The revisit ratio $R(U, m)$ in month m for a group of users U is defined as follows where $A(U, m)$ are users in U that access content (any URL in a given Web site) in month m:

$$R(U, n) = \frac{\mid A(U, m) \cap A(U, m + 1) \mid}{\mid A(U, m) \mid}$$

Where U_a is all users that access content in month m, $R(U_a, m)$ represents the total revisit ratio for month m active users. When the active users for month m are split into subgroups, U_1, U_2, ..., $R(U_1, m)$, $R(U_2, m)$, ... denotes the revisit ratio for each user group.

5.2 Data Set

The target of observation is a commercial news service on the mobile Internet. The service is available on three different mobile carriers, with a slightly different content menu. Each mobile carrier has different underlying network characteristics and different charging policies.

The user ID (UID), time-stamp, command name and content shorthand name are stored in the log. The services were launched between 2000 and 2001, and continue in use up to today. The target service provides 40 to 50 news articles per week on weekdays. The commercial mobile service charges the monthly subscription fee to users, approximately 3 US dollars per month. The UID is usually 16 or more unique alphanumeric characters long, e.g. "310SzyZjaaerYlb2". The service uses Compact HTML [6], HDML (an early version of WML) and MML (a proprietary dialect of a subset of HTML).

The log for each carrier includes 2,390,673 lines for carrier-A, 1,591.985 lines for carrier-B, and 397,373 lines for carrier-C. The number of unique users identified in each log is 60,311 users for carrier-A, 90,291 users for carrier-B, and 13,150 users for carrier-C.

Table 1. News-access only log data set characteristics

Carrier	Months (YYMM)	Clicks	Sum of Monthly Unique Users	Unique Users
A	0101-0105	196369	11610	4462
A	0201-0205	144767	7046	2442
B	0101-0105	86808	3163	1672
B	0201-0205	82815	2437	901
C	0101-0105	16050	1245	901
C	0201-0205	11610	914	329

The registration records include 12,462 unique users for Carrier-A, 2,954 unique users for Carrier-B, and 1,217 unique users for Carrier-C. In order to remove the non-news based additional services, which differ from carrier to carrier, the author filters all non-news, related transactions in logs from January to May in 2001 and from January to May in 2002. The data set characteristics are outlined in Table 1. Months are expressed as YYMM, for example, 0105 is May 2001.

5.3 Result

The author performed Welch's t test for prediction and reality data. R is used to perform the test with `t`.test() [7]. The test summary is depicted in Table 2. The (0,1) vector of all users represents the revisit reality. The test examines how significant the proposed method's identification capability is for regular users, in other words, users with multiple time slot visits in a window are compared to all users that access news in the given Web site in the month.

In the following tables, R(all) denotes the revisit ratio of all users the following month. The month is obvious from the month column, therefore, the second parameter month for R is omitted. R(RRN1) denotes the revisit ratio of users identified by the RRN1-method in the following month. The author performed a case study in 2001 and 2002 with 3-hour time slots in a 24-hour window. The threshold value is set to 2.

The observed service is a mobile commercial news service in Japan. The service is in commercial operation in 2007, but the recent log data were not available for this research.

Table 2. Welch's t test summary

Alternative hypothesis	True difference in means of two samples is not equal to 0
Sample 1	(0,1) vector of all users with multiple time slot visits in a window where 0 means no revisit in the following month 1 means a revisit in the following month
Sample 2	(0,1) vector of all users with news access in the month
Tool	R's `t`.test()

Table 3. Carrier-A results from January to April 2001 and from January to April 2002

month (YYMM)	R(RRN1)	R (all)	t-value	degree of freedom	p-value	signifi -cance
0101	89.51	66.59	-14.601	1322.5	0.0000	**
0102	90.86	70.93	-13.163	1423.5	0.0000	**
0103	87.78	67.11	-12.762	1382.7	0.0000	**
0104	90.91	70.93	-13.077	1480.4	0.0000	**
0201	93.72	74.65	-11.749	1273.6	0.0000	**
0202	94.74	74.72	-12.593	1365.1	0.0000	**
0203	93.42	73.15	-12.157	1337.5	0.0000	**
0204	93.83	75.19	-11.074	1224.2	0.0000	**

Note:
**: 1 % confidence level
*: 5 % confidence level

The result is depicted in Table 3. The t test gives a 1 % confidence level of significance in all the months under observation. The revisit ratio is 87.11 – 90.12 % range during January and April 2001, with an average of 88.74 %. It is in the 90.07 – 91.75 % range between January and April 2002, with an average of 90.00 %. It should be noted that the improvement is derived the increased revisit ratio in 2002. The revisit ratio increased from 66.59 – 70.93 %, during January and April 2001 to 73.15 – 75.19 % during January and April 2002. As time passed, the volatile users decreased and the remaining users tended to show a high revisit ratio.

The preliminary results for Carrier-B and Carrier-C showed similar results.

The relaxation of explicit splits among time slots with user visits impacts the revisit ratio when a user visits for a short period of time across the boundary between two time slots. In this case, the user visits for a very short period of time, resulting in two consecutive time slot counts. In the case study, this effect is negligible, to within 2-3 % true positive ratio, when the time slot size is a maximum of 3 hours. It should be noted that not all 2-3 % errors derive from this time slot crossing effect. The average user stay time on the mobile Internet is less then 10 minutes. Considering this factor, comparatively large time slots prevent error propagation from time slot boundary crossing patterns. The case study shows that 3 hours is sufficient to bring negligible effects. When the time slot size is smaller than 3 hours, it needs further validation tests.

6 Discussion

6.1 Advantages of the Proposed Method

The advantage of the RRN1-method is that it does not depend on the final state. When the count of time slots with user visits reaches a certain threshold value, all the later clickstream for the user can be safely discarded because the later results do not impact the final identification.

The method relies on the conjecture that a user with multiple time slots visits and a certain threshold will visit the Web site in the following month. This multiple count can be any day in the previous month. When the identification system captures multiple counts in a day, all the following clickstream can be safely discarded without impacting true positive ratio accuracy. This one-path nature of the RRN1-method fits stream mining with its constraints on storage.

Intuitively, this method has a drawback with true positive ratio by ignoring explicit splits between time slots. The past method (TCW-method with 4-hour time slot) shows an average revisit ratio of 92.01 % from January to April 2001, and 93.88 % from January to April 2002. The RRN1-method for 6-hour time slot shows an average revisit ratio with 88.74 % from January to April 2001, and 91.00 % from January to April 2002. It shows a loss of approximately 2–3 % true positive ratio with the trade-off of on-the-fly processing capability. This comes from the longer window size (4-hour vs. 6-hour) and not from one-path, distributed server tolerant characteristics. With the loss of 2–3 % true positive ratio, the RRN-1 method using 6-hour time slot gains 2 bits, which means that it can analyze four times more users with the same amount of memory.

It is assumed that the TCW-method and the RRN1-method show the equivalent classifier performance when they are used with the identical window size. The advantage of the RRN1-method is that it can allow hour-scale delay of clickstream log arrivals from distributed servers.

There is always a trade-off between high true positive ratio and wide coverage. When high true positive ratio is pursued, it will focus on the small group, therefore, the derived association rule can be applied to a small portion of the samples. When wide coverage is pursued, it is difficult to obtain the high true positive ratio of the derived rules. Considering this trade-off, the obtained 88.74 % and 91.00 % true positive ratio is acceptable for applications. When higher true positive ratio is required for applications, it will require the sacrifice of wide coverage.

This method can be applied to a wide range of mobile applications with time stamped logs. It is a key advantage of the proposed method.

6.2 Applications

It is important to identify what applications can use this measure to realize value-added services in the mobile Internet. For example, a high revisit ratio like 90 % can be used as a measure of the impact of new services or new user interfaces. The revisit ratio can be used as a litmus test to measure the effectiveness of new services or new user interfaces.

It is difficult to capture user feedback on the mobile Internet because the user interface is limited and the user does not want to perform additional input to give service feedback. It is feasible for content providers to differentiate between users with high retention and others in their services. It could improve the user capture and retention in mobile services.

NTTDoCoMo press-released that they will enable all content providers (both carrier-official and non-official web sites) to use ImodeId (their unique user identifier system). This public availability will start March 2008. In the past, the use of unique user identifier was restricted to only carrier-approved official sites. This new carrier movement will increase applicability of user-identifier-based research methods.

6.3 Limitations

Each service has its own characteristics. This research has limitations where it was performed on single service. The limitations include (a) service-specific (the result came from news services), (b) profile-specific (90 % users were male, most of them were age 20's and 30's), (c) time (2001 and 2002 data).

The periodic update of content during a day is a basic unchanged mobile service pattern that has not changed. The data obtained in 2002 is applicable as long as this basic service pattern persists.

It should be noted that the services observed are still commercially in operation and available now.

7 Conclusion

Mobile Internet business providers want to turn their raw data into new science, technology and business. The author conjectured that a user that show multiple visits to a mobile Web site on any day in a given month, has a high tendency to visit the site the following month.

The author described the mobile-specific requirements: efficient large-scale data mining requirement, one-path requirement, and distributed server configuration requirement. Considering the stream-mining requirement of the mobile Internet, the author relaxed the time slot count method in order to match large-scale data in the distributed environment. The data should be processed with one-path mechanism, having allowance of delay of log arrival.

Empirical observations in 2001–2002 with commercial news service subscribers show that this relaxation still match high accuracy of the following month revisit ratio prediction. One-path nature lost only 2–3 % prediction accuracy compared to the other methods which required all data available at an analysis time.

The RRN1-method can be applied to a wide range of mobile applications that use time stamps in logs. There is a trend that wireless carriers release their unique user identifier system to open public. It will increase importance and usability of user-identifier-based research methods.

References

1. Halvey, M., Keane, M., Smyth, B.: Predicting navigation patterns on the mobile-internet using time of the week. In: WWW 2005, pp. 958–959. ACM Press, New York (2005)
2. Gaber, M.M., Zaslavsky, A., Krishnaswamy, S.: Mining data streams: a review. ACM SIGMOD Record 34(2), 18–26 (2005)

3. Jiang, N., Gruenwald, L.: Research issues in data stream association rule mining. ACM SIGMOD Record 35(1), 14–19 (2006)
4. Yamakami, T.: A time slot count in window method suitable for long-term regularity-based user classification for mobile internet. In: MUE 2008, pp. 25–29. IEEE Computer Society Press, Los Alamitos (2008)
5. Yamakami, T.: A long interval method to identify regular monthly mobile internet users. In: WAMIS 2008 (AINA2008 Workshops), pp. 1625–1630. IEEE Computer Society Press, Los Alamitos (2008)
6. Kamada, T.: Compact HTML for small information appliances. W3C Note, February 9, 1998 (1998), http://www.w3.org/TR/1998/NOTE-compactHTML-19980209
7. R Development Core Team: R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria (2005) ISBN 3-900051-07-0

Preventing Illegal Information Flow Based on Role-Based Access Control Model

Tomoya Enokido[1] and Makoto Takizawa[2]

[1] Rissho University, Japan
eno@ris.ac.jp
[2] Seikei University, Japan
makoto.takizawa@st.seikei.ac.jp

Abstract. In the role-based access control (RBAC) model, authorized access requests are specified in roles. However, illegal information flow might occur as the well known confinement problem. We first define legal, independent, illegal, and possibly illegal types of information flow relations, $R_1 \Rightarrow R_2$, $R_1 \parallel R_2$, $R_1 \hookrightarrow R_2$, and $R_1 \rightarrow R_2$ among role families R_1 and R_2, respectively. Suppose a transaction T_1 with a role family R_1 precedes T_2 with R_2 in a schedule, i.e. for every pair of conflicting methods op_1 and op_2 from T_1 and T_2, respectively, op_1 is performed prior to op_2. Here, if $R_1 \Rightarrow R_2$ or $R_1 \parallel R_2$ hold, no illegal information flow occur. Otherwise, illegal information flow might occur. Hence, T_2 cannot be performed. In this paper, we discuss the locking protocol for synchronizing conflicting transactions so that no illegal information flow occur based on the information flow relations. In addition, we discuss when the role-based locks are released.

1 Introduction

Information systems have to be consistent and secure in presence of conflicting accesses from multiple transactions to resource objects. In the access control models [2], a subject s, i.e. user or program is allowed to issue a method op to an object o only if an access right (or permission) $\langle o, op \rangle$ is granted to the subject s. In the role-based access control (RBAC) models [9,10,13,15,16,17], a role is a collection of access rights, which shows what subjects playing the role can do on resource objects in an enterprise. However, illegal information flow among subjects through objects may occur even if every access request is authorized. There is a well-known *confinement* problem [3,12].

A subject s issues a transaction T to manipulate objects. In the RBAC model, the transaction T is assigned with a subfamily of the roles granted to the subject s. The subfamily is referred to as *purpose* of the subject s to perform the transaction T [5,6,7]. Let T_1 and T_2 be a pair of transactions with purposes R_1 and R_2, respectively. Suppose T_1 writes an object y after reading an object x and then T_2 reads y. Here, if T_2 is not granted an access right $\langle x, read \rangle$, T_2 *illegally* reads x's data from the object y, i.e. illegal information flow occur. If R_2 includes $\langle o, read \rangle$ for every read access right $\langle o, read \rangle$ in R_1, no illegal information flow

M. Takizawa, L. Barolli, and T. Enokido (Eds.): NBiS 2008, LNCS 5186, pp. 242–251, 2008.
© Springer-Verlag Berlin Heidelberg 2008

occur. Here, R_1 is referred to as *legally flow* into R_2 ($R_1 \Rightarrow R_2$). In addition, we define types of information flow relations, independent ($R_1 \parallel R_2$), illegal ($R_1 \hookrightarrow R_2$), and possibly illegal ($R_1 \rightarrow R_2$) relations in this paper. Here, suppose a transaction T_1 with purpose R_1 is performed and another transaction T_2 with purpose R_2 is initiated. Here, T_2 can be performed if $R_1 \Rightarrow R_2$ or $R_1 \parallel R_2$. Otherwise, since illegal information flow might occur, T_2 is aborted. We discuss a locking mechanism to prevent illegal information flow where each object just keeps a role family of a transaction which has most recently manipulated the object. Role-based locks can be released if information flow to the objects set obsolete. We discuss how to release locks.

In section 2, we present a system model. In section 3, we define the types of information flow relations. In section 4, we discuss the synchronization algorithm of multiple conflicting transactions to prevent illegal information flow.

2 System Model

2.1 Role-Based Access Control (RBAC) Model

An information system is composed of two types of entities, *subjects* and *objects*. A subject issues methods to objects. Users and programs are examples of subjects and databases are examples of objects. Let **S** be a set of subjects and **O** be a set of objects o_1, ..., o_m in a system. Each object o_i is assumed to support *read* and *write* methods for manipulating data in o_i ($i = 1, ..., m$). Only a subject s granted an access right (or permission) $\langle o, op \rangle$ is allowed to issue a method op to an object o in the access control models [2,4,9,10,14,15,17].

In the *role-based* access control (RBAC) models [9,10,15,17], a role is a set of access rights. A *role* shows a job function in an enterprise. Let **R** be a set of roles $\{r_1, ..., r_n\}$ in a system. Each role r_i is a collection $\{\alpha_{i1}, ..., \alpha_{il_i}\}$ ($l_i \geq 1$) of access rights. Let $SR(s)$ ($\subseteq$ **R**) be a family of roles granted to a subject s. That is, the subject s plays roles in $SR(s)$. If s initiates a transaction T, T is assigned with a subfamily $PR(T)$ of the roles granted to the subject s ($\subseteq SR(s)$). $PR(T)$ is referred to as *purpose* [5,6,7] of the subject s to issue the transaction T. T issues a method op to an object o for an access right $\langle o, op \rangle$ in the purpose $PR(T)$.

2.2 Transactions

A *write* method *conflicts* with *write* and *read* while *read* conflicts with *write* on an object since the result obtained by performing the methods depends on the computation order [8,11]. A *schedule* S of a transaction set **T** shows an execution sequence of methods which are issued by the transactions in **T**.

Definition. A transaction T_i *precedes* T_j in a schedule S ($T_i \rightarrow T_j$) iff a method op_i of T_i is performed prior to op_j of T_j where op_i conflicts with op_j or $T_i \rightarrow T_k \rightarrow T_j$ for some transaction T_k in the set **T**.

A pair of transactions T_i and T_j are *independent* ($T_i \parallel T_j$) iff neither $T_i \rightarrow T_j$ nor $T_j \rightarrow T_i$. A schedule S is *serializable* iff the precedent relation $\rightarrow$ of the transactions in S is acyclic [1]. Conflicting transactions should be performed in a serializable schedule [8] to keep objects consistent by the locking protocols like two-phase locking protocols [3,8,11,18] and the timestamp ordering (TO) schedulers [1].

3 Types of Information Flow Relations

3.1 Information Flow Relation on Transactions

Let $\mathbf{T}$ be a set of transactions issued to the system. For each role r in the role set $\mathbf{R}$, let $In(r)$ and $Out(r)$ show *input* and *output* sets of objects $\{o \mid \langle o, read\rangle \in r\}$ and $\{o \mid \langle o, write\rangle \in r\}$ ($\subseteq \mathbf{O}$), respectively. Let $In(R)$ and $Out(R)$ be a *input* set $\cup_{r \in R} In(r)$ and *output* set $\cup_{r \in R} Out(r)$ of objects for a role family R ($\subseteq \mathbf{R}$), respectively. Let $In(T)$ and $Out(T)$ be input and output sets of objects which a transaction T reads and writes, respectively. Here, $In(T) \subseteq In(PR(T))$ and $Out(T) \subseteq Out(PR(T))$ for a purpose $PR(T)$.

Suppose a transaction T_1 precedes T_2 in a schedule S ($T_1 \rightarrow T_2$).

1. T_1 *reads from* T_2 ($T_1 \vdash T_2$) iff $Out(T_1) \cap In(PR(T_2)) \neq \phi$ or $T_1 \vdash T_3 \vdash T_2$ for some transaction T_3.
2. T_1 and T_2 are *independent* ($T_1 \parallel T_2$) iff $Out(T_1) \cap In(PR(T_2)) = \phi$.

It is noted that the relation $\vdash$ is transitive but $\parallel$ is not transitive. Now, we define types of flow relations from T_1 to T_2 where $T_1 \rightarrow T_2$ in a schedule S:

1. T_2 *legally* reads from T_1 ($T_1 \Rightarrow T_2$) iff $In(T_1) \subseteq In(PR(T_2))$ and $T_1 \vdash T_2$.
2. T_2 *illegally* reads from T_1 ($T_1 \hookrightarrow T_2$) iff $Out(T_1) \subseteq In(PR(T_2))$ and $T_1 \vdash T_2$.
3. T_2 *possibly illegally* reads from T_1 ($T_1 \mapsto T_2$) iff $T_1 \vdash T_2$ but neither $T_1 \Rightarrow T_2$ nor $T_1 \hookrightarrow T_2$.

The illegal relation $T_1 \hookrightarrow T_2$ means that T_2 is not granted an access right to read some object read by T_1. That is, T_2 might read data in an object even if T_2 has no read access right on the object. In a schedule of the transaction set $\mathbf{T}$, illegal information flow occur iff there are a pair of transactions T_1 and T_2 such that $T_1 \hookrightarrow T_2$ and T_1 precedes T_2.

For a transaction T_1 in a schedule S, let $ST(T_1)$ be a subset $\{T_2 \mid T_2 \vdash T_1\} \subseteq \mathbf{T}$, transactions from which T_1 reads. $ST(T_1)$ is a *source* set of T_1. For each transaction T, $In(T)$ and $Out(T)$ might not be *a priori* defined before performing T while the purposes are *a priori* defined.

3.2 Information Flow Relations on Roles

Definition. A role r_1 *legally flows* into a role r_2 ($r_1 \Rightarrow r_2$) iff $In(r_1) \neq \phi$ and $In(r_1) \subseteq In(r_2)$ and $Out(r_1) \cap In(r_2) \neq \phi$ [Figure 1].

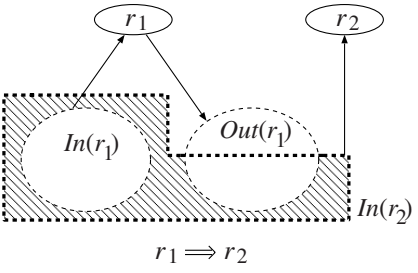


Fig. 1. Legal information flow relation

The relation $\Rightarrow$ is transitive. A role r_1 is *legally equivalent* with r_2 ($r_1 \equiv r_2$) iff $r_1 \Rightarrow r_2$ and $r_2 \Rightarrow r_1$. Based on the relation $\Rightarrow$ among roles, a *least upper bound* (*lub*) $r_1 \sqcap r_2$ of roles r_1 and r_2 is defined to be a role r_3 such that $r_1 \Rightarrow r_3$, $r_2 \Rightarrow r_3$, and there is no role r_4 such that $r_1 \Rightarrow r_4 \Rightarrow r_3$ and $r_2 \Rightarrow r_4 \Rightarrow r_3$. A *greatest lower bound* (*glb*) $r_1 \sqcup r_2$ can be defined similarly to the *lub*. Thus, an LIF lattice $\langle \mathbf{R}, \Rightarrow, \sqcup, \sqcap \rangle$ is defined.

Definition. A role r_1 is *legally independent* of a role r_2 ($r_1 \mid r_2$) iff $In(r_1) \neq \phi$ and $In(r_1) \subseteq In(r_2)$ and $Out(r_1) \cap In(r_2) = \phi$ [Figure 2].

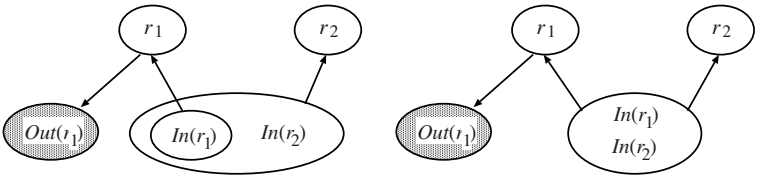


Fig. 2. Legally independent information flow relation ($r_1 \mid r_2$)

Definition. $r_1 \Rightarrow^+ r_2$ iff $r_1 \Rightarrow r_2$ or $r_1 \mid r_2$.

Theorem 1. For every pair of roles r_1 and r_2, $r_1 \Rightarrow^+ r_2$ if $r_1 \Rightarrow^+ r_3$ and $r_3 \Rightarrow^+ r_2$ for some role r_3.

Proof. $In(r_1) \subseteq In(r_3) \subseteq In(r_2)$ from the definition.

Definition. A role r_1 is *independent* of a role r_2 ($r_1 \parallel r_2$) iff 1) $In(r_1) = \phi$ or 2) $In(r_1) \neq \phi$ and ($Out(r_1) \cap In(r_2) = \phi$ and $In(r_1)$ - $In(r_2) \neq \phi$) [Figure 3].

It is noted that the relation $\mid$ is transitive but not symmetric while $\parallel$ is neither transitive nor symmetric. The lub $\sqcap^+$ and glb $\sqcup^+$ are defined for the relation $\Rightarrow^+$ in a same manner as $\sqcap$ and $\sqcup$.

Definition. A role r_1 *possibly illegally flows* into a role r_2 ($r_1 \mapsto r_2$) iff $r_1 \not\Rightarrow r_2$ and $r_1 \not| r_2$ and $r_1 \not\parallel r_2$.

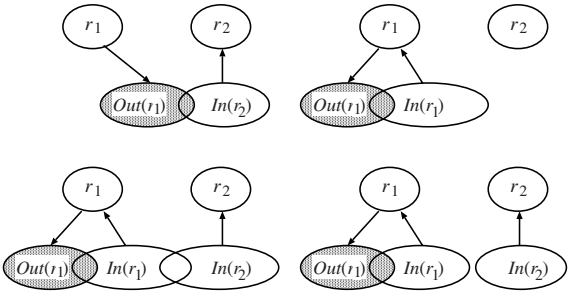


Fig. 3. Independent information flow relation ($r_1 \parallel r_2$)

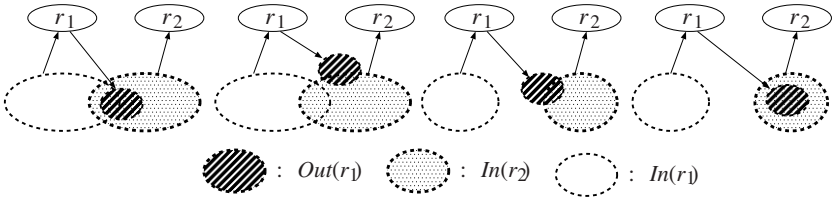


Fig. 4. Possibly illegal information flow relation ($r_1 \mapsto r_2$)

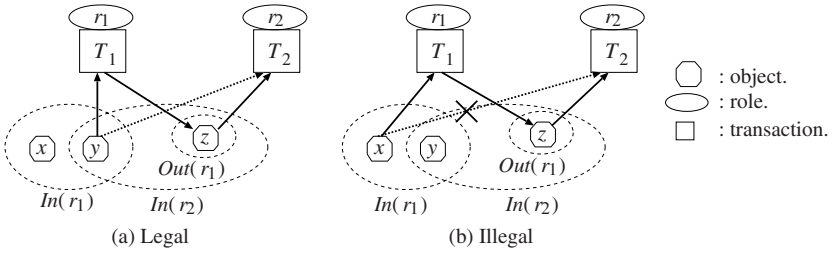


Fig. 5. Possibly illegal Information flow relations

The relation $r_1 \mapsto r_2$ satisfies the following conditions:

- $In(r_1) \neq \phi$ and $Out(r_1) \neq \phi$ and $Out(r_1) \cap In(r_2) \neq \phi$ [Figure 4].

Let us consider a pair of transactions T_1 and T_2 with roles r_1 ($=\{\langle x, read\rangle, \langle y, read\rangle, \langle z, write\rangle\}$) and r_2 ($=\{\langle y, read\rangle, \langle z, read\rangle\}$), respectively [Figure 5]. If T_1 writes the object z after reading y and then T_2 reads z, there is no illegal information flow since T_2 is granted a read right $\langle y, read\rangle$ in r_2. However, if T_1 writes z after reading x, illegal information flow occur since T_2 is not granted $\langle x, read\rangle$ in r_1. Thus, whether or not illegal information flow to occur depends on which objects are read and written in transactions.

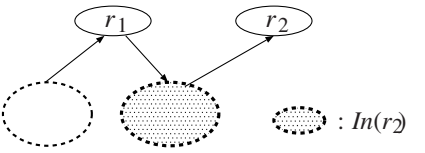


Fig. 6. Illegal information flow relation ($r_1 \hookrightarrow r_2$)

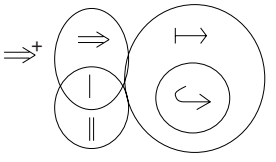


Fig. 7. Relation among information flow relations

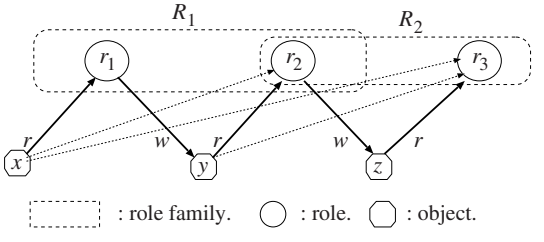


Fig. 8. $R_1 \Rightarrow R_2$

Definition. A role r_1 *illegally flows* into a role r_2 ($r_1 \hookrightarrow r_2$) iff $r_1 \mapsto r_2$ and $Out(r_1) = In(r_2)$ [Figure 6].

This means that the role r_1 necessarily illegally flows into r_2. The relations $\mapsto$ and $\hookrightarrow$ are neither transitive nor symmetric. Figure 7 shows how the relations $\Rightarrow$, $|$, $\Rightarrow^+$, $\|$,$\mapsto$, and $\hookrightarrow$ are related. For example, $\mapsto$ implies $\hookrightarrow$.

3.3 Information Flow on Role Families

A subject s can be granted one or more than one role $SR(s)$ ($\subseteq \mathbf{R}$) in a system. There are the following relations among role families R_1 and R_2:

- R_1 *legally flows* into R_2 ($R_1 \Rightarrow^+ R_2$) iff $\sqcup^+_{r_1 \in R_1} r_1 \Rightarrow^+ \sqcap^+_{r_2 \in R_2} r_2$. The relation $\Rightarrow^+$ is transitive from the definition.
- R_1 is *legally equivalent* with R_2 ($R_1 \equiv R_2$) iff $R_1 \Rightarrow^+ R_2$ and $R_2 \Rightarrow^+ R_1$.
- R_1 is *legally independent* of R_2 ($R_1 \parallel R_2$) iff $\sqcup^+_{r_1 \in R_1} r_1 \parallel \sqcup^+_{r_2 \in R_2} r_2$.

Suppose there are a pair of role families $R_1 = \{r_1, r_2\}$ and $R_2 = \{r_2, r_3\}$ as shown in Figure 8. Here, $r_1 = \{\langle x, read\rangle, \langle y, write\rangle\}$, $r_2 = \{\langle x, read\rangle, \langle y, read\rangle, \langle z, write\rangle\}$, and $r_3 = \{\langle x, read\rangle, \langle y, read\rangle, \langle z, read\rangle\}$. $r_1 \Rightarrow r_2 \Rightarrow r_3$. Here, $\sqcup^+_{r \in R_1} r$

$= r_1 \sqcup^+ r_2 = r_2$ since $r_1 \Rightarrow r_2$ and there is no role r' such that $r_1 \Rightarrow r' \Rightarrow r_2$. Similarly, $\sqcap^+_{r \in R_2} r = r_2 \sqcap^+ r_3 = r_2$. Then, $R_1 \Rightarrow R_2$ since $\sqcup^+_{r \in R_1} r \Rightarrow \sqcap^+_{r \in R_2} r$.

Theorem 2. Let R_1, R_2, and R_3 be role families. $R_2 \not\Rightarrow^+ R_3$ if $R_1 \Rightarrow^+ R_2$ but $R_1 \not\Rightarrow^+ R_3$.

4 Synchronization of Transactions

4.1 Role-Based Locking Protocol

A transaction T_t locks an object o before performing a method op on the object o. If another transaction T_u locks o, T_t waits until T_u unlocks o. In this paper, we assume every transaction takes the strict two phase locking protocol, i.e. a transaction T_t does not issue a lock request after issuing an unlock and every lock held by T_t is released at the end of T_t. Suppose T_t reads an object o_j in *read* before writing o_i in *write*. If an access right $\langle o_j, read \rangle$ is not in $PR(T_u)$, illegal information flow occur. Otherwise, no illegal information flow occur. Thus, whether or not illegal information flow occur depends on what method is performed on what object in what order. Suppose $read_{ui}$ for T_u is issued to o_i after $write_{ti}$ of T_t is performed on o_i. If $PR(T_t) \Rightarrow PR(T_u)$ or $PR(T_t) \parallel PR(T_u)$, no illegal information flow occurs. Otherwise, illegal information flow might occur. Hence, $read_{ui}$ cannot be performed, i.e. T_u is aborted.

Each time a method op_{ti} is issued to an object o_i, we check if illegal information flow might occur by performing op_{ti}. A variable $o_i.P$ denotes independent purposes of transactions which have most recently written o_i. $o_i.P = \phi$ in the initialization of the system. A variable $T_t.P$ is a set of independent purposes, initially $T_t.P = \{PR(T_t)\}$. Suppose a transaction T_t issues a method op_{ti} to an object o_i.

Perform(T_t, op_{ti}, o_i) {
  if an object o_i could be locked in a locking protocol, {
    if $op_{ti} = write$, {
      $o_i.P = T_t.P$;
      write o_i; }
    else { /* $op_{ti} = read$ */
      if $R \Rightarrow^+ R_t$ **or** $R \parallel R_t$ for every R in $o_i.P$ and every R_t in $T_t.P$ {
        read o_i; /* no illegal information flow */
        for each R **in** $o_i.P$ and R_t in $T_t.P$, {
          if $R \parallel R_t$, $T_t.P = T_t.P \cup \{R\}$; }
      } **else abort** T_t; }
  } **else wait**;
}

If op_{ti} is *write*, a purpose $PR(T_t)$ of the transaction T_t is stored in a variable $o_i.P$ and op_{ti} is performed on the object o_i. Next, suppose op_{ti} is *read* and $o_i.P$ shows a purpose $PR(T_u)$ of a transaction T_u which most recently writes

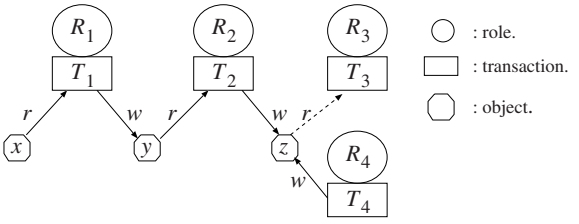


Fig. 9. Information flow check

o_i. If $o_i.P \Rightarrow^+ PR(T_t)$ or $o_i.P \parallel PR(T_t)$, T_t reads o_i. If $o_i.P \parallel PR(T_t)$, $T_t.P = \{PR(T_t),\ PR(T_u)\}$. Otherwise, T_t should be aborted. Here, if $o_i.P \mapsto PR(T_t)$ but $o_i.P \not\rightsquigarrow PR(T_t)$, T_t might not illegally read T_u. In order to decide whether $T_t \Rightarrow T_u$ or $T_t \hookrightarrow T_u$, we need more information than $o_i.P$ and $PR(T_t)$.

Suppose there are four transactions T_1, T_2, T_3, and T_4 with purposes R_1, R_2, R_3, and R_4, respectively, as shown in Figure 9. Suppose $R_1 \Rightarrow R_2$. T_1 reads an object x and then writes another object y. Then, T_2 reads y and then writes an object z. First, suppose T_3 reads z. In **Perform**(T_3, *read*, z), R_2 ($= PR(T_2)$) is stored in a variable $z.P$. If $z.P \Rightarrow PR(T_3)$, i.e. $R_2 \Rightarrow R_3$, T_3 can read z. Here, suppose $R_1 \not\Rightarrow R_3$ or $R_1 \not\parallel R_3$. Data in x might be brought to z through T_1 and T_2. T_3 should not read z since $R_1 \not\Rightarrow R_3$ or $R_1 \not\parallel R_3$. According to the theorem, $R_2 \not\Rightarrow R_3$ if $R_1 \not\Rightarrow R_3$ even if $R_1 \Rightarrow R_2$. Hence, $R_1 \not\Rightarrow R_3$ and $R_2 \Rightarrow R_3$ may not hold.

4.2 Release of Lock

Suppose a transaction T_t performs a *write* method on an object o_i and then $o_i.P = \{PR(T_t)\}$. Here, another transaction T_u issues a *read* method to o_i. If neither $o_i.P \Rightarrow^+ PR(T_u)$ nor $o_i.P \parallel PR(T_u)$, T_u is aborted in the protocol. Thus, any transaction T_u where $PR(T_t) \hookrightarrow PR(T_u)$ cannot be performed once T_t is performed on o_i. The role-based locks are released, i.e. $o_i.P = \phi$ if the following conditions hold:

1. It takes some time after $o_i.P$ is updated.
2. An object in $In(T_t)$ is updated by another transaction.

Each transaction T_t is assigned timestamp $T_t.TS$ which shows when T_t is initiated. Each time an object o_i is updated by T_t, $o_i.TS = max(o_i.TS, T_t.TS)$.

Suppose that a transaction T_t reads an object o_i and then writes another object o_j. T_t keeps in record what objects T_t has read. Since T_t reads the object o_i, $T_t.o_i$ shows that T_t has read an object o_i. Here, $T_t.o_i.TS$ indicates the timestamp of o_i when T_t has read o_i. Then, T_t writes o_i. The object o_i keeps in a record what objects the transaction T_t has read. Here, $o_i.P.o_j$ is recorded when P shows $T_i.P$. $o_i.P.o_j.TS = T_t.o_j.TS$.

Suppose another transaction T_u reads the object o_j. Here, T_u cannot read the object o_j unless $o_i.P \Rightarrow^+ PR(T_u)$ or $o_i.P \parallel PR(T_u)$. For every object o in $o_j.P$,

timestamp $o_j.P.o.TS$ is compared with $o.TS$. If $o_j.P.o.TS < o.TS$, the object o_j may have data in o but the data is obsolete. Here, the role families in $PR(T.P)$ are removed from $o_j.P$. Otherwise, if $(T_u.TS$ - $o.P.o.TS) > \delta$, the role families in P are removed from $o_j.P$ because the data from the object o is obsolete. δ is a constant which is decided based on the semantics of the object o.

5 Concluding Remarks

In information systems, multiple transactions issue conflicting *read* and *write* methods to an object. The RBAC models are widely used in information systems, especially database systems. In the access control models, there might occur confinement problem, i.e. illegal information flow occur even if only authorized subjects manipulate objects in authorized ways. We take a simple model where each object supports a pair of *read* and *write* methods. In this paper, we define a legal $\Rightarrow$, independent $\parallel$ and $|$, illegal $\hookrightarrow$, and possibly illegal $\mapsto$ information flow relations among role families. If $R_1 \Rightarrow R_2$ or $R_1 \parallel R_2$ for role families R_1 and R_2, no illegal information flow occur if a transaction with purpose R_1 is performed prior to a transaction with R_2. On the other hand, if $R_1 \hookrightarrow R_2$, illegal information flow surely occurs. If $R_1 \mapsto R_2$, illegal information flow might occur depending on which methods each transaction issues in which order.

We discussed the locking protocol to prevent illegal information flow. Then, a method is performed on an object only if no illegal information flow would occur. Otherwise, the transaction issuing the method is aborted. Differently from the locking protocols, role-based locks are not released even if the transaction terminates. We discuss how to release obsolete role-based locks based on timestamps of objects.

References

1. Bernstein, P.A., Hadzilacos, V., Goodman, N.: Concurrency Control and Recovery in Database Systems. Addison-Wesley, Reading (1987)
2. Bertino, E., Samarati, P., Jaodia, S.: High Assurance Discretionary Access Control in Object Bases. In: Proc. of the 1st ACM Conf. on Computers and Communication Security, pp. 140–150 (1993)
3. Chon, R., Enokido, T., Takizawa, M.: Inter-Role Information Flow in Object-based Systems. In: Proc. of IEEE 18th International Conf. on Advanced Information Networking and Applications (AINA 2004) (to appear, 2004)
4. Denning, D.E.: A Lattice Model of Secure Information Flow. Communications of the ACM 19(5), 236–343 (1976)
5. Enokido, T., Takizawa, M.: Concurrency Control Based-on Significancy on Roles. In: Proc. of the IEEE 11th International Conference on Parallel and Distributed Systems (ICPADS 2005), pp. 196–202 (2005)
6. Enokido, T., Takizawa, M.: Role-Based Concurrency Control for Distributed Systems. In: Proc. of the IEEE 20th International Conference on Advanced Information Networking and Applications (AINA 2006), pp. 407–412 (2006)

7. Enokido, T., Takizawa, M.: Concurrency Control using Subject- and Purpose-Oriented (SPO) View. In: Proc. of the 2nd International Conference on Availability, Reliability and Security (ARES 2007), pp. 454–461 (2007)
8. Eswaran, K.P., Gray, J.N., Lorie, R.A., Traiger, I.L.: The Notions of Consistency and Predicate Locks in a Database System. Communications of the ACM 19(19), 624–633 (2007)
9. Ferraiolo, D., Kuhn, R.: Role-Based Access Controls. In: Proc. of 15th NIST-NCSC National Computer Security Conf., pp. 554–563 (1992)
10. Ferraiolo, D.F., Kuhn, D.R., Chandramouli, R.: Role Based Access Control. Artech House (2005)
11. Gray, J.: Notes on Database Operating Systems. LNCS, vol. (60), pp. 393–481 (1978)
12. Izaki, K., Tanaka, K., Takizawa, M.: Information Flow Control in Role-Based Model for Distributed Objects. In: Proc. of IEEE International Conf. on Parallel and Distributed Systems (ICPADS 2001), pp. 363–370 (2001)
13. Oracle Corporation. Oracle8i Concepts Vol. 1, Release 8.1.5 (1999)
14. Sandhu, R.S.: Lattice-Based Access Control Models. IEEE Computer 26(11), 9–19 (1993)
15. Sandhu, R.S., Coyne, E.J., Feinstein, H.L., Youman, C.E.: Role-Based Access Control Models. IEEE Computer 29(2), 38–47 (1996)
16. Sybase. Sybase SQL Server, http://www.sybase.com/
17. Tari, Z., Chan, S.W.: A Role-Based Access Control for Intranet Security. IEEE Internet Computing 1, 24–34 (1997)
18. Watanabe, K., Sugiyama, Y., Enokido, T., Takizawa, M.: Moderate Concurrency Control in Distributed Object Systems. Journal of Interconnection Networks (JOIN) 5(3), 233–247 (2004)

Authentication Binding between TLS and HTTP

Takamichi Saito, Kiyomi Sekiguchi, and Ryosuke Hatsugai

Meiji University, Japan

Abstract. As Secure Socket Layer or Transport Layer Security (SSL/TLS) provides secure communication over the internet, there is no password-based authentication mechanism in original one. Then web application utilizes Hyper Text Transport Protocol (HTTP) authentication such like basic or digest with and over SSL/TLS. In the way, there are two authentications in HTTP over SSL/TLS session separately. It could be improper rather than inconvenient to establish secure communication. Then, we provide a way of authentication binding between them without modifying the protocols and implementation to show its effectiveness.

1 Introduction

Secure Socket Layer [1] or Transport Layer Security (TLS)[2] (SSL/TLS) is standard to make communication secure over the internet. SSL/TLS lies in the transport layer of the OSI model and provides *communication security* for upper layers. However, original SSL/TLS itself does not provide password-based authentication scheme, partially because a password-based authentication can be required to be in the application layer.

When web application server needs to identify a user, it utilizes Hyper Text Transport Protocol (HTTP) authentication scheme such like basic or digest authentication over SSL/TLS in *server authentication mode*. In the case, however, web server manages two security channels separately in each layer. Namely, web browser authenticates server by SSL/TLS while the server authenticates the user by HTTP authentication scheme.

This popular way could not be better to deploy in two reasons. As described above, the first reason is that there are two separated authentications in each layer. It means that there are two types of authentications in a session. Web application programmer must manage to obtain user's credential in a server side. The second one is that, since authentication of SSL/TLS is independent of that of HTTP, it can be vulnerable to man-in-the-middle (MITM) attack. When we need stronger security for communication, we select to apply mutual authentication by using SSL/TLS in the *client authentication mode*. However, as the mode enforces client to require its X.509 certificate with PKI (Public Key Infrastructure), most web-application server avoids utilizing the mode. Moreover, due to duplication of user authentication, one of user authentications should be eliminated from the scheme.

M. Takizawa, L. Barolli, and T. Enokido (Eds.): NBiS 2008, LNCS 5186, pp. 252–262, 2008.
© Springer-Verlag Berlin Heidelberg 2008

There are *built-in* ways of password-based authentication, modifying SSL/TLS Handshake protocol such like RFC standard [8] or using SRP [5,6]. However, this kind of solution incurs shortcomings. Since SSL/TLS lies in the transport layer, client such as web browser needs to prepare a user interface mechanism for obtaining its password when establishing SSL/TLS channel. A password itself to input into client can be stored to obtain before establish channel. Nevertheless, if it is wrong password in case of human error, web browser and SSL/TLS server need to handshake once more from the beginning. It wastes much computational cost for keeping *SSL/TLS context*, especially in web server.

In this paper, we review related works and concepts, provide a way of binding authentications in secure way without modifying protocols.

2 Preliminary

2.1 Notations

Hosts. SSL/TLS communication consists of client C and server S. In addition, attacker that could be active and/or passive is denoted as $\mathcal{I}$. In case of masquerading as server S, the attacker is denoted as $\mathcal{I}(S)$. In masquerading as client C, the attacker is denoted as $\mathcal{I}(C)$.

Exchanged Messages. A symbol Msg means an arbitrary message. For example, an expression that C sends Msg to S is denoted as the follow:

$$C \rightarrow S \ : \ Msg$$

When $\mathcal{I}$ masquerading as S sends Msg to C, it is denoted as the follow:

$$\mathcal{I}(S) \rightarrow C \ : \ Msg$$

Then, note that the following two expressions are different meanings:

$$\begin{array}{ll} C \rightarrow \mathcal{I}(S) & : \ Msg \\ C \rightarrow \mathcal{I} & : \ Msg \end{array}$$

The first one means that C regards the receiver as S, but in fact, C sends it to $\mathcal{I}$. On the other, the second means that C regards $\mathcal{I}$ as one of legitimate responders and sends it. Namely, C trusts $\mathcal{I}$ as a legitimate server.

Messages. A parameter generated by host C is expressed as X_C, which means an arbitrary message X is generated by C.

SA is *cipher suite list* composed of *protocol version*, *public key algorithm*, *compressed algorithm*, *bulk cipher algorithm* and *MAC algorithm*, which specifies supported ones by each host. RN consists of two components: *date*, *time* and 28 bytes *random number*. S_ID means *SSL/TLS session identifier*, which is held by each host, and it is referred in session resumption. A symbol *cert_list* is an *X.509 certificate chain*. CN is *common name* of server specified in an

X.509 certificate chain. A symbol *cert_auth* is a parameter to request client to send back the client certificate, and *cert_type* indicates a type of acceptable certificate. A symbol *pre_master_secret*, from which the session key is derived, is concatenation of 46 bytes *random number* and *protocol version*. A symbol *message_all* shows all messages exchanged between client and server so far. *Ack* is a parameter to inform the opposite side that all necessary messages were transmitted to establish SSL/TLS connections.

Symbol *user_id* and *password*, which are utilizing in HTTP authentication, denote the identifier of the user and its password respectively.

Encryption. A public key is denoted as P, then P_C means the public key owned with C. $\{Msg\}_Y$ means that Msg is encrypted with an encryption key Y. A secret key of corresponding to the public key P is denoted as P^{-1}, then $\{Msg\}_{P^{-1}}$ means that Msg is signed with P^{-1}. KS is a *session key* generated with RNs and *pre_master_secret*, e.g., KS_{CS} or KS_{SC} is a session key between C and S. $h(Msg_1)$ is a hashed value of Msg_1. $h(Msg_1, Msg_2, \cdots, Msg_i)$ is also a hashed value which is calculated after concatenating $Msg_1, Msg_2, \cdots$, and Msg_i.

2.2 HTTP Authentication over SSL/TLS

Although SSL/TLS supports RSA, DH (Diffie-Hellman) and Fortezza as *Key Exchange Algorithm*, all ones are not distinguished here. Therefore, we only discuss the case of SSL/TLS Handshake using RSA as Key Exchange Algorithm in server authentication mode.

The following run of protocol is SSL/TLS Handshake in server authentication mode:

Table 1. SSL/TLS Handshake in Server Authentication Mode

$M1)\ C \rightarrow S : SA_C, S\_ID_C, RN_C$
$M2)\ S \rightarrow C : SA_S, S\_ID_S, RN_S$
$M3)\ S \rightarrow C : cert\_list_S$
$M4)\ S \rightarrow C : Ack_S$
$M5)\ C \rightarrow S : \{pre\_master\_secret_C\}_{P_S}$
$M6)\ C \rightarrow S\ : \{h(KS_{CS}, h(message\_all_{CS}, C, KS_{CS}))\}_{KS_{CS}}$
$M7)\ S \rightarrow C\ : \{h(KS_{CS}, h(message\_all_{CS}, S, KS_{CS}))\}_{KS_{CS}}$

After SSL/TLS Handshake, since C and S share a session key KS_{CS} in secure manner, they can encrypt their communication with it.

In case of using server authentication mode, web server cannot identify a connecting user. Therefore, when it requires a user authentication, it can utilize two type of HTTP authentication after SSL/TLS Handshake. HTTP authentication can also be encrypted such like this:

Table 2. Basic Authentication of HTTP over SSL/TLS

$$
\begin{array}{lll}
M8) & C \rightarrow S & : \{[Request\ Some\_page]\}_{KS_{CS}} \\
M9) & S \rightarrow C & : \{[Authorization\ Request]\}_{KS_{CS}} \\
M10) & C \rightarrow S & : \{user\_id_C, password_C\}_{KS_{CS}} \\
M11) & S \rightarrow C & : \{[Requested\_page]\}_{KS_{CS}}
\end{array}
$$

Table 3. Digest Authentication of HTTP over SSL/TLS

$$
\begin{array}{lll}
M8) & C \rightarrow S & : \{[Request\ Some\_page]\}_{KS_{CS}} \\
M9) & S \rightarrow C & : \{[Authorization\ Request]\}_{KS_{CS}} \\
M10) & C \rightarrow S & : \{h(user\_id_C, password_C)\}_{KS_{CS}} \\
M11) & S \rightarrow C & : \{[Requested\_page]\}_{KS_{CS}}
\end{array}
$$

In addition, web server selects the protocol that password is hashed. This option can specified in the configuration of web server such like Apache.

However, this authentication is not secure after when using over SSL/TLS.

After establishing SSL/TLS connection, triggered by the password request message, e.g. `"401 Authorization Required"` from web server, web browser pops up a window to obtain a password of the user. After the user puts his/her ID and its password, the browser sends them to the server. Hereafter, when the authenticated client sends a request massage, it attaches the client's credential to the message.

2.3 Related Works

As described before, there are some proposals to integrate password-based user authentication into SSL/TLS protocol itself.

Protocol Extensions. One of the most promising proposals is SSL/TLS extension with using SRP [5,6], which is also discussing as IETF draft recently. In addition, there is more standardized one [8].

These schemes are just extensions of SSL/TLS Handshake to include password-based user authentication. Then, these approaches force SSL/TLS protocol to be modified in client and server. An authenticated SSL/TLS server can authenticate connecting user just in SSL/TLS Handshake. Therefore, when the user put wrong password, SSL/TLS Handshake must be negotiated from the beginning.

Another standardized choice is using SSL/TLS in client authentication mode (see Table 4). This is famous for most secure in the protocols. As you can find that client signs the message $M8$, client needs to prepare its public key as its certificate for the web server. However, it could be hard to introduce in every case, since user in the client has to obtain an X.509 certificate for the application.

Channel Binding. The basic concept of *channel binding* is to unify two security channels, integrating upper channels into a lower one [9]. For examples, when web server requires SSL/TLS authentication over IPsec such like Virtual

Table 4. SSL/TLS Handshake in Client Authentication Mode

$M1)\ C \rightarrow S : SA_C, S\_ID_C, RN_C$
$M2)\ S \rightarrow C : SA_S, S\_ID_S, RN_S$
$M3)\ S \rightarrow C : cert\_list_S, cert\_auth$
$M4)\ S \rightarrow C : cert\_type, Ack_S$
$M5)\ C \rightarrow S : \{pre\_master\_secret_C\}_{P_S}$
$M6)\ C \rightarrow S : \{h(KS_{CS}, h(message\_all_{CS}, KS_{CS})\}_{P_C^{-1}}$
$M7)\ C \rightarrow S : \{h(KS_{CS}, h(message\_all_{CS}, C, KS_{CS}))\}_{KS_{CS}}$
$M8)\ S \rightarrow C : \{h(KS_{CS}, h(message\_all_{CS}, S, KS_{CS}))\}_{KS_{CS}}$

Private Network (Virtual Private Network), this duplication of security wastes computing power and resource. Then, we aspire to their integration.

Since the concept focuses on channels, we here discuss and reconsider the binding. In the case of binding two channels, especially from the point of view of authentication, there are following three types:

1. Binding two mutual authentications
2. Binding a mutual authentications and a one-way authentication
3. Binding two opposite one-way authentications

The concept of the first and second are apparently duplication of security protocols. They could be integrated by the way of channel binding. Since we focus on authentication, however, we discuss the third case, namely HTTP authentication over SSL/TLS in server authentication mode. HTTP authentication itself does not encrypt its channel and client does not authenticate server, but server authenticates client. Then, binding the two opposite directions of authentication can make secure channel. In this paper, we call this integration as *authentication binding*, and propose a way of authentication binding between SSL/TLS and HTTP authentication in secure manner.

3 Design and Implementation

3.1 Requirements

For designing a new scheme, we enumerate requirements described in the followings:

1. Not modifying the protocols: SSL/TLS, HTTP and HTTPS
2. The timing to obtain a password is after establishing SSL/TLS connection.

Based on SSL/TLS, HTTP and HTTPS, we can inherit three good features such like performance, connectivity and security. Programmer of web applications can also utilize legacy softwares and libraries. Then, we here do not modify the protocols.

Client and server establish SSL/TLS connection in server authentication mode, and then the server authenticates the user by using HTTP authentication in arbitrary time. The server actually as web application can control timing to obtain

user's password although user authentication is required. This feature is more convenient than the case that the timing of authentication is fixed in SSL/TLS Handshake. Moreover, when utilizing a scheme with *reverse proxy* in server authentication mode, the proxy does not need a password file owing to separating password authentication from SSL/TLS protocol. Therefore, we decide not to extend SSL/TLS itself.

For satisfying these requirements, we need the following features in our proposed system:

1. Web application and browser can obtain SSL/TLS session information including *session key* shared with SSL/TLS Handshake.
2. Web application can obtain user's password in plain.

The first one is for binding the secret key in SSL/TLS layer and user's password in HTTP layer. In the second requirement, as this is contrary to standard management, such like a password file in UNIX system, web application needs to access a password in plain since client and server need to bind the above authentications, which is described in latter section.

3.2 System Preliminary

In this paper, we implement the proposed scheme by the following software component modules:

- Web browser with Network Security Service (NSS)
- Apache web server with `mod_ssl`
- Web application using PHP: Hypertext Preprocessor

Where NSS [10] is a library for developing web client and server equipped with security features. It supports to develop web browser with SSL, TLS, PKCS, S/MIME, X.509 certificate and other security standards. Using the library, we modify web browser to obtain the credential of user and server, and a session key shared by SSL/TLS Handshake. In server-side, we implement web application where a session key can be extracted via PHP [11] and `mod_ssl` [12]. Therefore, we prepare an interface to obtain them in each side.

3.3 HTTP Authentication over SSL/TLS

The proposed scheme consists of the following three steps:

1. Establishing SSL/TLS Handshake in server authentication mode
2. HTTP user authentication over the SSL/TLS channel
3. Exchanging and verifying a credential for binding authentications

For achieving this, server and client send the following messages after HTTP authentication over SSL/TLS in server authentication mode:

Table 5. Whole Image of Proposed Scheme

$M1)\ C \to S\ :\quad SA_C, S\_ID_C, RN_C$
$M2)\ S \to C\ :\quad SA_S, S\_ID_S, RN_S$
$M3)\ S \to C\ :\quad cert\_list_S$
$M4)\ S \to C\ :\quad Ack_S$
$M5)\ C \to S\ :\quad \{pre\_master\_secret_C\}_{P_S}$
$M6)\ C \to S\ :\quad \{h(KS_{CS}, h(message\_all_{CS}, C, KS_{CS}))\}_{KS_{CS}}$
$M7)\ S \to C\ :\quad \{h(KS_{CS}, h(message\_all_{CS}, S, KS_{CS}))\}_{KS_{CS}}$
$M8)\ C \to S\ :\quad \{[Request\ Some\_page]\}_{KS_{CS}}$
$M9)\ S \to C\ :\quad \{$ `HTTP/1.1 401 (PJ_Late_Bind_over_ADH-RSA)` $\}_{KS_{CS}}$
$M10)\ C \to S\ :\quad \{user\_id_C, hc_{C\to S}\}_{KS_{CS}}$
$M11)\ S \to C\ :\quad \{CN_S, hc_{S\to C}, [Requested\_page]\}_{KS_{CS}}$

where $hc_{C\to S} = h(CN_S, P_S, h(user\_id_C, password_C, KS_{CS}))$
$hc_{S\to C} = h(user\_id_C, password_C, h(CN_S, P_S, KS_{CS})))$

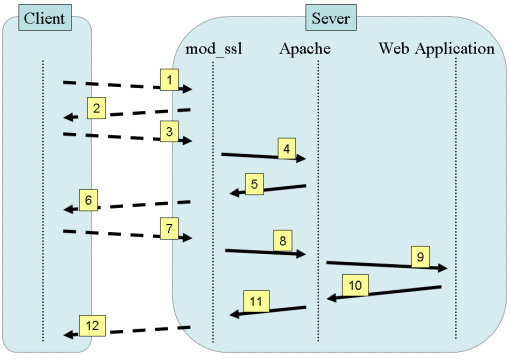


Fig. 1. Binding Process

After establishing SSL/TLS Handshake, client and server share a session key derived from $pre\_master\_secret_C$. The client is requested its credential by the server in application layer, then the client sends ID and $hc_{C\to S}$. In our proposed scheme for binding two authentications, server and client respectively compute a hashed value in local. They compare them: server computes $hc_{C\to S}$ and compare it with a receiving one in $M10$, while client also computes $hc_{S\to C}$ and compare it with a receiving one in $M11$. When they receive opposite-side of it, they verify to compare them. If the local value is same as received one in each side, we can decide that the channel is secure and integrated.

3.4 How the Messages Are Exchanged among Modules

In this subsection, we explain exchanging messages and related computations in modules. The numbers in the figure 1 are corresponding to those in the following descriptions. Where a dotted line is SSL/TLS communication, a solid is inside exchanges in plain.

[1,2] Web browser and server establish SSL/TLS channel by Handshake. These are corresponding to the messages from $M1$ to $M7$ of Table 5.

[3] Client sends the following message to request a page over SSL/TLS. In this case, authentication is required for obtaining `some_page`. This is corresponding to $M8$:

```
"GET <some_page>"
```

[4,5] After receiving it, Apache requests the user's credential by the following message, which is corresponding to $M9$ of Table 5:

```
"HTTP/1.1 401 (PJ_Late_Bind_over_ADH-RSA)"
```

Where `PJ_Late_Bind_over_ADH-RSA` is for a trigger of authentication binding.

[6,7] When receiving this trigger, client starts to bind authentication. Apache module `mod_ssl` encrypts to send it over SSL/TLS channel . Receiving it, the client extracts CN and a public key of the server from the server certificate received in SSL/TLS Handshake. With them, the client constructs a credential message of the user, i.e., $hc_{C \to S}$ to bind authentications. The client sends the following message with it:

```
GET <some_page> Authorization: Basic "XXX"
```

This is corresponding to $M10$ of Table 5. Where `XXX` is the hashed credential $hc_{C \to S}$ that is encoded by Base64.

[8] The module `mod_ssl` decrypts it, and delivers the credential via apache to web application.

[9,10] Web application decodes the Base64 data and extracts $hc_{C \to S}$ from it to compare with the local ones. If they are same, authentication binding is success. If not, failure or possibly attacked.

[11,12] When it is success, apache picks to pass the requested page into `mod_ssl`. It then sends the message $hc_{S \to C}$ encoded with Base64 over SSL/TLS channel. Receiving it, the client decodes Base64 and verifies to compare the receiving $hc_{S \to C}$ and the one computed in local. This message is corresponding to $M11$ of Table 5.

4 Security Considerations

4.1 MITM Attack

There is a possibility to leak a password after establishing SSL/TLS in server authentication mode. We here describe a way of attacking process that password is plundered by attacker masquerading as a legitimate server, while the proposed scheme can prevent it.

Attack Process over SSL/TLS Handshake. As precondition in here, client C trusts attacker $\mathcal{I}$ as one of legitimate servers, and then it connects with $\mathcal{I}$. Namely, C connects with $\mathcal{I}$ by its intent.

Especially, in this case, there is no warning message on the web browser since its certificate could be issued legally for the attacker itself. Therefore, the attacker can arrange the situation by cheating the client, for examples, setting up a *phishing site*.

1: C sends the HTTP request to $\mathcal{I}$.
`GET /index.html HTTP/1.1`
1′: $\mathcal{I}(\mathcal{C})$ forwards it to S.
2: S sends `index.html` to $\mathcal{I}(C)$ as the HTTP response:
`HTTP/1.1 200 OK`
2′: $\mathcal{I}$ forwards it to C.
3: According to a link written in `index.html`, C sends an HTTP request. Then, C is forced to connect with $\mathcal{I}$, and initiates SSL/TLS Handshake in server authentication mode:
`GET /basic_auth/test.html HTTP/1.1`
3′: After establishing SSL/TLS connection with C, $\mathcal{I}(C)$ forwards the HTTP request to S. Namely, $\mathcal{I}(C)$ also initiates another SSL/TLS Handshake with S.
4: S sends an HTTP response with the status code `401` to $\mathcal{I}(C)$ to inform that it requires $user\_id_C$ and $password_C$:
`HTTP/1.1 401 Authorization Required`
4′: $\mathcal{I}$ forwards it to C.
5: After receiving the HTTP response, C again sends the HTTP request with $user\_id_C$ and $password_C$ to $\mathcal{I}(S)$:
`GET /basic_auth/test.html HTTP/1.1`
`Authorization:Basic (`$user\_id_C$: $password_C$`)`
5′: $\mathcal{I}(C)$ forwards it to S. In here, the attacker $\mathcal{I}$ obtains the security credential of the user.
6: After receiving he security credential, S verifies them and sends the HTTP response to $\mathcal{I}(C)$:
`HTTP/1.1 200 OK`
6′: $\mathcal{I}$ forwards it to C.

The table 6 is an explanation of the attack process in more details.

Detecting the Attack. In this subsection, we explain how to detect the attack in our proposed scheme although preconditions are same in the system when the attack happens(see Table 7).

Receiving the message $M8$, the attacker has to send the message $M8'$ or $M8''$ to the server. However, it cannot construct a valid message to success the attack in the next step. In case of sending the message $M8'$, the server can detect the existence of attacking. On the other, the message $M8''$ cannot be constructed by the attacker since the attacker does not have or create the valid password of the user.

Table 6. Attack Process against SSL/TLS Handshake

$M1)\ C \rightarrow \mathcal{I} : SA_C, S\_ID_C, RN_C$
$M1')\ \mathcal{I}(C) \rightarrow S : SA_{\mathcal{I}}, S\_ID_{\mathcal{I}}, RN_{\mathcal{I}}$
$M2)\ S \rightarrow \mathcal{I}(C) : SA_S, S\_ID_S, RN_S$
$M2')\ \mathcal{I} \rightarrow C : SA_{\mathcal{I}}, S\_ID_{\mathcal{I}}, RN_{\mathcal{I}}$
$M3)\ S \rightarrow \mathcal{I}(C): cert\_list_S$
$M3')\ \mathcal{I} \rightarrow C : cert\_list_{\mathcal{I}}$
$M4)\ S \rightarrow \mathcal{I}(C) : Ack_S$
$M4')\ \mathcal{I} \rightarrow C : Ack_{\mathcal{I}}$
$M5)\ C \rightarrow \mathcal{I} : \{pre\_master\_secret_C\}_{P_{\mathcal{I}}}$
$M5')\ \mathcal{I}(C) \rightarrow S : \{pre\_master\_secret_{\mathcal{I}}\}_{P_S}$
$M6)\ C \rightarrow \mathcal{I} : \{h(KS_{C\mathcal{I}}, h(message\_all_{C\mathcal{I}}, C, KS_{C\mathcal{I}}))\}_{KS_{C\mathcal{I}}}$
$M6')\ \mathcal{I} \rightarrow C : \{h(KS_{C\mathcal{I}}, h(message\_all_{C\mathcal{I}}, S, KS_{C\mathcal{I}}))\}_{KS_{C\mathcal{I}}}$
$M7)\ \mathcal{I}(C) \rightarrow S : \{h(KS_{\mathcal{I}S}, h(message\_all_{\mathcal{I}S}, C, KS_{\mathcal{I}S}))\}_{KS_{\mathcal{I}S}}$
$M7')\ S \rightarrow \mathcal{I}(C) : \{h(KS_{\mathcal{I}S}, h(message\_all_{\mathcal{I}S}, S, KS_{\mathcal{I}S}))\}_{KS_{\mathcal{I}S}}$
$M8)\ C \rightarrow \mathcal{I} : \{user\_id_C, password_C\}_{KS_{C\mathcal{I}}}$
$M8')\ \mathcal{I}(C) \rightarrow S : \{user\_id_C, password_C\}_{KS_{\mathcal{I}S}}$

Table 7. Detecting and Preventing the Attack

$M7)\ \mathcal{I}(C) \rightarrow S : \{h(KS_{\mathcal{I}S}, h(message\_all_{\mathcal{I}S}, C, KS_{\mathcal{I}S}))\}_{KS_{\mathcal{I}S}}$
$M7')\ S \rightarrow \mathcal{I}(C) : \{h(KS_{\mathcal{I}S}, h(message\_all_{\mathcal{I}S}, S, KS_{\mathcal{I}S}))\}_{KS_{\mathcal{I}S}}$
$M8)\ C \rightarrow \mathcal{I} : \{user\_id_C, h(CN_S, P_S, h(user\_id_C, Password_C, KS_{C\mathcal{I}}))\}_{KS_{C\mathcal{I}}}$
$M8')\ \mathcal{I}(C) \rightarrow S : \{user\_id_C, h(CN_{\mathcal{S}}, P_{\mathcal{S}}, h(user\_id_C, Password_C, KS_{C\mathcal{I}}))\}_{KS_{\mathcal{I}S}}$

$M8'')\ \mathcal{I}(C) \rightarrow S : \{user\_id_C, h(CN_S, P_S, h(user\_id_C, Password_C, KS_{\mathcal{I}S}))\}_{KS_{\mathcal{I}S}}$

4.2 Off-Line and Replay Attack

The proposed scheme does not reveal a password owing to SSL/TLS encryption. In the scheme, since the password is actually hashed as an authentication key, it cannot be deprived by the attacker if SSL/TLS connection is compromised.

Moreover, each SSL/TLS session uses a different key to prevent the attacker from guessing a session key. Replay attack then cannot be worked against the scheme.

Therefore, the scheme is not vulnerable to these attacks.

4.3 Misleading Attack

This kind of attack seems to be prevented by a user's carefulness when connecting a web server prepared to deploy by the attacker. Since the attacker utilizes the legitimate certificate and there is a lot of web site, users cannot decide if it is attacker or not.

As shown in the section 4.1, a server in the scheme can detect the attacker even when using phishing site in same conditions. Therefore, the proposed scheme is one of the solutions for this kind of compromising.

5 Conclusion and Future Work

In this paper, we introduce the concept of authentication binding between SSL/TLS and HTTP authentication without modifying protocols, and implement our proposed scheme to show its effectiveness. It provides yet another method to establish secure channel for web application.

References

1. Frier, A., Karlton, P., Kocher, P.: The SSL 3.0 Protocol, Netscape Communications Corp., (November 18, 1996)
2. Dierks, T., Allen, C.: The TLS Protocol Version 1.0, RFC 2246 (January 1999)
3. Rescorla, E.: HTTP Over TLS, RFC 2818 (May 2000)
4. Housley, R., Ford, W., Polk, W., Solo, D.: Internet Public Key Infrastructure: Part I: X.509 Certificate and CRL Profile, RFC 2459 (January 1999)
5. Wu, T.: SRP-6: Improvements and Refinements to the Secure Remote Password Protocol (October 2002), http://srp.stanford.edu/srp6.ps
6. Wu, T.: The SRP Authentication and Key Exchange System, RFC 2945 (September 2000)
7. Franks, J., Hallam-Baker, P., Hostetler, J., Lawrence, S., Leach, P., Luotonen, A., Sink, E., Stewart, L.: HTTP Authentication: Basic and Digest Access Authentication, RFC 2617 (June 1999)
8. Eronen, P., Tschofenig, H.: Pre-Shared Key Ciphersuites for Transport Layer Security (TLS), RFC4279
9. http://sec.ietf.org/saag/13-November-2003/cbindings.pdf
10. http://www.mozilla.org/projects/security/pki/nss/
11. http://www.php.net/
12. http://www.modssl.org/

Embedding Legacy Keyword Search into Queries for the Ubiquitous ID Database

Tetsuo Kamina, Noboru Koshizuka, and Ken Sakamura

The University of Tokyo,
7-3-1, Hongo, Bunkyo-ku, Tokyo, 113-0033, Japan
{kamina,koshizuka,ken}@sakamura-lab.org

Abstract. Ubiquitous ID is a general purpose framework for implementing context-aware ubiquitous computing applications, where identifiers (called ucode numbers) and their relations are maintained in a large scale distributed database called UCRDB. Since we have to maintain a huge amount of data in UCRDB, it is sometimes desirable to delegate some subqueries to the "legacy" text-base search engines. In this paper, we propose a new query language construct for UCRDB (but can be applied to any other similar technologies such as RDF databases) that enables such a dynamic linking. Using this approach, context of the real world and legacy contents exist in the digital space can be seamlessly combined, and we can view UCRDB and legacy search engines as a single hybrid database so that no programming to "hard-wire" existing services is required. Furthermore, in our system, configurations of dynamic linking are described as a rule base stored in UCRDB itself, thus the resulting system is very simple but highly flexible and extensible.

1 Introduction

In a ubiquitous computing environment, it is important to enable computing devices to *identify* objects and places in the real world; one of the promising ways to satisfy this requirement is to assign an identifier to each object and place that we want to identify, and store this identifier into some digital formats (such as RFID, barcode, and so on) that computing devices can read. Furthermore, to realize the property of *context awareness* [2], we can define the relations among objects and places that describe the context of the real world by relating each identifier.

Ubiquitous ID [1] is a general purpose framework for implementing such context-aware ubiquitous computing applications. In this framework, each object and place is identified by a *ucode number*, a unique identifier that is actually a 128-bit length integer carrying no semantic information and thus it can be assigned to *everything*. Even logical entities such as relations of ucode numbers may be identified by ucode numbers. All the ucode numbers and their relations are stored in a large scale distributed database called UCRDB (UCode Relation database). These relations form a very large directed graph like RDF [15], where each node is a ucode number or a string literal (that is an attribute of

M. Takizawa, L. Barolli, and T. Enokido (Eds.): NBiS 2008, LNCS 5186, pp. 263–272, 2008.
© Springer-Verlag Berlin Heidelberg 2008

the ucode number), and each edge is a relation between a ucode number and another ucode number or a string literal. Queries to UCRDB is performed by adapting the technology of graph pattern matching. A query graph pattern is constructed by a mobile terminal acquiring ucode numbers from its surrounding environment and combining them with some contextual information (such as access histories); this pattern is then sent to the UCRDB and all the matched results are returned.

To implement many information services based on UCRDB, we have to prepare a huge amount of digital contents, which can easily be a very time-consuming task. On the other hand, there already exists a huge amount of information in World Wide Web, and its amount is explosively increasing. Therefore, to effectively reuse these contents, it is sometimes desirable to link UCRDB with services provided by the "legacy" text-base search engines (regardless to say that they serve documents on the Internet or a local document repository).

In this paper, we propose a new query language construct that enables such a dynamic linking. Even though this proposal is based on UCRDB, our approach can be applied to any similar database technologies such as RDF databases. Actually, our query language is implemented as a simple extension of SPARQL [16], a standard query language for RDF and thus details of the underlying implementation of database systems are out of scope of this paper. However, to clarify our motivations, in the rest of this paper we argue our approach based on UCRDB.

The proposed query language has a new feature of interfacing and querying external text-base search engines *on behalf of the query execution of UCRDB*. In general, the scope of a query language is closed inside its targeting database system, and its only way to communicate with external systems is using standard interfaces such as JDBC and ODBC, through which execution of query is *passively* invoked from general purpose host languages such as Java and C++. Since many query languages do not provide interfaces where they can *actively* communicate with the external systems, the aforementioned dynamic linking have had to be implemented in the level of host languages. Our approach is useful in that this dynamic linking is performed in the level of query languages.

In our approach, not only a linking between UCRDB and legacy contents but also which external search engines to be accessed is determined at run-time. To implement this feature, we use UCRDB itself as a key technology; using UCRDB, we define which external engines to be accessed and how to communicate with these engines as a rule base. Our query system understands this rule base and automatically delegates search request to the appropriate external engines. Thus, in our approach a query is very concise and simple but its behavior is flexibly configurable and extensible.

So far, our contributions can be summarized as follows:

– In our approach, context of the real world and legacy contents exist in the digital space are seamlessly combined so that application developers can easily make use of legacy contents in new context-aware ubiquitous information services.

– A simple query language enabling aforementioned dynamic linking is implemented as a simple extension of SPARQL.
– By using this query language, we can view UCRDB and legacy search engines as a single hybrid database and thus no programming at the level of host programing language is required.
– The resulting system is very simple but highly flexible and extensible.

2 Motivating Applications

2.1 Ubiquitous ID Framework

In this section, we introduce the core technologies of ubiquitous ID as a background. Ubiquitous ID is a general purpose framework that is designed to implement context-aware ubiquitous computing applications. To realize the property of context-awareness, in this framework every object, place, and even concept that we want to identify is assigned a unique identifier called a ucode number. A ucode number can be stored in many kinds of tags such as passive RFID tags, active RFID tags, infrared markers, barcodes, 2-D codes, and so on. How to embed ucode numbers to these tags and access protocols to these tags are standardized so that many kinds of computing devices can read ucode numbers attached to the real world.

The context of the real world is described by relating each ucode number. Each relation is also identified by using a ucode number, thus this relation is described as a triple of a subject ucode number, a relation ucode number (i.e. predicate), and an object ucode number, which is exactly the same format of RDF [15] triples except that each element of a triple is not an URI but a ucode number[1]. Actually, each ucode number can be represented as an URI by using a namespace; furthermore, in UCRDB we may assign an *alias name* (in the form of URI) for each ucode number. Therefore, we can query over UCRDB by using SPARQL, a standard query language for RDF databases[2].

2.2 Applications

Ubiquitous ID is a general purpose framework in that each ucode number does not carry any semantic information, and UCRDB is semi-structured so that we can freely define application specific schemata. Since we have to maintain a huge amount of data in UCRDB, when we implement an information service using UCRDB it is sometimes desirable to delegate some subqueries to the legacy text-base search engines. In the following subsections, we describe such scenarios.

[1] As in RDF, the object part of a triple may also be a string literal. Note that this string literal may be a URI string. For example, we may relate a ucode number with a URL that stores further information of the ucode number.
[2] In our implementation, each alias name is reduced to the corresponding ucode number in the preprocessing phase of SPARQL.

Site-specific Information Systems. By using ucode numbers, we can identify places or sites. Furthermore, by adapting this identification technique, we can also define the relations among places such as "that is the 4th building past the intersection," "these two intersections are connected," and so on. By maintaining these relations in UCRDB, we can construct a pedestrian navigation system [4]; in this system, an active RFID tag announces the ucode number of a place, and the mobile terminal reads the ucode number and queries the route information to the UCRDB. This mechanism enables more fine-grained pedestrian navigations than that are built using GPS technologies. Another advantage of this approach is we can provide a pedestrian navigation for interior regions such as museums and shopping malls, where GPS technologies cannot be applied.

In a pedestrian navigation system, it is also convenient to link the navigation system with other information services; for example, some users may want to be notified the nearest shop information or sightseeing information from where they are. Since creating such contents from scratch is a very time-consuming task, such information should be retrieved from the Web search engines using the name of the place (and other auxiliary information) as keywords. Therefore, some linking mechanism between UCRDB and Web search engines will be useful.

Equipment Management. Preserving digital archives of records of the equipments in the real world is a persistent requirement. By applying the Ubiquitous ID framework, we can also construct such an equipment management system. In this system, an RFID tag containing a ucode number is attached to each equipment, and we can get the records of equipments on site using a mobile terminal that reads the RFID tags and queries the database storing the information.

Such equipment management database can be constructed by relating each ucode number to their information. However, there are huge amount of records of equipments, thus constructing a database of such records is a very time-consuming task. On the other hand, to construct a digital archive, we may also use a full-text search engine. In this case, we do not have to construct a database; we just index each document stored in the file system so that each document is searched using keywords. To make use of such a full-text search engine from UCRDB, we have to implement some communication mechanisms between them.

3 Design and Implementation

In the aforementioned applications, we *conceptually* view the UCRDB and text-base search engines serving World Wide Web or file systems as a single hybrid database. We introduce a new feature of interfacing and querying external engines *on behalf of the query execution* into the SPARQL query language. In this extension, we impose the following requirements:

Extensibility: There are many kinds of text-base search engines. The number of search engines is still growing, and each search engine may evolve so that its interface changes. Thus, our extension should be able to incorporate with new kinds of external engines.

```
@prefix ex: ... .
ex:pointA ex:latitude "..." .
ex:pointA ex:longitude "..." .
ex:linkAB ex:node ex:pointA .
ex:linkAB ex:node ex:pointB .
ex:linkAB ex:length "34.5" .
ex:pointA ex:nearestShop ex:abcDepartmentStore .
ex:pointA ex:nearestShop ex:xyzMusic .
ex:abcDepartmentStore ex:name "ABC Department Store" .
ex:xyzMusic ex:name "XYZ Music" .
```

Fig. 1. A simple RDF data for pedestrian navigation system

Lightweight Language: SPARQL is a very simple query language and thus adding unnecessary complexity is undesirable. Therefore, the new language constructs that are added to SPARQL should be very simple.

To achieve these requirements, our new extension of SPARQL has the following features:

Defining the Mapping to External Engines in UCRDB: In our extension, how queries are executed on each external engine is user-definable using UCRDB. To implement this feature, we store attributes of external engines those are written as UCRDB relations mapping each external server's ucode number to their attribute values. Each of these relations is assigned a ucode number and also given an alias name in the form of URI for human readability.

Developing the Interpreter: We introduce a new pattern `SEARCH EXTERNAL` into SPARQL. A `SEARCH EXTERNAL` pattern interprets the relations mapping each external server to its attribute values and sends a query to the appropriate external engine. This pattern may contain variables appear in other triple patterns.

In the following sections, we assume that every ucode number appears in our example is assigned an alias name. Thus, in the following examples every triple of UCRDB appears in the form of an RDF triple. For simplicity, we also use a convention to write a URI by using a namespace prefix such as `ex:`.

An Example. To explain the feature of our proposal, we use a simple example that describes UCRDB data of a pedestrian navigation system. Fig.1 shows a piece of the UCRDB database in the form of N3 [3]. The values `ex:pointA` and `ex:pointB` are points in the pedestrian network that we want to identify to represent routes (e.g. intersections); these points are connected by the link `ex:linkAB` whose length is 34.5m. Besides the route for the goal, this pedestrian navigation system also shows some useful information around the point, if necessary. For example, the descriptions in Fig.1 show that `ex:pointA` is close to a department store and a music shop.

Consider that a pedestrian reads a ucode of `ex:pointA` using a mobile terminal. The following query acquires the name of shops close to the variable

```
extern:myServer exvocab:url "http://search.yahoo.ac.jp/search" .
extern:myServer exvocab:query_param "p" .
... (other specifications for HTTP request parameters)
```

Fig. 2. A specification of an external engine written in RDF

`?this` (whose value is implicitly given as `ex:pointA`) and sends a request to the external search engine by using the acquired shop name as a keyword:

```
SELECT ?y ?ext WHERE {
  ?this ex:nearestShop ?x .
  ?x ex:name ?y .
  SEARCH EXTERNAL ?ext { ?y -> extern:myServer } }
```
(1)

The `SEARCH EXTERNAL` pattern is the new syntax we add to SPARQL. This pattern may appear in any places where a SPARQL pattern (such as a triple, a `FILTER` pattern, and so on) may appear, but inside the block of this pattern has a special syntax; we may not write any SPARQL patterns there. Instead, we put a search request descriptor (just a variable `?y` in the above example) and a search engine descriptor (a resource `extern:myServer` in the above example), which means that the search request `?y` is sent to the resource `extern:myServer`. The response of this request (a list of pair of found resource's URL and title) is stored in the variable `?ext`[3]. As in the original SPARQL, the result of this query is a list of tuples whose columns are indexed as `?y` and `?ext`, respectively. A request is sent to `extern:myServer` for each matched value with `?y`[4].

We can use multiple variables in a search request descriptor by combining them by logical operators such as `&&` (and) and `||` (or). We can also restrict the search results by specifying filters such as the content's language (written `lang="jp"`, for example), maximal counts of the results (written `max=10`, for example), and so on. Note that filters can only be combined using the `&&` operator.

Behavior Mapping in UCRDB. In the above example, how the query keywords are sent to the external engine is not specified. In our system, this mapping is not hard-wired, to support a dozen of external engines and to incorporate with a new kind of external engines that is not taken into account at the first time of development. For this purpose, we introduce a new UCRDB vocabulary (i.e., a set of ucode numbers assigned to relations specifying attributes of external engines). This vocabulary is used to map a resource identifying an external engine (`extern:myServer` in the above example) to its attributes such as its base URL, a HTTP request parameter name that captures query keywords, and so on.

[3] Note that the variable `?ext` does not match any of subject, predicate, and object of triples, since it does not have a type of URI or string literal.

[4] However, we may restrict the total number of requests to the same external engine in one query, whose default value is 100.

For example, a specification of the external engine `extern:myServer` can be written as shown in Fig. 2. It shows that the base URL of the location of `extern:myServer` is `http://search.yahoo.co.jp/search`, and the HTTP request parameter name used for queries to this search engine is `p`. Thus, the query `(1)` results in the following sequence of HTTP requests, when the variable `?this` matches `ex:pointA`:

```
http://search.yahoo.co.jp/search?p=%22ABC+Department+Store%22
http://search.yahoo.co.jp/search?p=%22XYZ+Music%22
```

We can also specify other external engines in the similar way. Furthermore, we can change the specification of `extern:myServer` to make use of the other external engine without modifying the query. Our query engine understands the UCRDB vocabularies and delegates queries for the appropriate external engine.

Note that the format of results from each external engine (in general written in HTML) also differs from other external engines. This means that how to extract the list of pairs of found document's URL and title from the results of each external engine is also differs from that of other external engines. Therefore, we also have to specify the rules for such extraction using UCRDB vocabularies. For example, which anchor tag contains URL of the query result can be determined by observing its `class` attributes' value, its enclosing tag and attributes, depth from the root `html` tag, and so on.

Implementation. A data flow of our query processor is shown in Fig.3. Firstly, the extended query processor parses the query and constructs a SPARQL statement from which `SEARCH EXTERNAL` patterns are removed. Then, it passes this statement to the original SPARQL processor. The SPARQL processor executes the SPARQL statement and returns the results to the extended query processor. By inspecting the results, the extended query processor then assigns values to the variables appear in `SEARCH EXTERNAL` patterns to construct a query for each external engine. This process iterates over all the combinations of variable

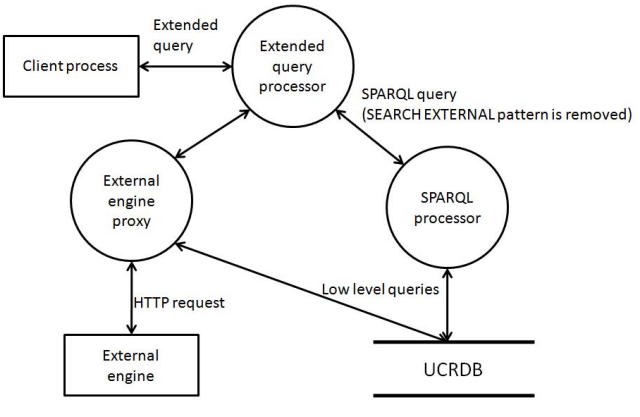


Fig. 3. Data flow of our query processor

```
String query = ...; /* SPARQL statement shown in section 3.1 */
DatabaseResult rs = ucode.execSparqlQuery(query);
while (rs.hasNext()) {
  ExternalResultList lst = (ExternalResultList)rs.get("?ext");
  while (lst.hasNext()) {
    ExternalResult ext = lst.next();
    String url = ext.getURL();
    String title = ext.getTitle(); ... } }
```

Fig. 4. An example code using our API

assignments. The external engine proxy constructs HTTP request parameters for the external engine whose location is acquired by inspecting the UCRDB, and sends them to the external engine. The response from the external engine is also inspected by looking up the UCRDB rule base to extract the search results (i.e. found document's URLs and titles). Finally, the extended query processor merges the results from the original SPARQL processor and external engines, and returns them to the client.

Our current implementation is based on Java API for UCRDB. This API is object-oriented, in that every query is executed via message passing. A receiver of messages is a ucode object; besides simple lookup methods such as "getting all the triples whose subject is the receiver ucode," a SPARQL execution method is also implemented as a member of `Ucode` class. The method signature of `execSparqlQuery` is as follows:

```
DatabaseResult Ucode.execSparqlQuery(String query);
```

The formal parameter `query` is an (extended) SPARQL statement. This statement may include a special reserved variable `?this`, which refers to the ucode number of its receiver.

A result of SPARQL statement is a list of tuples, whose columns are indexed by the names of variables. The class `DatabaseResult` encapsulates the internal structure of tuples and provides access methods. In general, a value of each column has a type of ucode or string literal (or blank node), but some values have a type of "lists of pairs of found document's URL and title." Consider the code shown in Fig.4. In this code, the SPARQL statement `(1)` is executed, and its result is assigned to the variable `rs`. For each row of `rs`, we get a value of column indexed as `?ext`. Since this value is a result of `SEARCH EXTERNAL`, it contains a list of found document's URL and title. The types of `?ext` and its elements are represented as classes `ExternalResultList` and `ExternalResult`, respectively.

So far, the aforementioned requirements are met; owing to UCRDB's ability of self-description, we may set each parameter describing external engines in UCRDB itself, thus the resulting system is highly flexible and extensible. The only language construct we add to SPARQL is `SEARCH EXTERNAL` pattern, whose semantics is also straightforward. The only burden we add to the underlying SPARQL implementations is the process time of external engine proxy, which includes response time of HTTP request that likely be a bottleneck. This

response time varies depending on the external engines and network conditions, but in most of the cases it should be acceptable.

4 Related Work

Extensive research efforts have been made for exploiting how to combine classical search techniques with semantic model described as metadata or ontology [6,7,11,12]. For example, Rocha et al. [11] show an approach of using semantic model of a given domain to calculate weights of links that measure the strength of the relation. Spread activation techniques are used to find related concepts in the ontology given an initial set of concepts and corresponding initial activation values, which are obtained from the results of classical search. In general, these approaches are useful when there is rich metadata associated with web pages.

Our approach, on the other hand, aims to combine classical search techniques with semantic search, regardless to say that there are metadata associated with web pages or not. In this sense, our approach is more similar with business process execution languages such as BPEL4WS [5], which is a result of merging previously developed WSFL [9] and XLANG [13]. In these approaches, however, the description of service partners is done via WSDL portType definitions [14], which prevents solely matching of WSDL messaging interfaces[5]. Furthermore, semantically the same services cannot be combined unless the import and export interfaces are exactly matched. Therefore, these approaches do not provide a flexible mechanism to combine each service as presented in our approach.

Mandell and McIlraith [10] propose a bottom-up approach to integrate Semantic Web technologies into Web services. Based on BPEL4WS, they present integrated Semantic Web technology for automating customized, dynamic binding of Web services together with interoperation through semantic translation. Our approach, on the other hand, is based on SPARQL and provides much closer looking at database queries. In our approach, combination of UCRDB and legacy search engines is performed in a declarative way.

5 Concluding Remarks

This paper presents a new query language construct that enables dynamic linking of UCRDB and legacy text-base search engines. This feature is designed and implemented at the top of UCRDB, but the same approach may be applicable to any other similar technologies such as RDF databases. Using this approach, context of the real world and legacy contents exist in the digital space can be seamlessly combined, and we can view UCRDB and legacy search engines as a single hybrid database so that no programming to hard-wire existing services are required. Furthermore, in our system, configurations of dynamic linking are described as a rule base stored in UCRDB itself, thus the resulting system is very simple but highly flexible and extensible.

[5] Kamina and Tamai [8] propose a method of structural matching of WSDL messaging interfaces.

References

1. Ubiquitous ID Center, http://www.uidcenter.org/
2. Abowd, G.D., Dey, A.K., Brown, P.J., Davies, N., Smith, M., Steggles, P.: Towards a better understanding of context and context-awareness. In: Gellersen, H.-W. (ed.) HUC 1999. LNCS, vol. 1707, pp. 304–307. Springer, Heidelberg (1999)
3. Tim Berners-Lee. Notation 3 (1998), http://www.w3.org/DesignIssues/Notation3.html
4. Bessho, M., Kobayashi, S., Koshizuka, N., Sakamura, K.: A space-identifying ubiquitous infrastructure and its application for tour-guiding service. In: SAC 2008, pp. 1616–1622 (2008)
5. Curbera, F., Goland, Y., Klein, J., Leymann, F., Roller, D., Thatte, S., Weerawarana, S.: Business process execution language for web services, http://www.ibm.com/developerworks/library/ws-bpel/
6. Davies, J., Weeks, R.: QuizRDF: Search technology for the Semantic Web. In: Proceedings of the 37th Hawaii International Conference on System Sciences, page 40112 (2004)
7. Ding, L., Finin, T., Joshi, A., Pan, R., S̃cott Cost, R., Peng, Y., Reddivari, P., Doshi, V., Sachs, J.: Swoogle: A search and metadata engine for the Semantic Web. In: CIKM 2004, pp. 652–659 (2004)
8. Kamina, T., Tamai, T.: Loosely Connected RPC: An approach for extendable interface of Web Services. In: Proceedings of the 1st International Workshop on Web Services: Modeling, Architecture and Infrastructure (WSMAI 2003), pp. 62–73 (2003)
9. Leymann, F.: Web services flow language, http://www-3.ibm.com/software/solutions/webservices/pdf/WSFL.pdf
10. Mandell, D.J., Mcllaith, S.A.: Adapting BPEL4WS for the Semantic Web: The bottom-up approach to web service interoperation. In: Fensel, D., Sycara, K.P., Mylopoulos, J. (eds.) ISWC 2003. LNCS, vol. 2870, pp. 227–241. Springer, Heidelberg (2003)
11. Rocha, C., Schwabe, D., de Aragao, M.P.: A hybrid approach for searching in the semantic web. In: WWW 2004, pp. 374–383 (2004)
12. Stojanovic, L., Stojanovic, N., Volz, R.: Migrating data-intensive Web Sites into the Semantic Web. In: ACM SAC 2002, pp. 1100–1107 (2002)
13. Thatte, S.: XLANG: Web services for business process design, http://www.gotdotnet.com/team/xml_wsspecs/xlang-c/default.html
14. W3C. Web Services Description Language (WSDL) 1.1 (2001), http://www.w3.org/TR/2001/NOTE-wsdl-20010315
15. W3C. Resource Description Framework (RDF): Concepts and Abstract Syntax (2004), http://www.w3.org/TR/2004/REC-rdf-concepts-20040210/
16. W3C. SPARQL Query Language for RDF (2008), http://www.w3.org/TR/2008/REC-rdf-sparql-query-20080115/

Secure Ubiquitous Health Monitoring System

Arjan Durresi[1], Mimoza Durresi[1], and Leonard Barolli[2]

[1] Indiana University Purdue University Indianapolis, IN 46202, USA
durresi@cs.iupui.edu
http://www.cs.iupui.edu/~durresi
[2] Fukuoka Institute of Technology, Fukuoka, Japan

Abstract. We propose a distributed and secure system that enables global & ubiquitous health monitoring. The biomedical data, collected by wearable sensors will be transmitted using cell phones towards the corresponding Health Monitoring Centers via various wireless networks. The security of personal medical data is paramount. Therefore, we propose a new scheme that enable security communications among cell phones and other wireless networks. In particular, we propose a hybrid authentication scheme that involves both the cellular network and the Health Monitor Centers. Therefore, we take advantage of the strong authentication and non-repudiation provided by the cellular networks. Consequently, our solution guaranties confidentiality, privacy and defences against denial of service attacks.

1 Introduction

Information technology can be instrumental in supporting health care services, including increasing the quality and containing the costs of such services. The cost of the health care is a growing problem, for example, expenditures in the United States for healthcare will grow to 15.9% of the GDP ($2.6 trillion) by 2010 (Digital 4Sight's Healthcare Industry Study) as a result of the accumulative impact of chronic degenerative diseases in the elderly and their increasing dependence on the health care system.

There are several research projects in the field of health care services. *I-Living* [1,2], an assisted-living supportive system, beng developed by researchers at the University of Illinois at Urbana-Champaign. In [3] are used infrared sensors, computers, bio-sensors, and video cameras. The *Aware Home* project at Georgia Tech [4] targets to create a home environment that is aware of its occupants' whereabouts and activities. The smart in-home monitoring system at University of Virginia [5] focuses on data collection with the use of a suite of low-cost, non-intrusive sensors. The major industry research effort is perhaps led by the age-in-place advanced smart-home system at Intel [6].

We have proposed in [7] an architecture for a distributed ubiquitous health monitor system. Differently from the above mentioned strategies, our goal [7] is to provide ubiquitous health monitoring, at home, and outside it, all the time. People need to have their health conditions under control not only when they are at home, but everywhere.

M. Takizawa, L. Barolli, and T. Enokido (Eds.): NBiS 2008, LNCS 5186, pp. 273–282, 2008.
© Springer-Verlag Berlin Heidelberg 2008

In this paper we expand the security aspect of our health monitoring system. In particular we develop a new scheme that will enable secure communications among the patients' cell phone, which collect the sensed data, and the various wireless networks.

The rest of the paper is organized as follows. Section 2 provides an overview of our system architecture. In Section 3 we discuss the communication architecture. Section 4 presents the related work. In Section 5 we discuss the elements of our scheme. We present our security scheme in 6 and we discuss its security aspects in Section 7. We conclude in Section 8.

2 Architecture Overview

In Fig. 1 we show an overview of the architecture that we are proposing for the Ubiquitous Health Monitoring System.

The biomedical data will be generated by multiple sensors that can be integrated in devices wearable by patients, such as necklaces, bracelets, etc. Sensors could include Carotid Dopplers, ECG Nanotrodes, biochip based on tissue/fluids Quantum Dots, etc. The measured biomedical data could include vital signs, such as the glucose level, blood pressure, heart bit rate, arterial oxyhemoglobin saturation level, etc.

The measured biomedical data will be transmitted via multiple complementary wireless networks, through the Internet, towards the appropriate Health Monitoring Center (HMC), where this data will be integrated with the permanent medical data of the given patient. Therefore, the medical personnel at HMC will be able to monitor various vital signs at desirable time granularity. Should

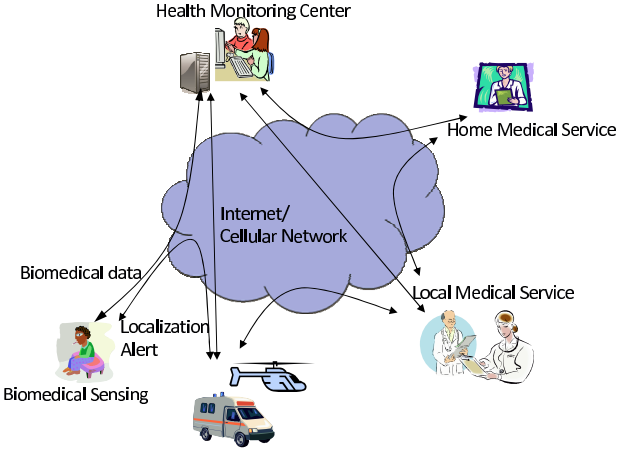


Fig. 1. System Architecture. Patient's biomedical data will be sent via multiple wireless networks to the appropriate Health Monitor Center, which will coordinate the activities among the Home and Local Medical Services and various emergency services.

the readings suggest any abnormal health situations, medical instructions can be given and actions can be taken before the situations deteriorate. More details about the proposed architecture can be found in [7].

3 Mobile Health Monitoring

When patients are away from their homes, again we propose cellular phone to be used as the first choice for the access network.

The reason to select cell phones as a collecting device for personal medical data is that cell phones are already ubiquitous devices. So, there are over 2 billion cell phones worldwide [8]. The features and resources available to cell phones has been increasing at a staggering pace [9,10]. Many of the current cell phones come with advanced operating systems. External memory cards are available which takes the storage capacity of these phones into the order of Gigabytes. If the average communication message is 1KB, a memory of 1GB would allow more than million such messages. The life of the batteries being used in these phones has also increased tremendously. New cell phones, called *dual phones* [9,10], are able to communicate over cellular networks and WLANs, such as IEEE 802.11 family. To increase the reliability of communications, cell phones can also communicate as ad hoc nodes.

4 Related Work

We have proposed a protocol which can be used to establish non-refutable communication for cellular phones in the ad hoc mode [11]. These networks have the special property unlike other ad hoc networks, which is that the nodes of the network are part of the centralized network before they enter the ad hoc mode. This allows the cellular base station to send all the key information securely using the secure cellular infrastructure. We achieve secure non-repudiation in these networks by the use of Identity based cryptography.

In 2-2.5 G cell phone systems such as GSM, each cell phone has a Universal Integrated Circuit Card (UICC), also referred to as SIM card [12,13]. It acts as the ID of the phone and stores all the information related to the working of the phone. To authenticate a mobile phone the cellular base station sends a 128-bit random challenge ($RAND$). Each phone has a 128-bit Individual Subscriber Authentication Key ($ISAI$) stored in the SIM. Using the $RAND$ and $ISAI$ the mobile phone generates a response to the challenge posed by the cellular base station along with the session key for the communication. This session key is used to encrypt the communication between the mobile station and the cellular base station. In this paper we propose to assign a new ID and the corresponding private key to the mobile phone when it is authenticated.

Identity based cryptography was first proposed by Shamir in 1984 [14]. In this paradigm, the users' ID, like phone number or email ID, has a one-on-one mapping with the public key of the user. This reduces the system complexity and cost for establishing and managing public key infrastructure [15,16]. The Private Key Generator (PKG) creates its master and public key.

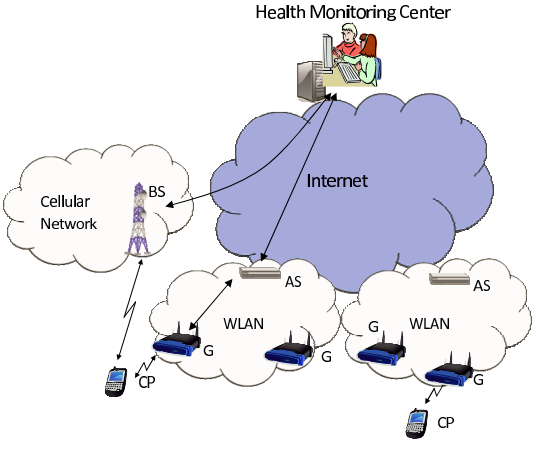


Fig. 2. Our network architecture. HMC: Health Monitoring Center, AS: Authentication Server, BS: Cellular Base Station, CP: Cell Phone.

Eschenauer and Gligor proposed the random key predistribution scheme [17], also known as the basic scheme. Many variations of this scheme have been proposed [18,19,20,21,22,23,24]. We have presented a key predistribution based approach that allows a mobile sensor node to communicate with the stationary networks of multiple stationary networks [25].

Several existing research papers deal with with intrusion detection systems in mobile ad hoc networks [26]. The possibility of deploying the mobile nodes in inaccessible regions and difficult terrain makes node capture a suitable approach to attack the network. This means that intrusion prevention techniques like encryption and authentication can not guarantee security. The primary assumption of an intrusion detection system are that the behavior of a user can be observed. It is also assumed that the normal behavior and the abnormal behavior of a system can be differentiated. A intrusion detection system collects secure audit data about the system and determines if the system is under attack and points to the compromised parts of the system. Two popular categories on intrusion detection systems are anomaly detection and misuse detection. Anomaly detection deals with the uncovering of abnormal patterns in the behavior of a system. Misuse detection on the other hand makes use of specific attack patterns to detect attacks. We believe that the interaction of the cellular infrastructure with sensor networks may be used to constantly transmit these attack signatures to all the nodes making the implementation of intrusion detection systems possible.

5 System Model

In this section we present the proposed architecture for connecting cell phones to wireless networks securely, as shown in Fig. 2. Wireless networks could consist of various technologies, including the IEEE 802.11 family.

5.1 System Components

We list the different components in our proposed architecture.

Cell Phones: Cell phones are popular hand held devices with the functionality of the traditional phone. These phones connect to a cellular infrastructure of base stations to communicate with each other. We present schemes to allow these cell phones to connect to multiple WLANs. The cellular network is used to authenticate the cell phone and provide the key information using the secure channels of cellular communication.The use of cell phones as document scanners and input devices has been proposed [27].

Gateways: These nodes act as interfaces between the WLAN and the cell phones. The gateways interact with the cell phones to authenticate them and collect from them the information sensed by the wearable patient's sensors. A WLAN may use multiple gateways. These gateways are connected to the Authentication Server which provides them with the key information.

Health Monitor Center (HMC): This is the main server used to establish security in the communication links between the cell phones and various WLANs. The HMC provides the cell phone relevant key information using the cellular infrastructure. The HMC also interacts with all the Authentication Servers (AS) to provide and refresh the key information in the multiple WLANs and gateways (refer to Section 5.1). For this scheme we assume that the HMC is secure and not vulnerable to capture. To HMC is distributed over the territory, to increase the reliability and the scalability of the system.

Authentication Server (AS): There is one authentication server per WLAN. The AS is in charge of all the security management in the WLAN. It has a secure link with all the gateways of WLAN. This link is compromised only when the gateway is captured. The AS provides the gateways with the keys required to securely communicate with the cell phones. Our architecture assumes the AS to be secure.

5.2 Security Architecture

We present a security architecture using the different components described in Section 5.1. The gateways different WLANs use different sets of keys because in that case security is increased as the compromise of one network does not lead to the compromise of all the networks.

When the cell phone is within the communication range of the gateways of a WLAN, through which it wishes to report the medical data, it sends a request to the HMC through the cellular network. The HMC then provides the key information to the cell phone, which it can use to communicate with the given WLAN. The HMC then communicates with the AS in the WALNA and provides it the key information which can be used by the gateways to communicate with the particular cell phone. The AS then sends the relevant information to the

gateways and the communication is achieved. The AS can either send this key information to all the gateways or provide a on demand scheme. If the key information is on demand, the overhead is reduced because the key information would have to be sent to only one gateway. On the other hand giving the key information to all the gateways is less secure because even if one of the gateways is compromised, the key would be compromised. A tradeoff between the two approaches can be drawn where the AS gives the key information to the gateways which are likely to be in the neighborhood of the cell phone.

6 Our Schemes

In this section we present our scheme that enable secure communication among the cell phones to multiple WLANs. The proposed scheme make use of asymmetric key cryptography between the cell phones and the gateways.

We first present the different steps which are required for the cell phone to establish a secure session key with the cellular base station. The Cellular Base Station (BS) sends a random challenge (RAND) to the Cell Phone (CP). The function f_1 implements the $A3$ algorithm which generates the 32-bit signed response using the 128-bit $RAND$ and the 128-bit $ISAI$ (refer to Section 4). This response is used to authenticate the mobile phone to the cellular base station. The cell phone then generates the session key K_s using function f_2, which implements the $A8$ algorithm. After this stage the cellular base station can establish a secure communication with the mobile phone using the session key K_s [12,13].

$$BS \rightarrow CP : RAND \tag{1}$$

$$CP \rightarrow BS : RESP = f_1\,(RAND, ISAI) \tag{2}$$

$$CP : K_s = f_2\,(RAND, ISAI) \tag{3}$$

The key K_s is used to establish secure communication between the cell phones and the Cellular Base Station.

We assume that the cell phone and the Health Monitoring Center already have established end-to-end secure communications using asymmetric encryption. When a cell phone (CP) wishes to communicate with a gateway of a given WLAN, it sends a request to the Health Monitor Center (refer to Section 5.1). This request is sent to the HMC through the cellular network.

The HMC responds to this message by sending to the cell phone: the public key of the given gateway of that WLAN (this information is provided to HMC by the respective AS), the temporary ID_C and the corresponding private key SK_{ID_C} for the cell phone to be used in its communication with the given WLAN. On the other hand the HMC sends to AS the pair: ID_C, public key SK_{ID_C} of the given cell phone.

Let ID_G and ID_N be the ID of the gateway and the given WLAN; and ID_C the ID of the cell phone. Let K_s be the session key that the cell phone shares with the cellular base station.

$$CP \rightarrow BS : K_s\,[ID_G, ID_N] \tag{4}$$

The BS then passes this request to the HMC that responds with the public key of the given gateway (PK_{PKG}) to the cell phone via the cellular network. HMC sends also via secure communications to the respective AS the ID and public key of the given cell phone.

$$BS \rightarrow CP : K_s \left[ID_G, ID_N, ID_C, SK_{ID_C}, PK_{PKG}\right] \tag{5}$$

The cell phone can use his ID_C and SK_{ID_C} to communicate securely with the WLAN. To send a message to a gateway G, the cell phone encrypts the data with the public key of G and signs it with its own private key.

$$CP \rightarrow G : ID_C, SK_{ID_C} \left[PK_G \left[Data, ID_G, ID_N\right]\right] \tag{6}$$

The gateway acknowledges the data by using the public key of the cell phone and then signing it with its private key.

$$G \rightarrow CP : ID_G, SK_G \left[PK_{ID_C} \left[Reply, ID_C, ID_N\right]\right] \tag{7}$$

The AS assigns an ID and private key to each gateway. This key information is transmitted to the gateways using a secure communication channel. This paper assumes that information can be communicated between the AS and gateways securely. This information is available to the attackers only when they capture the particular gateway.

Using the cellular network for communications authenticates the cell phone. To lie about its ID the cell phone would have to fake the number with which it connects to the cellular network. Therefore, the scheme benefits from the strong authentication of the cell phones.

The cell phone and the HMC use end to end asymmetric encryption to guaranty the security of data in case gateways are captured.

7 Security Discussion

Node capture is a major attack on this architecture and security protocols. In node capture, the attacker can physically capture a network component and compromise all the information stored in it. In the architecture that has been proposed in this paper, we have three components which are vulnerable to capture. These components are cell phones, gateways and sensor nodes.

We discuss the capture of nodes, possible mechanisms of detection of capture and methods to revoke the nodes which are captured.

7.1 Capture of Cell Phones

Every person can buy a cell phone which makes the cell phone easy to capture. The information stored in the phone belongs to the owner of the cell phone. The two security schemes presented in this paper ensure that the information stored in a cell phone can not be used to fake the communication from another cell

phone. In our scheme, the cell phone has the private key corresponding to the ID assigned to it by the HMC.

In case of a cell phone being lost or stolen, it can be captured by a malicious person. In such a situation, the user of a cell phone can have the base station of the cellular network revoke the phone. If the base station of the cellular network revokes a cell phone, it will be unable to communicate with the HMC to obtain the keys and hence the phone would be revoked.

7.2 Capture of Gateways

The gateways are nodes which are deployed as interfaces between the cell phones and sensor networks. These nodes are deployed in the open and may be captured by the attackers. Since the ID used by the cell phones to communicate with the sensor nodes is not related to the real phone number of the cell phone, we assume that the captured gateways would not be able to target any particular cell phones. Popular techniques like [26,28] can be used to detect intrusions into gateways. We believe that the availability of secure AS would be helpful in testing if any of the gateways have been captured. This detection can be done by sending online queries through other gateways which may pretend to be cell phones. The response of the gateway to these queries may be used to determine if it is captured or not. Other intrusion detection schemes may also use the AS to improve their performance.

7.3 Denial of Service

Another type of attack cold be aim at wireless resources in a WLAN and therefore, deny the service (connectivity) to legitimate patients, who need to use their cell phones to send their medical data via the WLAN.

A denial of service attack would be launched by captured cell phone that already are registered with HMC. However, it is expected the such lost cell phones will be reported to the cellular service provider and the HMC.

7.4 Privacy

Privacy is also a strong requirement for systems dealing with personal medical data. In our system, we guaranty privacy by using temporary ID for the cell phone when they report medical data though WLANs. Only HMC will know the relation between temporary cell ID and cell numbers.

8 Conclusion

We propose a secure architecture for a ubiquitous distributed Health Monitoring System. Our architecture is based on multiple complementary wireless communication access networks between the patient and the system, through the Internet and cellular networks.

The biomedical data will be generated continually by wearable sensing devices. Then the data will be collected by personal cell phones, which can use both cellular networks or other wireless networks, such as WLANs. In the later case, we propose a hybrid authentication scheme that involves both the cellular network and the Health Monitor Center. Therefore, we take advantage of the strong authentication and non-repudiation provided by the cellular networks. Consequently, our solution guaranties confidentiality, privacy and defences against denial of service attacks.

References

1. Assited Living Project Web page at University of Illinois at Urbana-Champaign, http://lion.cs.uiuc.edu/assistedliving/index.html
2. Wang, Q., Shin, W., Liu, X., Zeng, Z., Oh, C., Li, B.A., Caccamo, M., Gunter, C., Gunter, E., Hou, J., Karahalios, K., Sha, L.: I-Living: An Open System Architecture for Assisted Living. In: Proceedings of IEEE SMC (2006), http://lion.cs.uiuc.edu/assistedliving/publications/I-Living.pdf
3. University of Rochester, Center of Future Health, http://www.futurehealth.rochester.edu/news/
4. Georgia Institute of Technology, Aware Home, http://www.cc.gatech.edu/fce/ahri/
5. University of Virginia, Smart In-Home Monitoring System, http://marc.med.virginia.edu/projectssmarthomemonitor.html
6. Intel Corporation, Age-in-Place, http://www.intel.com/research/prohealth/cs-aginginplace.htm
7. Durresi, A., Durresi, M., Barolli, L.: Integrated Biomedical System for ubiquitous Health Monitoring. In: Enokido, T., Barolli, L., Takizawa, M. (eds.) NBiS 2007. LNCS, vol. 4658, pp. 397–405. Springer, Heidelberg (2007)
8. Cellularonline, http://www.cellular.co.za
9. http://www.networkworld.com/news/2004/072604avaya.html
10. http://www.pcworld.com/news/article/0,aid,116334,00.asp
11. Durresi, A., Bulusu, V., Paruchuri, V.: Security in ad hoc networks on cellular phones for emergency situations. Ad Hoc Networks Journal 5(1), 126–133 (2007)
12. Lo, C., Chen, Y.: Secure communication mechanisms for gsm networks. IEEE Transaction on Consumer Electronics 45(4), 1074–1080 (1999)
13. Mehrotra, A., Golding, L.: Mobility and security management in the gsm system and some proposed future improvements. Proceedings of the IEEE 86(7), 1480–1497 (1998)
14. Shamir, A.: Identity-based cryptosystems and signature schemes. In: Blakely, G.R., Chaum, D. (eds.) CRYPTO 1984. LNCS, vol. 196, pp. 47–53. Springer, Heidelberg (1985)
15. Boneh, D., Lynn, B., Shacham, H.: Short signatures from the weil pairing. J. Cryptol. 17(4), 297–319 (2004)
16. Horwitz, J., Lynn, B.: Toward hierarchical identity-based encryption. In: Knudsen, L.R. (ed.) EUROCRYPT 2002. LNCS, vol. 2332, pp. 466–481. Springer, Heidelberg (2002)
17. Eschenauer, L., Gligor, V.D.: A Key-Management Scheme for Distributed Sensor Networks. In: Proceedings of the 9th ACM conference on Computers and communications security, November 18-22, 2002, pp. 41–47 (2002)

18. Spencer, J.: The Strange Logic of Random Graphs ISBN: 3-540-41654-4, August 9, 2001. Springer, Heidelberg (2001)
19. Chan, H., Perrig, A., Song, D.: Key Distribution Techniques for Sensor Networks. Wireless Sensor Networks, 277–303 ISBN:1-4020-7883-8
20. Chan, H., Perrig, A., Song, D.: Random Key Predistribution Schemes for Sensor Networks. In: IEEE Symposium on Security and Privacy, May 11–14, 2003, pp. 197–213 (2003)
21. Du, W., Deng, J., Han, Y.S., Chen, S., Varshney, P.K.: A Key Management Scheme for Wireless Sensor Networks using Deployment Knowledge. In: Proceedings of the IEEE INFOCOM 2004, March 7-11,2004, pp. 586–597 (2004)
22. Liu, D., Ning, P.: Location-Based Pairwise Key Establishments for Static Sensor Networks. In: ACM workshop on Security in Ad Hoc and Sensor Networks (2003)
23. Liu, D., Ning, P., Li, R.: Establishing Pairwise Keys in Distributed Sensor Networks. In: 10th ACM conference on Computers and Communication Security (CCS 2003), October, pp. 52–61 (2003)
24. Zhu, S., Setia, S., Jajodia, S.: LEAP: Efficient Security Mechanisms for Large-Scale Distributed Sensor Networks. In: 10th ACM conference on Computers and Communication Security (CCS 2003), pp. 62–72 (October 2003)
25. Durresi, A., Bulusu, V., Durresi, M., Paruchuri, V., Jain, R.: Key distribution for heterogeneous mobile wireless networks. In: Proceeding of Globecom 2006, San Francisco, November 27 - December 1 (2006) WSN09-4: 1–5
26. Zhang, Y., Lee, W., Huang, Y.A.: Intrusion detection techniques for mobile wireless networks. Wireless Networks 9(5), 545–556 (2003)
27. Ballagas, R., Rohs, M., Sheridan, J., Borchers, J.: The smart phone: A ubiquitous input device. IEEE Pervasive Computing 5(1), 70–77 (2006)
28. Chen, T.M., Venkataramanan, V.: Dempster-shafer theory for intrusion detection in ad hoc networks. IEEE Internet Computing 9(6), 35–41 (2005)

COMANCHE: An Architecture for Software Configuration Management in the Home Environment

Sara Grilli[1], Andrea Villa[1], and Christoforos Kavadias[2]

[1] CEFRIEL, Center for ICT Excellence, via Fucini 2, 20133 Milano, Italy
{sara.grilli,andrea.villa}@cefriel.it
[2] TELETEL SA, 124, Kifissias Ave., 11526 Athens, Greece
c.kavadias@teletel.eu

Abstract. The number of "intelligent" devices with computing and communication capabilities, that surround people in the home environment, are continuously growing. This kind of devices is characterized by software running on them but this needs to be configured and updated from time to time. The user who wants to configure and update these devices has to face the variety of devices configuration procedures and to find the software updates from different manufacturers. This intrinsic difficulty in configuring and updating devices, together with their spreading, make the necessity of an automatic configuration system grow. The COMANCHE project has proposed a complete architecture for software configuration management in the home environment, which aims to overcome the heterogeneity of configuration and upgrading procedures and will allow the deployment of value added services. This paper will give a screenshot of the defined architecture, comparing it also with other configuration and updating existing solutions, like OMA and TR-69.

Keywords: Ambient Intelligence, Software configuration management, pervasive computing, semantic modeling, home automation.

1 Introduction

In these last years, the trend towards nanotechnologies and distributed computing has increased rapidly, involving a great number of devices, sensors and tools. Starting from mobile phones and electronic devices, this evolution has now reached also the home environment with white goods, audio-video devices, set top boxes and game consoles.

Furthermore, the increased use of networks and the Internet has lead to connect together all these devices and sensors in order to better exploit their enhanced capabilities and collect information for services personalization. The common objective is our daily life improvement.

While networked home environment offers great opportunities for developing and deploying value added services, in which also user's position in the house and his preferences can be exploited, the enhanced computing and networking capabilities of home devices have the drawback to put complexity in them. This complexity is unusual for such type of devices, which are considered "commodities" and have to be simple to use.

M. Takizawa, L. Barolli, and T. Enokido (Eds.): NBiS 2008, LNCS 5186, pp. 283–292, 2008.
© Springer-Verlag Berlin Heidelberg 2008

Furthermore, their enhanced computing capabilities require that some pieces of software run on them. This software must be updated in order to fix bugs and to add new services and functionalities, but it also needs to be properly configured and personalized when it is inserted in the specific home environment.

This task, at the moment, is strongly dependent from manufactures: each defines his own configuration and upgrade procedure. As a consequence, the software upgrade, the installation of new services and the configuration of "smart" home devices are complex and usually left to technicians. The necessity of an automatic mechanism for performing the above mentioned task is becoming more and more important, also because some of the existing approaches cannot be used with every type of devices.

The COMANCHE project addresses this issue, aiming to develop and to validate a generic framework for Software Configuration Management (SCM), which will pave the way to the realization of technically and commercially viable private spaces incorporating ambient intelligence features. The COMANCHE approach allows obtaining complete ambient configurations, based on users' wishes, and guarantees a continuous update of devices' software in the home environment. The entire process is invisible to the user, whose only duty is to connect the device to the home network.

2 The COMANCHE Architecture

The COMANCHE infrastructure is based on three main technical components [4]: a distributed knowledge, a component based software architecture and a formal modeling methodology for consistency validation.

The distributed knowledge, coming from a huge number of Home Environment, needs to be managed by a proper framework. For this reason, the COMANCHE Ontology provides the means for conceptualizing, organizing and exploiting effectively the tremendous amount of attribute information, pertaining to the management of home services.

The component-based software architecture and an adequate design methodology are used to address the engineering and to manage the run-time configuration of intelligent devices. This approach guarantees modular software easier to model and to reconfigure.

A formal modeling methodology is used to capture and to analyze the structure of the distributed software system. The generated model is the input of a consistency validation framework that analyzes the run-time systems behavior. This approach aims to preserve the integrity of the target networked services environment across present complex and multi-vendor private spaces.

The COMANCHE specific architecture and entities are depicted in Fig. 1 [4][5]. The main actor in the architecture is the home user who interacts with "intelligent devices" existing in his home. He makes use also of his private services, which he has subscribed to or aims to subscribe through the SCM Service Provider, the Identity Provider, and the different Software/Services Providers and Attribute Providers.

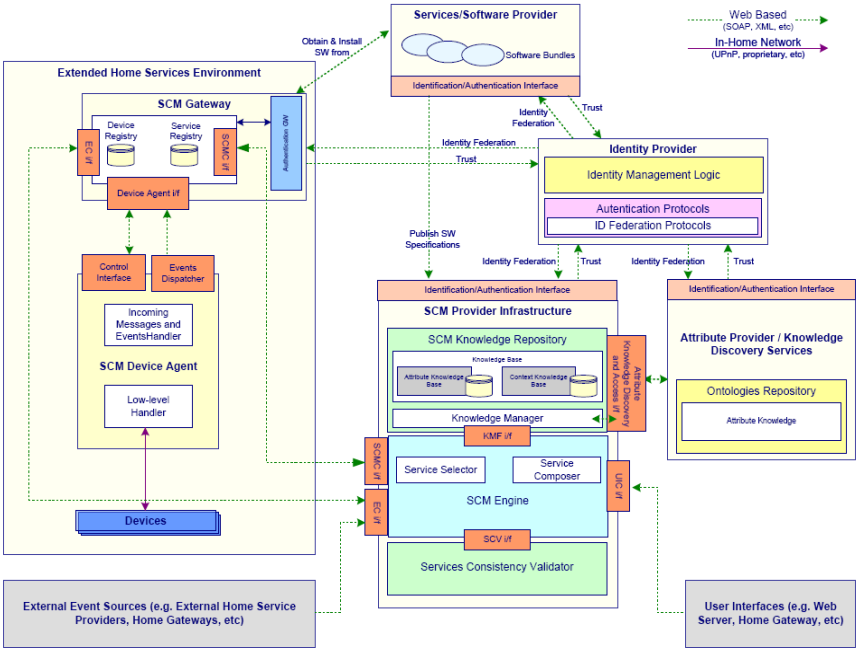


Fig. 1. COMANCHE architecture with components details

2.1 Attribute Providers

The Attribute Provider stores and manages the diverse knowledge required for consistent software configuration management. This includes for example: information on user profile, legal binding issues and technical interoperability between software and hardware subsystems. An Attribute Provider may maintain web sites for allowing users to personalize their services and devices (user-profile manager) or it can become a trusted party that handles user subscription between different Service Providers.

Attribute Providers publish persistent knowledge, structured using the COMANCHE ontology, relating to services, user profiles and business domain relations. The access to the knowledge is performed through the use of appropriate Knowledge Discovery Services realized in Knowledge Registries. The privacy protection regarding the published information is ensured by the Identity Providers.

2.2 Software/Services Providers

The Software/Services Provider supplies software for the user devices. It maintains software repositories containing different device-specific or service-specific software components and firmware updates to be downloaded and installed on the devices.

2.3 SCM Services Framework

The SCM Service Framework is the part of the COMANCHE architecture that is dedicated to the assembly, validation, deploy and delivery of services for the Home Environment. The SCM Service Framework involves a lot of components and constitutes the heart of the COMANCHE project. In the following sections the main components of this framework will be described, focusing on their architecture and their function.

2.3.1 SCM Services Providers Infrastructure

The SCM Service Provider manages the software configuration of the user service environment. By contacting Attribute Providers, it discovers user profiles, legal binding information, business-domain information, technical procedures relating to software configuration, etc. Furthermore, it is able to retrieve and to exploit dynamic context information, such as triggers from the user or external services or faults occurrence. The SCM Service Provider is intelligent enough for assessing the impact of software (re)configuration across complex multi-vendor environments from the user private spaces. This important feature is achieved through the Service Consistency Validator, described in section 2.3.1.3.

The SCM Service Provider is composed of a SCM Engine, a Knowledge Management Framework and a Service Consistency Validator.

2.3.1.1 SCM Engine. The SCM Engine is the central part of the SCM Service Framework and it is responsible for the coordination of SCM procedures and for the derivation of SCM decisions. In order to elaborate SCM decisions, the Engine strongly utilizes the COMANCHE ontology (sec. 2.5), accessing to the acquired knowledge through the use of interfaces with the Knowledge Manager. This information will be properly used to select the adequate service components for a home environment and to elaborate a configuration. Finally, the Engine, validates the derived service compositions prior to deployment through the interworking with the Service Consistency Validator.

The selected service components are synthesized in Composite Service Descriptions and sent to the Device Agent through the SCM Gateway interface. A Composite Service Description (CSD) is the specification of the software configuration of Home Environment. It describes the components that need to be installed, on which devices and how these devices have to be configured.

The Engine will also be capable of interacting with users, through the User interface, to receive user requests regarding, for example, the available services or the installation of a specific service.

As it is shown in Fig. 2, the SCM Engine is composed of a central element, called SCM Flow Manager, which is responsible for the coordination of the performed operations. The Flow Manager interacts with the Service Selector and the Service Composer in order to compose services. The first, strongly interacting with the Knowledge Repository through the KMF Interface, selects the most appropriate services considering functional and non-functional requirements, context information and user specified selection criteria. The user wishes are acquired by a home device and communicated through the UIC (User Input Collection) Interface. The Service Composer then

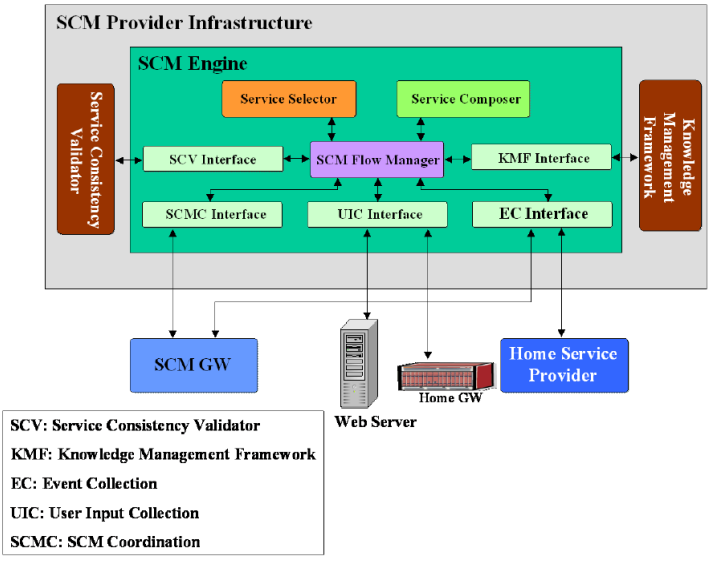


Fig. 2. Functional architecture and interfaces of SCM Engine

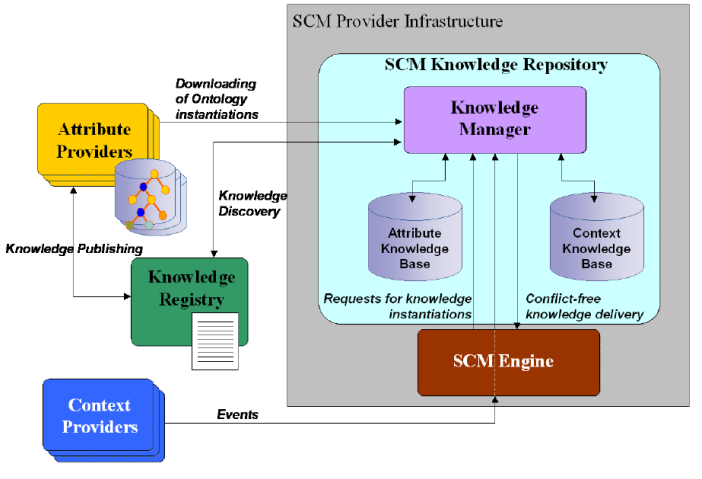


Fig. 3. Functional architecture of the COMANCHE Knowledge Management Framework

takes the selected services to compose Composite Service Descriptions (CSDs) to be sent to the SCM Gateway through the SCMC Interface. Before delivering the CSD, the Engine sends the software configuration to the Consistency Validator, using the SCV (Service Consistency Validator) Interface in order to validate it.

The described SCM procedure can start as result of a user request of new service or after the notification of events. The events, received through the EC (Event Collection)

Interface, are collected from Home Service Providers (e.g. a request for service upgrade), home gateways (e.g. a malfunctioning detection) and SCM Gateways (Service Execution Errors).

2.3.1.2 Knowledge Management Framework. The main purpose of the Knowledge Management Framework is to formulate and to maintain valid and consistent instantiations of the COMANCHE ontology.

As shown Fig. 3 shows, the main entity of this framework is the Knowledge Manager. This element receives from the SCM Engine requests for the delivery of knowledge instantiations and answers, querying the Attribute and Context Knowledge Base with conflict-free services lists. The Attribute Knowledge Base contains the COMANCHE ontologies and static information on software and devices. The information stored into this component comes from queries formulated to the Knowledge Registry, resolving the potential conflicts, while the Knowledge Registry discovers knowledge across the Attribute Providers. The Context Knowledge Base consists of the dynamic part of Home Environment and Context Ontology. Its information is the result of the mapping of events coming from the Context Providers on the Home Environment and Context Ontology.

2.3.1.3 Service Consistency Validator. This component interacts with the SCM Engine to validate SCM decisions prior to software deployment for the integrity of the target environments to be preserved. The validation process is performed executing appropriate off-line consistency validation tests using SDL-based service models. The Service Consistency Validator receives from the SCM Engine a Composite Service Description, discovers and downloads the consistency validation models of all the services contained in it and communicates the tests' result to the SCM Engine.

2.3.2 Extended Home Services Environment

The COMANCHE architecture is also designed to incorporate functional entities that reside in the user environment and enable the user to take advantage of the COMANCHE framework.

One important entity is the Software Configuration Manager Gateway (SCM Gateway), which coordinates software configuration management in the context of a services island on the basis of the SCM specifications obtained from the SCM Service Provider. The SCM Gateway is a software component that runs on a processing and communication capable device and has two main functions:

- Triggering of an SCM procedure towards the SCM Service Provider, as a consequence of an error reported by a Device Agent or of the connection of a new device to the Home Environment.
- Localized coordination of SW configuration actions during the course of SCM procedures. The SCM Gateway receives CSDs from the SCM Engine and translates them into Device Configuration Scripts (DCS) and Device Configuration Package (DCP). The difference between the two scripts is that the first one is only a list of commands while the second one contains also code or binary files.

The SCM Gateway is designed also to retrieve from the respective SCM Device Agents and to provide to the SCM Service Provider dynamic context information.

The SCM Gateway will probably run on a processing and communication capable device located in the Home Environment, maybe the Home Gateway or a home PC. A SCM Device Agent is a software program associated with a device and running on the device itself or on another processing capable device able to command the associated device. The Device Agent receives commands from the SCM Gateway and communicates them to the device, translating them into proper device specific commands.

The SCM Gateway maintains also two registries: an SCM Gateway Registry, which contains the devices known to be connected to the Home Environment, and an SCM Gateway Service Registry, which is a collection of the Software Services known to be installed on each device.

2.4 Identity and Security Management Framework and the Identity Providers

The Identity Provider aims to overcome problems arising from the fact that each user has many serial numbers and keys for accessing services and installing software. The Identity Provider authenticates and identifies all the aforementioned entities as well as verifies the identity of devices and software. Having a trust relationship with the Identity Provider, players and devices don't need to trust each other.

As sensitive information, concerning users, need to be dynamically shared across different business domains, the Identity Provider needs to control information exchanges in order to verify that the user privacy preferences are respected.

The Identity Management and Security Framework is the part of the project dedicated to the management of the COMANCHE components identities in respect of security and privacy principles. This functionality is realized through one or more Identity Providers and consists in authenticating all entities involved in SCM and performing identity federation across different providers.

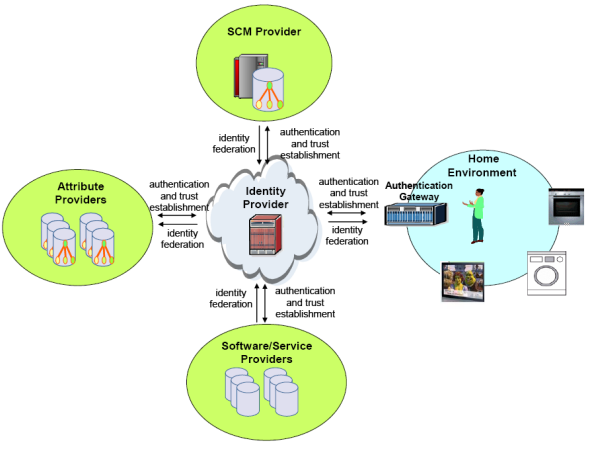


Fig. 4. COMANCHE Identity Management and Security Framework

The entities involved in the framework that are described in Fig. 4, are:

- Identity Provider: authenticates the entities involved in SCM
- Resource Providers: architecture entities that host resources necessary for SCM
- Requestors: architecture entities that request resources from Resource Providers
- Authentication Gateway: software program, running on a processing and communication capable device, responsible for authenticating users.

2.5 COMANCHE Ontology

As described in paragraph 2.3.1.2, the COMANCHE ontology is strongly used by the SCM Provider infrastructure in order to represent the huge amount of information linked to the system. The COMANCHE ontology is described in the following sections and is composed of three elements: Service Ontology, Home Environment and Context Ontology, User & Business Domain Ontology.

2.5.1 Service Ontology

The Service Ontology, based on the OWL-S specification [3], is an ontology developed to represent the services treated by COMANCHE. These can be of two types: SW Services or Internet Services. The first type represents a collection of technical functions installed on a device, while the second type is a collection of function to be accessed through the Internet and hosted by an Internet Service Provider. Each Service is described through a Service Profile and a Service Model. The Service Model tells how to use the service, by detailing the semantic content of requests, the conditions under which particular outcomes will occur and, where necessary, the step by step processes leading to these outcomes. The Service Profile, instead, tells "what the service does", in a way that is suitable for a service-seeking application to determine whether the service meets its needs.

2.5.2 Home Environment and Context Ontology

This ontology is used to represent information about software and services related to Home Environment. The represented information consists of:

- Static information about the compatibility and interoperability aspects of Software, Services and Devices
- Home Environment status, i.e. which devices it contains and what is installed on them
- Users' requirements to be satisfied, i.e. functions to be deployed
- Events that affect and/or trigger SCM procedures
- The SCM outcomes, i.e. services to be installed on each device

2.5.3 Business Domain and User Ontology

The Business Domain and User Ontology is used for organizing and exploiting information relating to business relationship, user identity management and user preferences. The ontology regarding the Business Domain is represented by the Business Organization entity and describes the business components used to deploy services and their interconnections. The User ontology is described through the Person entity

and represents the user subscriptions and privileges. A particular attention is given to the privacy which is connected with the treated data and handled through a proper identity management.

3 Related Works

The problem of devices remote configuration has been already handled in some specific communication technology fields.

A work in that direction has been done by the DSL Forum [1] with the definition of the TR-69 CPE WAN Management Protocol (TR-69 CWMP). TR-69 is a bidirectional SOAP/HTTP based protocol that sets out a common method for Customer premises equipment (CPE) devices to communicate with an Auto configuration Server (ACS). A CPE is a device located on the customer's premises that needs to be connected to the Internet and requires a complicated configuration because of the increasing Internet access possibilities. The performing of the required configuration takes place through the communication with an ACS, that provides configuration and firmware upgrades.

Another interesting protocol for Device Management (DM) is called OMA DM and has been specified by the Open Mobile Alliance (OMA) [2]. This protocol has been designed for management of small mobile devices, such as mobile phones, PDA's and palm top computers, in terms of configuration, software upgrades and fault management. Considering the particular target, the protocol takes into account of the memory restrictions of the devices, the possibly constrained communication bandwidth and the tight security solutions related to the configuration procedures.

The COMANCHE approach is more complete than the presented ones. In comparison with the OMA DM protocol, COMANCHE takes into account of every device typology, while the OMA DM is developed in function of mobile devices. TR-069, even if it takes into consideration different types of devices, is focused on the connection configuration and is designed to configure only one device at time. The COMANCHE project defines a complete configuration management framework with a new approach. In the configuration process COMANCHE considers the whole home environment context, obtained retrieving information from the devices, and it can configure every device in the house. These characteristics open the possibility of offering services that involve different devices, actually realizing a "smart" home environment.

4 Conclusions

With the increasing proliferation of smart home devices, the need of upgrading and configuring their software to enable enhanced capabilities is becoming urgent. On one hand manufacturers would like to have a low-cost automated method for software configuration procedures, on the other hand Services Providers would like to offer value-added and innovative services, which deployment is completely automated, without bothering the user with complex installation and configuration procedures.

The proposed COMANCHE approach is a valid solution to configure and update the increasing number of devices surrounding people in everyday life. However, the

COMANCHE framework looks beyond, managing the ubiquitous computing in order to offer services that involve more devices transparently to the user. COMANCHE will go beyond the used solutions TR-69 and OMA, offering a complete Software Configuration Management Framework that simplifies devices' maintenance and allows the deployment of innovative services towards a number of home devices. COMANCHE offers also a knowledge organization solution, that embraces different knowledge area and source, but it allows the automated generation and validation of the software installation and configuration scripts to be sent in the home environment.

References

1. DSL Forum, http://www.dslforum.org/
2. Open Mobile Alliance, http://www.openmobilealliance.org/
3. http://www.w3.org/Submission/OWL-S
4. Grilli, S., Makri, E., Mouratidis, N., Steblovnik, K., Efremidis, S., Mähönen, P., Meshkova, E.: Software Configuration Management for ambient intelligence: the COMANCHE approach. In: IEEE International Symposium on Ubiquitous Computing and Intelligence, May 21-23, 2007 (accepted for publication, 2007)
5. COMANCHE Deliverable D-2.3 COMANCHE Architecture and Functional/Technical Specifications, http://www.ist-comanche.eu/

Graphic Drawing Tools for Network Traffic Simulation

Shingo Nomoto[1], Kensuke Fukuda[2], Minoru Uehara[1], and Hideki Mori[1]

[1] Department of Open Information Systems, Graduate School of Engineering,
Toyo University, 2100 Kujirai, Saitama, Japan
gz0700132@toyonet.toyo.ac.jp, uehara@toyonet.toyo.ac.jp,
mori@toyonet.toyo.ac.jp
[2] National Institute of Informatics
2-1-2 Hitotsubashi, Chiyodaku, Tokyo, Japan
kensuke@nii.ac.jp

Abstract. We propose a visualization tool for better understanding AS-level Internet Topology. Traditional graph drawing tools do not focus on the AS-level policy, thereby restricting depiction of the direction of traffic flow. The proposed tool provides information about the relationship of AS links, as well as routing tables for nodes.

Keywords: Scale-free, small-world, graph drawing.

1 Introduction

The Internet is a network of Autonomous Systems (ASes), which in reality, are ISPs, companies and universities. It is difficult to understand the network structure precisely, because the internet does not have a central controlling entity. However, many studies have recently used approaches from statistical physics to gain an understanding of the structure of networks. AS level network structures are very different to random graph, small diameter and small-world networks that are good at connecting nodes. AS level networks also have scale-free structures with large deflection and power-law decrement for large numbers of nodes [1].

Information traffic is simulated in the AS-level internet topology, with dynamic query scheduling to evaluate whether servers balance load well and are suitably located. Given the complexity of this topology, graphical drawing tools are necessary when depicting such networks.

It is known that network structure greatly influences the efficiency of information traffic flow [2]. Building an appropriate model is essential for various research fields such as routing, congestion control and load balancing. However, as only simple connection relations between nodes are known natively, it is difficult to evaluate control algorithms for information traffic in unknown network topologies. Even if nodes are connected, information traffic is based on the routing policy between ASes and not only on the connectivity.

Gao classified policies between ASes into provider-customer and peer-peer relationships, and suggested an algorithm for estimating these relations [3]. A

M. Takizawa, L. Barolli, and T. Enokido (Eds.): NBiS 2008, LNCS 5186, pp. 293–302, 2008.
© Springer-Verlag Berlin Heidelberg 2008

Customer is an AS that requires a Provider to translate information traffic. ASes are peers if they translate information traffic for each other. In Provider-Customer relations information traffic is only translated by the Provider for the Customer. These policies greatly influence routing complexity.

We propose graphic drawing tools to understand how these policies influence the routing of information traffic between AS level network objects. With these tools users are able to understand information traffic intuitively. We apply Gao's provider-customer and peer-peer classifications to create realistic routing tables.

2 Small-World and Scale-Free Networks

Small-world networks are characterized by two parameters [4]: their characteristic path length L, and a clustering coefficient C. L is defined as the average of the shortest path between two nodes for all nodes in the network. C is defined as the average clustering of all nodes in the network, where Eq. (1) is used to calculate C and Fig. 1 depicts this calculation.

$$_{k_v}C_2 = k_v (k_v - 1)/2. \quad (1)$$

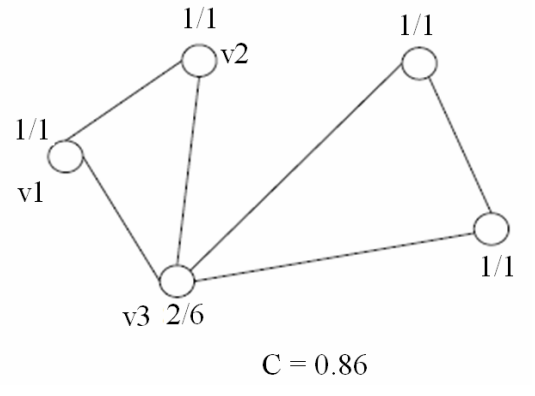


Fig. 1. The neighbors of v1 are v2 and v3. The probability of linking between v1's neighbors is 1. v2 and v3 are linked. The value of C for v1 is 1/1.

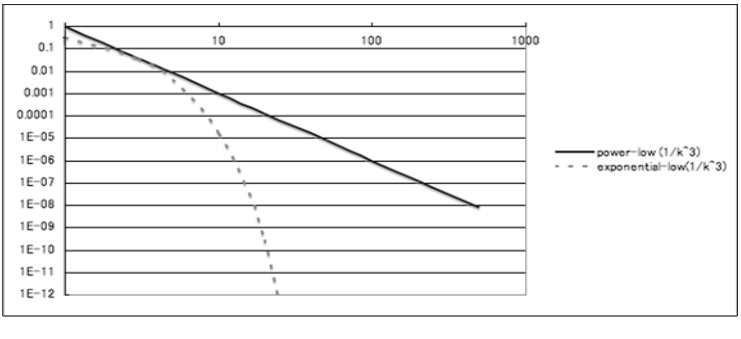


Fig. 2. Power-law and exponential-law graph. The power-law has a long tail.

For large numbers of nodes, scale-free networks have large deflection and the power-law decrement. Figure 2 compares the power- and exponential-laws.

3 Routing

The major routing algorithms are dynamic rather than static, and dynamically update routing tables with path information. Dynamic routing can easily produce routing tables, and can change the network topology as well, but it does have some drawbacks: requiring CPU power at the router, and the large volume of information traffic required in comparison to static routing. However, dynamic routing is the most used routing scheme, because it is difficult to configure routing tables manually for static routing. Various routing protocols are applied both intra and inter AS networks.

3.1 Intra and Inter AS Routing Protocols

An AS is a group of network objects, and uses different routing protocols intra and inter the AS. Routing algorithms can be classified into three types:

i. Distance vector
Distance vector algorithms are based on the Bellman-Ford algorithm and calculate the route to a destination from a neighbor node's information. The cost of the route calculation is hop times at the router.

ii. Link state
Link state algorithms are based on Dijkstra's algorithm and calculate the route to a destination using information from all the nodes. The cost of the route calculation is the weight of the links.

iii. Path vector
In distance vector algorithms, the best route is determined from the number of hop times at the router. In path vector algorithms, the best route is decided from the number of transited nodes. For example, in BGP routing, the best route is determined from the lowest transit times within the AS-PATH list. This scheme is highly reliable and reduces both processing cost compared to distance vector algorithms, and the information required compared to link state algorithms. The routing table is created by the ASes that add their AS numbers to the AS-PATH list recurrently. BGP has further properties, other than the AS-PATH; most importantly, that the path vector type can reflect the policy.

3.2 BGP Routing Algorithm

Figure 3 depicts the BGP routing algorithm. As shown, each AS adds its own AS number to the AS-PATH list.

Separate lists exist for the paths through AS-400 and AS-200, and in turn, each AS adds its AS number to the appropriate list. At the destination node, AS-500, the best route is decided based on how many AS numbers appear in each AS-PATH list. BGP routing is based on policy, because there is routing information like the AS-PATH, routing priority, etc.

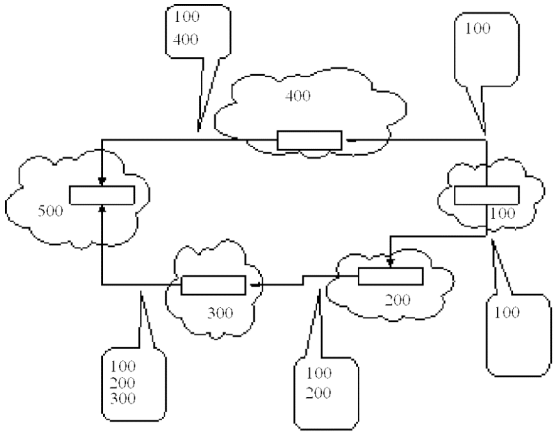


Fig. 3. BGP routing algorithm. Starting from AS-100, each AS adds its own AS number to the list.

Gao suggested a method that uses a valley-free rule to classify the routing policy [4]. This rule estimates provider-to-customer, and peer-to-peer relations from the direction of the routing information. The term ”valley-free” means that traffic cannot get over a valley. There are four conditions in the Valley-free rule. And if the route had some condition, it is not in the routing tables.

A) The route does not have more than two peer-to-peer edges between ASes.
B) The route does not have a customer-to-provider edge after a provider-to-customer one between ASes.
C) The route does not have a peer-to-peer edge after a provider-to-customer one between ASes.
D) The route does not have a customer-to-provider edge after a peer-to-peer one between ASes.

By applying these conditions, we can estimate relations from the routing information list. Figure 4 depicts the Valley-free rule.

4 Graph Layout

The effects of abstracting and visualizing information depend on the graphical layout. Standards for evaluating good layout include rules governing the placement of nodes and lines, distances between nodes and restrictions on line crossing. These can be summarized into an ‘aesthetic standard’. There are both theoretical and mechanical model based algorithms to assist in adhering to the aesthetic standards.

Drawing techniques using mechanical models include spring models as suggested by Eades [6]. Spring models are a simulative approach from system mechanics. Edges are changed to springs. If the distance between nodes becomes large, the spring pulls

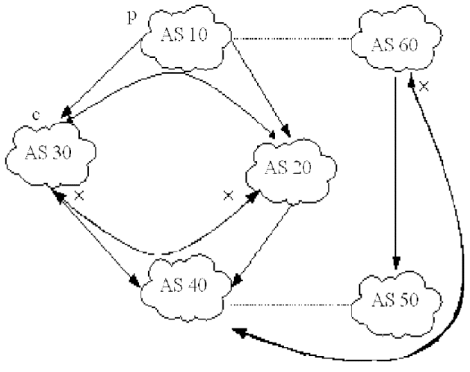


Fig. 4. Valley-free rule in a simple topology network. Dotted lines denote peer-to-peer edges, while straight lines denote provider-to-customer edges. Curved lines depict information traffic.

the far nodes inwards. If the distance between nodes becomes small, the spring pushes the nodes apart. In the simulation, each spring calculates its length and the resulting forces, and iteratively changes the layout until a stable compliant representation is achieved.

5 Design and Implementation

We have proposed functions for graphic drawing tools in a previous work [7].

i. Indication of routing tables.
ii. Indication of best and second best route.
iii. Visualization of links policies.
iv. Visualization of AS hubs.

We want to know the end to end routing of information traffic flow for an AS-level internet topology. So, we apply the properties of a small-world structure network, scale-free structure network and Valley-free rules for AS-level internet routing. We have found that the small-world structure network's parameters, characteristic path length L and clustering coefficient C, the scale-free structure network's, large deflection and power-law decrement for large numbers of nodes, and the Valley-free rules are very effective in routing.

The routing tables for the first function are contructed using the BGP algorithm and Valley-free rules. All nodes have the routing tables of each other. When we click on a node, we can see the node's routing table and get the traffic flow information.

The second function is performed to highlight lines for the best and second best route. This is because important nodes, apart from hub nodes, surely have more than one route to the internet backbone.

The third function is performed to highlight the provider-to-customer and peer-to-peer edges as arrows or colorful lines. The Valley-free rules give us the information traffic flow.

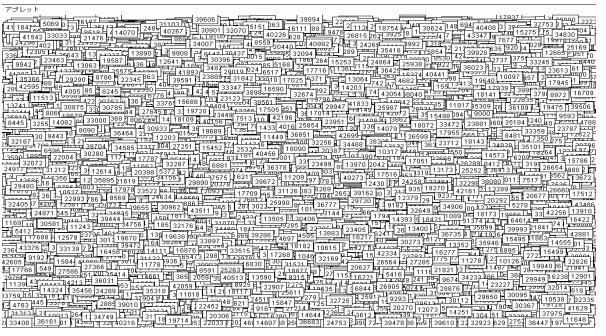


Fig. 5. Depicting 1/1000 ASes

As the final function, a hub AS is an important factor in small-world and scale-free structure networks. Saying that the hub ASes make the routing tables is not an exaggeration, because hub nodes make the path lengths between two nodes shorter for all nodes in the network. Hub nodes therefore have a large influence on routing.

Overall we will be able to gain a better understanding of the AS-level network information traffic flow.

Input data is made to conform to CAIDA's format with Valley-free rules; that is, "first-AS-number, second-AS-number, relation". The relation is that of the first AS, being provider, customer or peer, to the second. Peer-to-peer relationships are 4002 of the 26023 edges. We have constructed graphs from this data.

Figure 5 shows almost all the AS data being drawn for 26023 AS input values. We illustrate 1/1000 of the ASes in Fig. 5.

Nodes completely fill the drawing field. This resolution provides no understanding of how nodes are linked to each other. When spring models, that coordinate the distances between nodes, are applied to the drawing, Fig. 6 is obtained.

In Fig. 6, a large number of nodes are obscured by the hub node. Therefore we reduce the scanning rate depth from the start node and Fig. 7 is obtained.

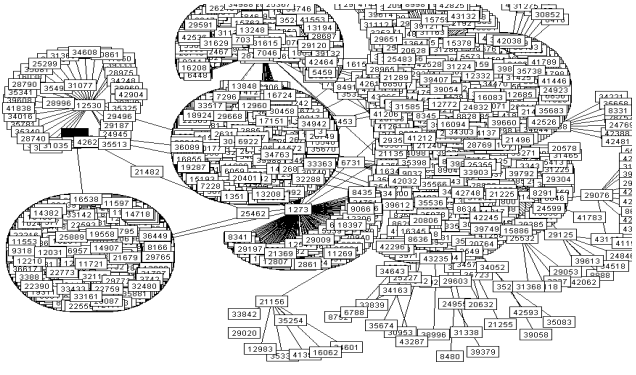


Fig. 6. Spring model drawing

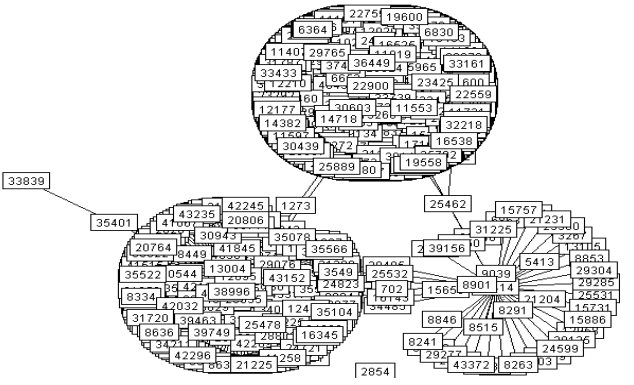


Fig. 7. Setting limits in drawing depth

Here, nodes form crowded zones because the network has large hub ASes from scale-free structures and a small diameter from the small-world structure. Therefore, We adjust the power of springs and scanning over 200 degree, Fig.8 is obtained.

We set the scanning limit of degree over 200, because if scanning limit set over 3 or 2, we got a graph like Fig.5. We can see several node which as a bridge between hub ASes and ASes that have many edges on visualization. Next, we have got a minimum spanning tree graph as shown in Fig.9, though Fig.8 is an undirected graph.

Finally, we display the AS-level graph representation based on the Valley-Free rule.

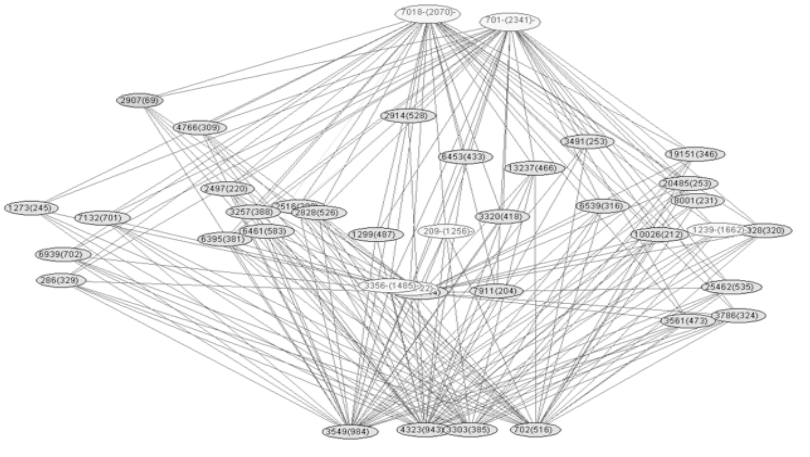


Fig. 8. Undirected graph of ASes. Setting the starting node at scanning as AS2907 and scanning over 200 degree nodes. And springs pull the hub ASes that have many edges on visualization to upside, and pull the ASes that have many edges on visualization to downside. AS nodes are containing “number (degree)”, and hub AS nodes are containing “number –(degree)-” and light color.

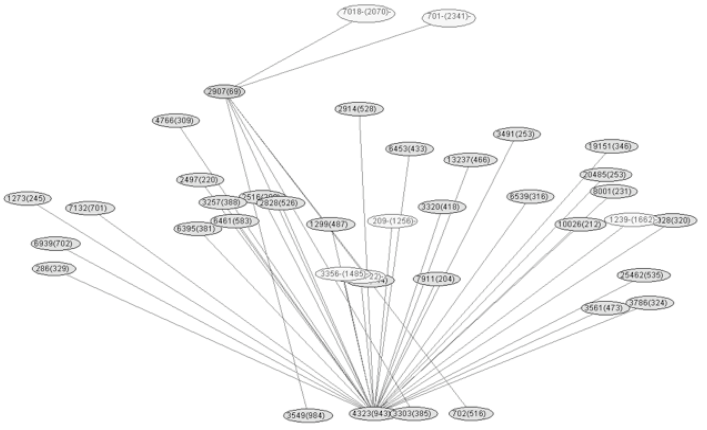


Fig. 9. Minimum spanning tree graph of ASes. Setting the starting node at scanning as AS2907 and scanning over 200 degree nodes.

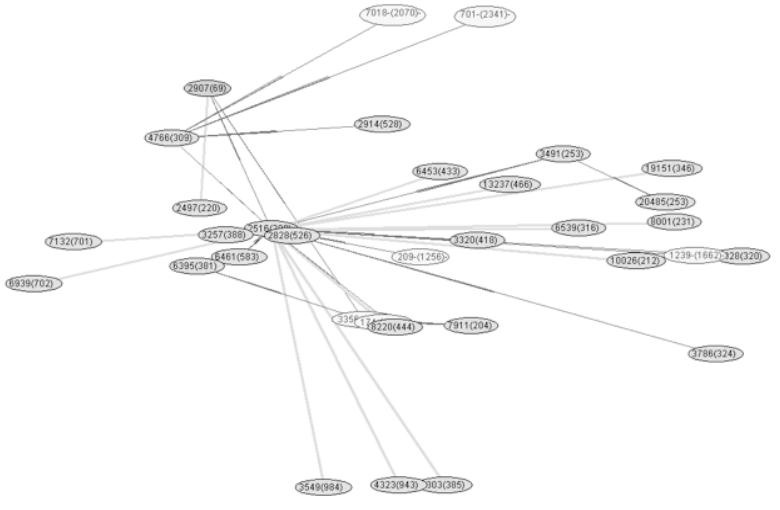


Fig. 10. Valley-Free graph. Setting the starting node at scanning as AS2907 and scanning over 200 degree nodes.

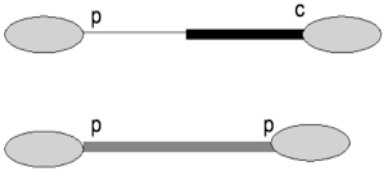


Fig. 11. Legend of the Valley-free relationship shown in Fig.10. Upper link corresponds to the provider-customer relationship, and lower link is the peer-peer relationship.

Comparing minimum spanning tree graph with Valley-Free graph, we can easily confirm that many edges in the minimum spanning tree are rewired in the Valley-free graph. In BGP, all nodes have these routing tables, so we calculate routing tables for all nodes. Edges of the graph are highlighted with peer-to-peer and provider-to-customer relationships of Valley-free rules (see Fig.11). Two figures clearly demonstrated the importance of the Valley-free rule for information transfer for large-scale network simulation. Thus, ignorance of this rule might yield to underestimation/overestimation of any proposed algorithm.

6 Conclusion

We have designed graphic drawing tools. In this work we implemented graphic drawing using a 'mechanical spring' algorithm. We also proposed functions that highlight properties of small-world, scale-free structure and Valley-free rules. Hub nodes from small-world and scale-free structure have a great influence on routing. Hub nodes typically have a large degree, with many links to leaf nodes and some to transit nodes. If we draw the network normally as shown in Fig. 5, we do not obtain a usable graph of the network topology. Thus, we used a mechanical spring model to visualize the network topologies, as shown in Fig. 6. In this figure, we can see that there are very few hub nodes, but that they have many links connected to leaf nodes and a few links connected to transit nodes. Some of our proposed functions are suited to hub nodes. AS hubs are very important when drawing an AS level network, because it is normally very congested around hub nodes.

In this paper, we have proposed a new function that controls hiding leaf nodes and visualize AS networks. By setting the scanning rate depth at nodes, we obtain the graph, shown in Fig. 7, in which the large number of leaf nodes hide the hub nodes and transit nodes. Instead of a simple graph, we see only the hub node links with a mass of leaf nodes. In scale-free structures, almost all nodes have a small degree, and very few nodes have a large degree. In other words, almost all nodes are leaf nodes, and few nodes are hub nodes. We can see several hub nodes in Fig.8.

We implement some ideas for scale-free structure in graphic drawing tools. In Fig.8, hub nodes are pull to upside and nodes that have many links on visualization are pull to downside. And we set the scanning rate, over 200 degrees, at scanning nodes. And we make Fig.9, minimum spanning tree graph of ASes. In Fig.9, we can see simply linking of ASes. Moreover, We make Fig.10, Valley-Free graph. Many minimum spanning tree's edges are vanishing in Fig.10, Valley-Free graph. Valley-Free graph have different structure from minimum spanning tree graph, and more complexity connecting than minimum spanning tree graph.

References

1. Pastor-Satorras, R., Vespignani, A.: Evolution and Structure of the Internet. Cambridge University Press, Cambridge (2004)
2. Albert, R., Jeong, H., László Barabási, A.: Error and attack tolerance of complex networks. Nature 406, 378–382 (2000)

3. Gao, L.: On Inferring Autonomous System Relationships in the Internet. IEEE/ACM Transactions on Networking 9(6), 733–745 (2001)
4. Watts, D.J., Strogatz, S.H.: Collective dynamics of 'small-world' networks. Nature 393, 440–442 (1998)
5. Cooperative Association for Internet Data Analysis, http://www.caida.org/
6. Eades, P.: A Heuristic for Graph Drawing. Congressus Numerantium 42, 149–160 (1984)
7. Nomoto, S., Fukuda, K., Uehara, M., Mori, H.: Design of Graphic Drawing Tools for Network Traffic Simulation, IPSJ SIG Technical Reports, 2007-DPS-33, pp.37-41 (2008)

A Methodology for the Enterprise Information and Communication Infrastructure Design Process

Natalia Kryvinska[1], Lukas Auer[1], Christine Strauss[1], and Peter Zinterhof[2]

[1] Department of Business Administration, University of Vienna,
Bruenner Strasse 72, A-1210 Vienna, Austria
n-v-kryvinska@gmx.net, a0700895@unet.univie.ac.at,
christine.strauss@univie.ac.at
[2] Department of Scientific Computing, University of Salzburg,
Jakob-Haringer-Str. 2, 5020 Salzburg, Austria
peter.zinterhof@sbg.ac.at

Abstract. Current information and communication services are growing rapidly in directions very different from those pursued in the past. The strong drive for these services' fast development and deployment has not been accompanied by a proper understanding of their engineering and design principles, and for the most part, the services' satisfactory qualitative performance is being obtained by the strength approach of over-provisioning. Thus, there is a strong need for a deeper quantitative understanding of the services' functioning, an understanding that leads to clear design and control methods that it turn make the services' efficient deployment and exploitation possible. The research performed within this paper covers analytical and design aspects of technical and economical challenges emerging in multiservice networks and was conducted from a methodological perspective of evolving network infrastructure. The paper focuses on challenging issues in the planning and design of enterprise communication infrastructure and also determines strategies for the successful implementation converged networks.

Keywords: Enterprise Converged Network, Design Methodologies, Delay Sources, QoS, VoIP.

1 Introduction

Open systems and new industry standards have helped to broaden the field of information and communication services that fit business needs. As a result, the challenge of how to best navigate these choices and ensure that an enterprise achieves new efficiencies from IP-based communications solutions has become more prominent. Thus, the key goal of this paper lies in determining strategies and tactics for successfully implementing converged networks [1, 2].

Specifically, this paper will examine the conceptual background and operation of these converged networks, which are much more complex than single-purpose voice or data networks. Because of this complexity, the design, implementation and testing of the infrastructure require rigorous attention to details in order to present end users with an acceptable Quality of Service (QoS) [3].

M. Takizawa, L. Barolli, and T. Enokido (Eds.): NBiS 2008, LNCS 5186, pp. 303–312, 2008.
© Springer-Verlag Berlin Heidelberg 2008

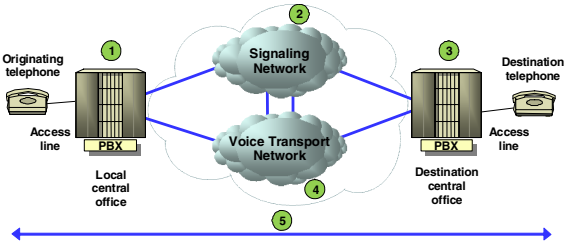


Fig. 1. Voice call processing flows

The problems encountered by early adopters have been largely resolved and for mainstream organizations, the decisions that must be made no longer involve "if VoIP", but rather "when VoIP". Consequently, we provide a roadmap for optimized VoIP implementations. Industrial research shows that sufficient experience with what works across a large variety of organizations, locations and business applications now exists. As a result, we can synthesize the lessons to move to VoIP cost-effectively. It is not a simple recipe to follow, but it outlines a thoughtful approach containing an admixture of the ingredients that make a successful plan [4 ÷ 7].

2 An Analysis of Network Performance Features

An analysis of the performance features of existing as well as developing networks is the first step towards planning and optimizing performance. To this end, we compare the protocol flows in Intelligent Network (circuit-switched) and VoIP (packet switched) voice call scenarios (Fig. 1).

The IN/PSTN telephone call involves five components: end user equipment, such as telephone sets; connections to the local exchange central office; local switching offices; a signaling network; and a transport network. A classic telephone call through an IN engages five key steps:

1. A call establishment request from an analog-telephone is recognized at the local central office, which returns a dial tone and accepts the destination telephone number.
2. A local central office generates a call setup message that passes through the signaling network, identifying a path to the desired destination.
3. A destination central office signals the destination telephone of an incoming call using ringing tone.
4. As the destination party answers, the signaling network starts billing, and the voice network connection becomes active.
5. Two parties carry on a conversation, and when completed, hang up their telephones. The on-hook signals are recognized at the local and destination central offices.

A call establishment over a packet network is similar to the IN case. Five key steps are involved in this scenario:

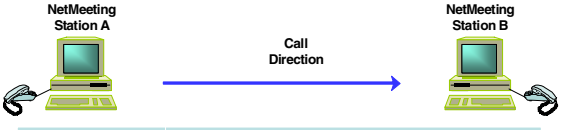
1. An analog telephone places a call to a remote telephone connected to an IP network. The signaling network within the PSTN receives the calling user's destination number, which is then passed to the VoIP gateway.
2. A call signaling information is passed from the gateway to the gatekeeper requesting admission to the network. This signaling information is sent using TCP/IP for greater reliability.
3. While the destination telephone resides on another network, the respective gatekeepers communicate signaling information requesting call completion.
4. The destination user and remote gatekeeper exchange signaling information.
5. When all call signaling is completed, the two end stations exchange media information using RTP/UDP/IP. After the information transfer is completed, additional signaling messages are used to disconnect the call.

When comparing the IN and VoIP call scenarios, we can see two network operations are involved in each case: signaling (for call establishment, management and termination) and media information transport (from the source to the destination). Analyzing these operations is the central theme of our next sub-sections [8 ÷ 10].

3 Voice Call with the "Microsoft NetMeeting"

Analyzing voice/video call scenarios over IP sessions is more complicated than doing so for LAN/WAN applications, because IP sessions involve inter-related protocols. As an example, we consider the simple network consisting of two workstations running the Microsoft application NetMeeting, as shown in Fig. 2. In this scenario, NetMeeting Station A initiates a call to NetMeeting Station B, and thus initiates the protocol interactions we study.

The protocol operations for these workstations follow a sequence of four functions that are found with most voice and data communication processes. These functions are detailed in Fig. 2 and include the direction of the information flow:



Direction	Description
←	TCP: Open Connection for H.225
→	H.225: SETUP
←	H.225: CONNECT
←	TCP: Open Connection for H.245
→	H.245: Terminal Capability Set
←	H.245: Terminal Capability Set
←	H.245: Master/Slave Determination
→	H.245: Master/Slave Determination
←	H.245: Open Logical Channel
→	H.245: Open Logical Channel
↔	RTP: Media Session (G.723)
←	RTCP: Report messages
→	H.245: End Session
→	H.225: Release Complete
→	TCP: RST

Fig. 2. H.323 NetMeeting-to-NetMeeting topology and connection protocol sequence

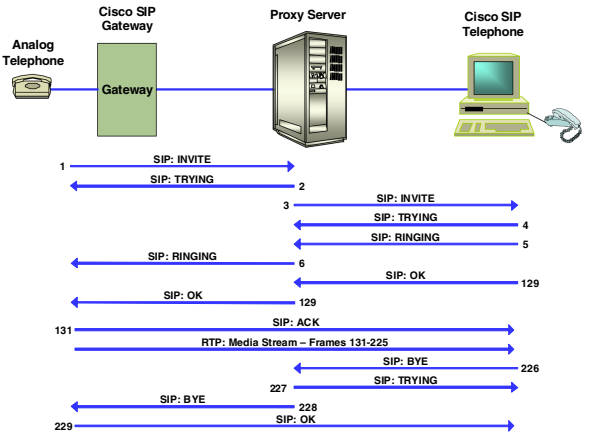


Fig. 3. SIP Control Message Details

- Connection establishment using TCP and H.225;
- Parameter exchange using TCP and H.245;
- Information transfer using RTP and RTCP;
- Connection termination using H.245, H.225, and TCP.

A number of protocols, including TCP, H.225, H.245, RTP, and RTCP, are involved in a multimedia communication session over a converged network. From the protocol interactions illustrated, and the complexities of some of the messages, it is clear that the network manager's job is made easier by having appropriate tools, such as a network analyzer, to identify and resolve H.323 call control issues [11 ÷ 13].

4 SIP Phone Voice Call Associations

As another example, we consider SIP (Session Initiation Protocol) phone voice call associations, which provide the signaling functions necessary to establish, manage, and terminate a connection between multimedia endpoints. For the SIP protocol analysis, consider the network shown in Fig.3, in which an analog telephone connected to a Cisco Systems SIP gateway initiates a call to a Cisco SIP telephone connected to the same network. A proxy server exists on this network and is involved in the call establishment and termination procedures. The SIP control messages flow between the SIP devices and the proxy server, and the RTP media information flows between the SIP gateway and the SIP telephone.

The SIP control messages can be broken down into three general categories. These functions are detailed in Fig. 3 and include the direction of the information flow:

- Connection establishment using SIP;
- Information transfer using RTP;
- Connection termination using SIP.

A SIP-based network includes two key elements: user agents and servers. The user agents exist on an end system, such as a SIP-based telephone, and can establish connections with other SIP peer devices. Four different types of servers, which may be logical (instead of physical) devices, are also defined within the SIP architecture: The proxy server acts as an intermediary, forwarding requests on behalf of other devices. The redirect server forwards a client to the appropriate location to complete a task. The registrar server keeps track of end stations in conjunction with a database known as a location server (or the location service).

To sum up, connections using SIP signaling are not as complex as those using the ITU-T H.323 protocol suite, which explains the growing interest in SIP and SIP-supported products [14, 15].

5 Converged Network Design, Realization Challenges

The network design and implementation phase is divided into steps, beginning with a clear definition of the objectives for the converged network and defining the applications that this network must support. Most customers have had quite positive experiences with voice calls over the PSTN, having experienced fast and reliable call setups and clear connections. These positive experiences have raised the bar for end-user expectations, which means that the end-user experience for converged network must equal, if not exceed, these experiences in order to be considered successful. Many factors can influence QoS, including the bandwidth of the communication channels involved, network loading factors, and latency or delay. These factors must be considered while implementing converged network principles into enterprise architecture.

The challenges are consequences of the fundamental differences between voice and data networks: voice networks are connection-oriented, while data networks are connectionless. As a consequence, these two networks approach the flow of information quite differently.

Besides the differences in reliability, traditional voice and data networks were designed to provide different information transport services. Human communication is sensitive to end-to-end transport delay, as well as variations in that delay, which is known as “jitter”.

In contrast, traditional packet-switched networks do not establish a fixed end-to-end path but provide routing on a per-packet basis. Delay and jitter are of less concern for PSTNs, since it is likely that the large messages will have been divided into multiple packets. These two processes are typically part of the packet-switched architecture. However, if one of the packets in this sequence is lost, then the original message cannot be reassembled and passed to the application.

Next, a preference for which architecture is to be deployed must be established. There are two fundamental types, as well as some hybrid variations on these themes. The first architecture is based on a voice switching system known as a private branch exchange, or PBX. This architecture derived its origins from voice architectures such as PBX networks connecting end users within a single company location, or inter-location trunks connecting branch offices back to headquarters.

A further alternative is a router-based architecture. Minimal reconfiguration is typically required to upgrade the existing routers with VoIP capabilities, such as traffic management and Quality of Service (QoS) features, access and security functions, and other VoIP processing requirements. Routers are placed at higher density sites where traffic is aggregated for transport to other locations. Leased lines connect these core routers with each other, plus links to other networks such as the PSTN or Internet. Single or remote users gain access to the core network through broadband Virtual Private Network (VPN) connections.

The IP PBX-based architecture is more centralized, while the router-based architecture is more distributed. As would be expected, each approach has advantages and disadvantages. For the IP PBX network, a reconfiguration may involve fewer subsystems and intermediate steps. For the router network, sites can be migrated individually, with less disruption likely on the network as a whole. With possibly more sites to migrate, the implementation time may be lengthened.

Both of these approaches, IP PBX and router-based, share a common concern: will the new system interoperate with existing systems, such as a voice mail processor, automatic call distributor or call detail recorder? We address this concern by first defining the steps necessary to design and implement a VoIP network [3, 16, 17].

6 Planning and Design Methodology

The converged network is intended to support both voice and data communications and the design process must consider both of these requirements. While each network presents unique challenges, several common factors must be considered in any event (Fig. 4):

1. *Definition of design objectives*, including the applications to be supported, and the objectives presented as a short mission statement that allows all parties involved to understand the challenges. In addition, the document can include other factors that are driving the project.
2. *Examination of existing voice and data operating environments*, containing tangible loading factors on LAN and WAN segments, as well as the anticipated bandwidth requirements for any new data applications. Current call patterns and anticipated growth during busy periods is also taken into consideration.
3. *Specification of system requirements*, delineates functions the network must provide and includes the appropriate design objectives. Involves the determination if an architectural infrastructure, such as the ITU-T H.323 or IETF SIP, is preferred for this network application. Also takes into consideration any changes that will be necessary to the telephone dialing plans, IP subnets or other issues that could impact end users.
4. *Contemplation of support systems and tools that will maintain this network*, e.g., for network management, network analysis, security, and end-user help desks. This factor also involves the verification of any of these support systems, such as the network management console or protocol analyzer, are also prepared to support the new environment.
5. *Discussion with and feedback from vendors*, in order to obtain their input on the systems requirements.

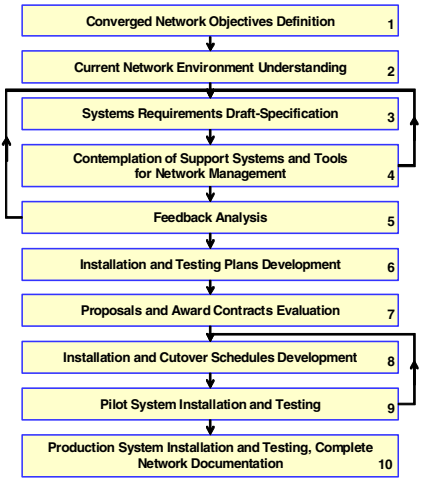


Fig. 4. Network Design and Implementation Plan

6. *Systems approval and installation plans progress*, including interoperability testing between various components and verification of all signaling protocols between dissimilar networks.
7. *Evaluation of proposals from different vendors and the awarding of the contract* , it is necessary to hold two rounds of proposals: one that is open to all interested vendors, and a second round that is limited to the most-qualified ones. This activity also includes the additional support systems, such as network management and troubleshooting tools, so that acceptance testing of the system can proceed once installation is complete.
8. *Development of the installation and cutover schedule* with the vendors, and other organizations, such as carriers, that are also part of the project. Verification of the critical paths in the schedule and accomplishing agreements from all involved parties regarding the importance of this schedule.
9. *Pilot system installation and testing of its operation and applications* before expanding to the larger network. At this stage, it is necessary to use a protocol analyzer to document successful call setups. This documentation, done in a controlled environment, may be useful if deployment problems arise down the road. Before rolling out the implementation to hundreds or thousands of end users, it is also necessary to verify that all dialing plan and network address changes have been successfully migrated to the new environment with this small group of stations.
10. *Installation of the production network*, verification that all network components operate as planned, including interoperability between multivendor systems and supporting systems, such as voice mail processors, and procedures documenting, completion of network drawings and other appropriate documentation [3, 18].

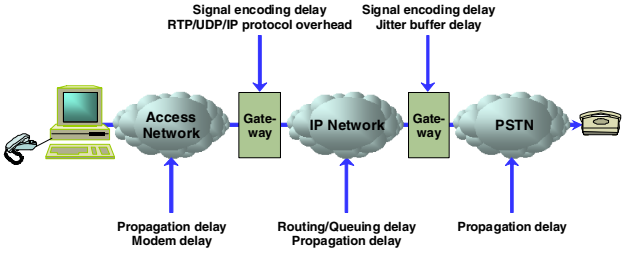


Fig. 5. The Sources of Delay in Converged Network

7 VoIP Service Quality Evaluation and Analysis of Delay Sources in Converged Network

There are relatively big differences in the reliability of voice and data networks, with voice networks designed for very high reliability on account of life-safety issues, whereas data networks operate from the perspective of "best efforts" service. When the upgrade of an existing voice network with a VoIP infrastructure is planned, one cannot expect end users to accept a downgrade in system performance, known as quality of service (QoS). Consequently, the factors that affect QoS must be addressed and the industry standards that have been developed to measure those factors must be considered before advancements activities [3, 19 ÷ 24].

We perform a network analysis in order to find all the possible sources of latency and to increase the quality of VoIP service, with our main concerns being problems related to network delay in VoIP [19].

A number of network design and operations factors affect QoS, including: packet loss and delays, the available bandwidth, the WAN protocols in use and their efficiencies, the presence of echo that is caused by impedance mismatches within the PSTN, and the use of silence suppression that optimizes utilization of the communication facilities. However, the amount of latency, or the delay of the end-to-end transmission, is possibly the most important factor affecting the QoS. A maximum delay of 150 milliseconds in one direction of transmission has been the accepted standard for many years, and generally provides acceptable results for end-users. If this delay is exceeded significantly, then conversation becomes more difficult. As part of the network design, a delay budget may be developed to identify components of this end-to-end delay (Fig. 5):

- Signal encoding algorithm, usually 15-37.5 milliseconds at both the origin and destination;
- Protocol processing overheads in components to include RTP, UDP, and IP information plus echo cancellation, typically less than 5 milliseconds;
- Bandwidth and utilization of channels, which may introduce queuing delays in the range of 5-25 milliseconds, depending on the transmission rate;
- Routing, queuing, and propagation delays across the WAN that depend on transmission media and distance, typically 10-40 milliseconds;

- Jitter, since the packetized voice samples may take different paths through the packet-switched network, arrival rates of those packets may vary, typically 20-40 milliseconds.

Many of the factors above are fixed and cannot be improved, while others are options that can affect network performance. Many encoding algorithms have been designed, some of which are published as ITU-T standards, and others that are proprietary to a particular vendor. These techniques differ in their underlying mathematical algorithms as well as in their end results. Thus, the network manager faces a tradeoff of bandwidth for delay, and must balance the higher cost of bandwidth against the reality of lower QoS from increasing delays.

For many years, the telephone industry has employed a subjective rating system known as the Mean Opinion Score (MOS), to measure the quality of telephone connections. This technique are defined in ITU-T P.800, and are based on the opinions of many testing volunteers who listen to a sample of voice traffic and rate the quality of that transmission. The volunteer subjects are asked to provide their opinion of the connection they have just been using, based on a five point scale (e.g., Quality Rating: Excellent -5, Good -4, Fair -3, Poor -2, Bad -1).

Since the test subjects are human, and the testing therefore subjective, some variation in the scores is expected. A MOS of 4 is considered “toll quality” within the telephone industry, and is generally the accepted standard that is targeted for quality VoIP implementations.

Another ITU-T standard, P.862- Perceptual Evaluation of Speech Quality (PESQ), defines telephone transmission quality tests that are more objective. It addresses the effects of filters, variable delay and coding distortions, and is thus applicable for both speech codec evaluation and end-to-end measurements. The PESQ algorithm is fairly complex, but nevertheless produces a summary output that is useful for comparison purposes. In summary, the P.800 MOS test is a subjective evaluation, while those tests defined in P.862 are objective measurements, with both widely used within the industry [3, 19 ÷ 24].

8 Conclusions

The studies performed within this paper encompass analytical and design aspects of technical and economical challenges emerging in enterprise multiservice converged networks that support new voice and multimedia services, mobility and internetworking. The studies are conducted both from a methodological perspective and from the point of view of the evolving network infrastructure and the traffic carried by it.

Specifically, we have addressed the challenging issues involved in providing an appropriate planning, design, implementation and further maintenance model for converged network. We have performed an analysis and have determined the strategies and tactics for the successful implementation of converged networks. In addition, the design, implementation and testing of the infrastructure requires rigorous attention to the QoS, to address the increased complexity of converged network.

References

1. Technical Paper: Full Service Broadband Architecture, 284 23-3098 Uen Rev, Ericsson (December 2006)
2. An Executive Briefing Paper: Strategies for IP Telephony Evaluation and Migration: Best Practice Considerations for Deploying IPT in the Enterprise, InfoTech (April 2005)
3. Miller, M.A.: Implementing the VoIP Network: A technical briefing series on VoIP and converged networks. DigiNet Corporation 3 (August 2005)
4. Pierce, L.: Obstacles to Migrating to New Telecom Services, Forrester Research (September 2005)
5. Miller, M.A.: Introduction to Converged Networking A technical briefing series on VoIP and converged networks. DigiNet Corporation 1 (August 2005)
6. White Paper: Convergence Can Be Cruel, Network Physics (2007)
7. Lazar, I.: IP Telephony System Manageability: Architecture Matters, Collaboration and Convergence, Nemertes Research, Special Report, Network World (2007)
8. Miller, M.A.: Protocols for the VoIP and Converged Network A technical briefing series on VoIP and converged networks. DigiNet Corporation 2 (August 2005)
9. Wallingford, T.: Switching to VoIP, 1st edn., June 2005. O'Reilly, Sebastopol (2005)
10. Varshney, U., Snow, A., McGivern, M., Howard, C.: Voice Over IP. Communications of the ACM 45(1) (January 2002)
11. Miller, M.A.: Managing Call Flows Using H.323 A technical briefing series on VoIP and converged networks. DigiNet Corporation 4 (September 2005)
12. Rungta, S., Ben-Shalom, O.: Enterprise Converged Network - One Network for Voice, Video, Data, and Wireless. Intel Technology Journal 10(01) (February 15, 2006)
13. White paper: Bringing the Telephone into the 21st Century, The Real Value of Converged Applications, Citrix Systems (2005)
14. Miller, M.A.: Managing Call Flows Using SIP A technical briefing series on VoIP and converged networks Mark. DigiNet Corporation 5 (September 2005)
15. Kolbehdari, M., Lizotte, D., Shires, G., Trevor, S.: Session lnitiated Protocol (SIP) Evolution in Converged Communications. Intel Technology Journal 10(01) (February 15, 2006)
16. White paper: Extending VoIP to Remote Locations: Challenges and Solutions, Quintum Technologies (2006)
17. Herrell, E.: Enterprise IP Telephony Plans in 2006, Forrester Research (June 2006)
18. Basart, E.: Building Reliable IP Telephony Systems How Architecture and Design Differentiate, ShoreTel (October 2006)
19. Boutremans, C., Le Boudec, J.Y.: Adaptive joint playout buffer and FEC adjustement for Internet Telephony. In: Proceedings of IEEE Infocom, San Francisco, California (April 2003)
20. Technical Paper: Convergence: Preparing the Enterprise Network, Hewlett-Packard Development Company, L.P. 4AA0-0740ENW, 06 (2005)
21. Pierce, L., Thomas, B., Herrell, E., Bartolomey, F.: The Forrester Wave: US Enterprise-Class VoIP Services, Q1 2007, AT&T and Verizon Are Ahead Of the Pack, Forrester Research (February 9, 2007)
22. Sacker, S.M., Santaiti, M., Spence, C.: The Business Case for Enterprise VoIP, Intel Corporation (February 2006)
23. Technical Paper: Thinking About Enabling Convergence, Taking Advantage of Services Over IP, AT&T Knowledge Ventures, 07/12/06 (2006)
24. Turek, M.: Voice and Video over IP: Leveraging Network Convergence for Collaboration, Nemertes Research, Issue Paper (2006)

A New Networked Surveillance Video System by Combination of Omni-Directional and Network Controlled Cameras

Yousuke Sato, Koji Hashimoto, and Yoshitaka Shibata

Faculty of Software and Information Science, Iwate Prefectural University
152-52 Sugo, Takizawa, Iwate, Japan 020-0193
g231f007@edu.soft.iwate-pu.ac.jp, hashi@iwate-pu.ac.jp,
shibata@iwate-pu.ac.jp

Abstract. In recent years, in the surveillance system which observes the behavior of human invasion to building or indoor, it is required not only to capture the high quality and wide area image, but also to automatically track to the specific suspicious person in real-time to reduce the number of the required surveillance cameras. While these installations have included more video streams, they have been also placed in contexts with limited personnel for monitoring. Using the suggested system, the location of the target motion objects in wide area with 360 degrees surround it can be detected and tracked by capturing high quality images in real-time.

Keywords: surveillance system, video streaming, omni-directional camera.

1 Introduction

In surveillance system to keep safety and security for humans, it is required not only to capture the high quality and wide area image, but also to automatically track to the specific suspicious person in real-time to reduce the number of the required surveillance cameras. With the conventional surveillance system which uses one-directional cameras, a number of one-directional cameras and their recording devices must be installed to cover wide area. This system leads to a complicated camera network system to monitor the images by operators and to record for a long time.

In this paper, we introduce a new surveillance system which is based on a combination of omni-directional camera [1] and networked Pan-Tilt-Zoom camera, so called PTZ camera. Using this system, the moving objects in wide area with 360 degree can be detected by the omni-directional camera and the coordinates of the position of the moving object is extracted. Then the relative pan and tilt angle for the PTZ camera can be calculated from the current position of the extracted the position of the moving object and the PTZ camera cane automatically controlled by properly zooming to follow and identify the object operation.

M. Takizawa, L. Barolli, and T. Enokido (Eds.): NBiS 2008, LNCS 5186, pp. 313–322, 2008.
© Springer-Verlag Berlin Heidelberg 2008

2 System Configuration

In our system, we call a combined system of an omni-directional camera and a PTZ camera and multiple directional microphones as Telegnosis system. In public, many different omni-directional cameras with various lens size from 1/3 to 3 inches in diameter, interfaces such as IEEE1394, USB2.0, camera link can be used depending on the applications, purpose and applied environments. Also the omni-directional camera can be reversely installed at upside down depending on the physical condition.

The system configuration is as shown in Figure 1 and consisted of multiple number of camera units TGs1~TGsn with camera server, omni-directional camera and PTZ camera and client TGc1. The omni-directional camera with PAL lenze is attached to DV, HDV or C-mounted USB camera and connected to the camera server through IEEE1394 or USB interfaces. On the other hand, the video output signal NTSC of the PTZ camera is converted by A/D converter to DV format and connected to the camera server through IEEE1394.

On the camera server, moving object detection function by which the location and the size of the moving object can be detected from the image and automatic motion tracking function by which the PTZ camera automatically can track to the detected moving object can be performed. The video images captured from the PTZ are sent to the client as a live video stream. At the same time, the captured images are automatically recorded and replayed on the client. The client can also interactively access to the server to replay, stop and forward the captured video images.

Furthermore, by introducing multiple numbers of camera units and servers, the whole captured images from various places can be received on the client at the same time and can be observed on the same display monitor or on the individual monitors on the client as a surveillance system. It is also possible to automatically display the particular video portion only when the moving detection events are generated.

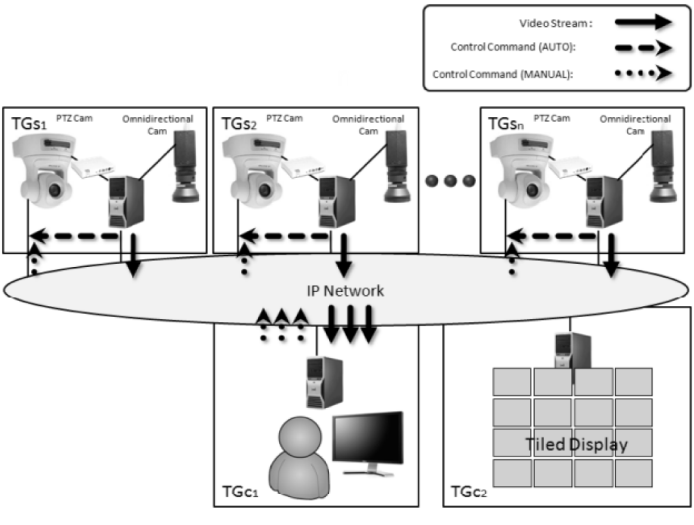


Fig. 1. System Configuration

On the other hand, the captured original omni-directional images can be directly sent and converted to the panorama images. By selecting the any positions on the panorama images and controlling the PTZ camera, the images from the PTZ camera equivalent to the selected positions can be displayed. In this case the automatic image tracking function is temporally released. Thus, the format conversion processing of omni-directional images to panorama images can be selectively performed based on CPU loads of the server or the client. Furthermore, the received video images from multiple locations can be output on the tiled display at the same time to attain the higher resolution and presence.

3 System Architecture

In the proposed system, three middleware functions are developed as function libraries. First, Midfield [2] is developed to transmit the omni-directional and PTZ vide images on IP network, record into files and control remotely the video stream. Second, omni-directional middleware is developed to convert the omni-directional images to the equivalent panorama images. Third, PTZ middleware is developed to control the PTZ camera images. We call those system functions as Telegnosis system.

The system architecture of the proposed Telegnosis system is shown in Fig. 2 and constructed as a middleware on top of Midfield System which will be explained in session 3.1. This middleware includes two layers and three planes including interface layer, multimedia control layer and view control plain, event process plane and system management plane respectively.

In interface layer, the information with user's interactive operations from user to remote site is processed and transmitted to the under layer. The whole system state, multimedia stream state, development process state from the original ring image to

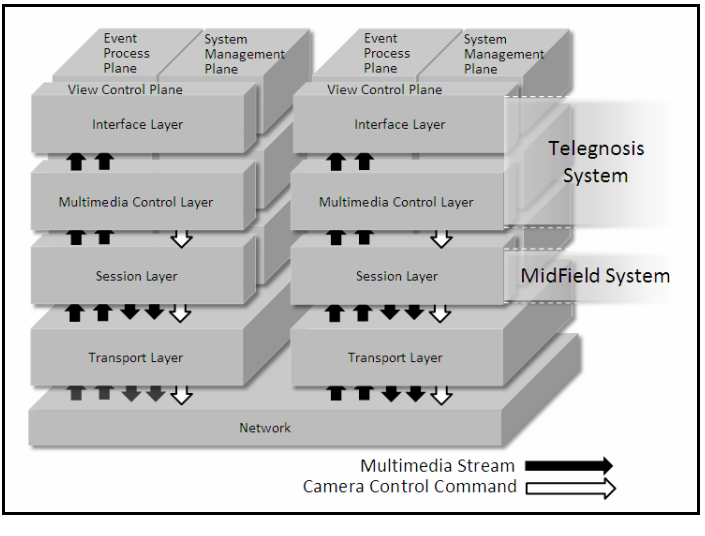


Fig. 2. System Architecture

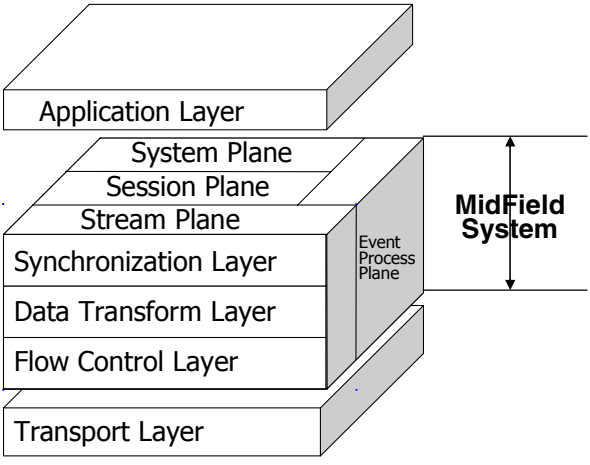


Fig. 3. MidField System

panorama image control state of PTZ camera and it's a set of static and dynamic parameters are managed in this interface layer. In multimedia control layer, actual development process from the omni-directional camera to panorama image, the computation of voice direction and PTZ camera control process by user's operations and video stream creation and control functions are performed. In event processing plane, various events generated during a video session are detected and processed. In view control plane, the PTZ camera state, such as vertical and horizontal angles and zooming positions by pan, tilt and zoom operations is controlled and managed. In the system management plane, the session of teleconferencing by video and audio is maintained.

3.1 MidField System

As shown in Figure. 3, the MidField [3] system is located between the Telegnosis system and transport layer as a session layer. Midfield offers multimedia communication functions to the application layer. Furthermore, Midfield is divided into three layers and two planes. Stream Plane is constructed by synchronization, data transform and media flow control layer, and performs multimedia stream processing. Session Plane performs management of communication sessions. System Plane monitors network traffic and CPU rate in the local host, and performs admission tests for QoS requirements from system user. Event Process Plane processes various events that are created in the system. The system is able to construct intercommunication environment on computer networks dynamically according to the environment of users and QoS requirements from users.

3.2 Omni-Directional Video Middleware

Omni-directional video middleware [3] performs the functions of Telegnosis system, including both interface and multimedia control layers and processes real-time image

Fig. 4a. Omni-Directional Ringed Image

Fig. 4b. Omni-Directional Panorama Image

development from the original ringed image as shown in Figure 4a from omni-directional camera to panorama image as shown in Figure 4b.

Any pixel formats and various video interfaces including IEEE1394, USB 2.0 and camera link interfaces can be available according to applications and its environments. However, when the higher pixel resolution video format such as HDV with 1920x1080 and more is used, the processing load of the whole middleware system increases, the original frame rate cannot be maintained, and packet loss on the video stream and block noise occurs, eventually the QoS of the video stream may decrease. In order to avoid this problem, the video frame rate is controlled by sub-sampling the frame according to the CPU load of the sending host or the receiving host in system management plane.

4 Automatic Tracking Function

First, the captured ringed images from omni-directional camera are converted to the equivalent panorama images. Then the moving object detection in the panorama images is examined. When the motion object is detected, the equivalent coordinates of top, left, right and bottom of the moving area are founded and the central coordinate of the moving area is calculated. Then the central coordinate for panorama camera system is converted to the PTZ camera system using the position parameters of both omni-directional camera and the PTZ camera which are calculated in advance. The calculated coordinate values is sent to the PTZ camera control middleware and the position of the PTZ camera can be precisely controlled by pan and tilt functions.

4.1 Moving Object Detection Function

In order to extract the moving object from the video images, several method including background difference method, inter-frame difference method pixel density distribution method. Those methods include both advantage and disadvantage and there is no the best method. In our system, from computational simplicity, realtime processing and easy implementation points of view, we applied background difference method. The omni-directional ringed images are converted to the panorama images to extract the moving object. The moving object detection processing is developed as motion detection filter by directShow filter implemented on Microsoft Windows O.S. environment to process realtime video stream [4]. By imbedding this motion detection filter to the omni-directional image middleware, the center of the position of the moving area can be obtained as a pixel position of moving object.

On motion detection filter using background image model [5], moving pixels are extracted by subtracting the background image from the current frame. Since the background model images are influenced by fluctuation of brightness surround environment, the pixel values on past several frames are averaged and updated on every frame. As the number of the average frames are increased, the accuracy of the moving object extraction increases but waste computing resources. In our system, the average interval is controlled by moving rate of object to save computing power on the camera server. The moving pixels are decided based on the predetermined threshold which is adaptively applied depending on the size of the varied area and its deviation when the background difference computation is carried out.

When the moving pixels are extracted, the four corners, top, right, bottom and the central position of the moving area are found as shown in Fig. 5. From this central position of moving object, the pan and tilt angle are determined and zooming rate is determined by the width of the moving pixel area. By introducing multiple reference lines by multiple vertical and horizontal lines to interace whole panorama images and dividing the moving pixel area into the sub areas, multiple moving objects can be extracted.

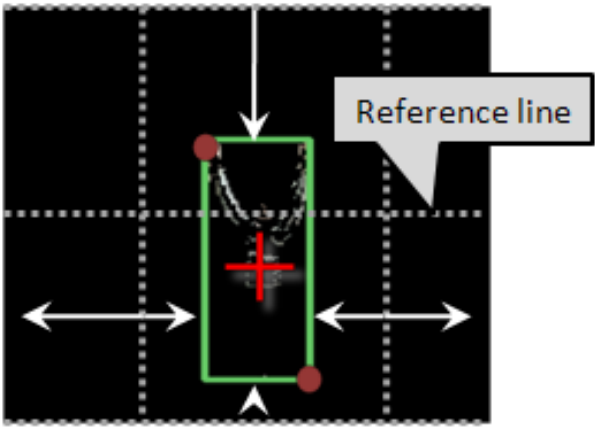


Fig. 5. Moving Pixels Area Extraction

4.2 Coordinate Conversion

When the PTZ camera is controlled for automatic tracking mode or manual operation mode, the coordinate of any pixel location on the panorama must be converted to the equivalent pan and tilt angle as shown in Fig. 6. The (Px, Py) on panorama pixel image is the pixel coordinate for automatic tracking mode and display window coordinate and the equivalent PTZ camera angle is calculated in the following equation [6][7].

$$PTZ\theta x = 360 \times Px/W + Ex \tag{1}$$

$$PTZ\theta y = omniVRange \times Py/H + Ey \tag{2}$$

$$\text{omniVRange} = |\text{omniTop}| + |\text{omniBottom}| \tag{3}$$

Where Ex and Ey are the gaps in the vertical and horizontal directions between the omni-direction camera and the PTZ camera and adjusted after installing at the place and omniTop and omniBottom are upper and lower limits of PTZ camera and H and W are height and width of the panorama image.

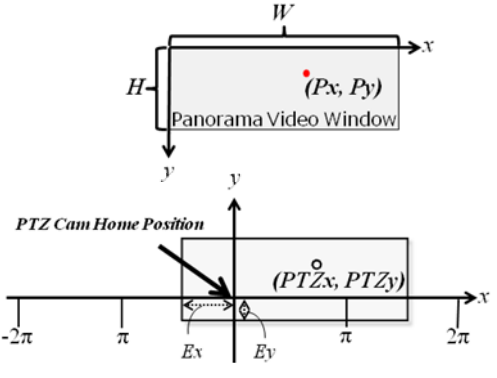


Fig. 6. Panorama Image and PTZ Camera Angle

5 Performance Evaluation

In order to evaluate performance of motion detection and tracking video methods by the proposed system, field experimentation was examined as shown in Fig. 7. In this experiment, both omni-directional camera and PTZ camera are installed at the ceiling, vertically 4.5m high from the floor. Horizontally the original point is set at the camera position. A testee walks at constant speeds from 20 m way from the original point. Then the motion detection start positions and frame out positions of the testee from PTZ camera were observed. Four different walking speeds were examined 5 times for each speed. Here, tracking start position means that motion detection starts to extract a moving object by omni-directional camera. The tracking limit means that the tracking speed of the PTZ camera exceeded its limited speed because of increase of the relative speed of moving object.

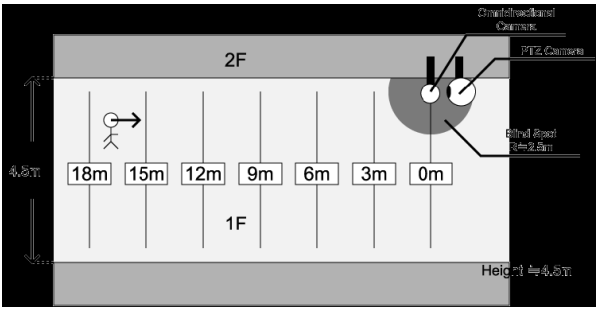


Fig. 7. Field Experimentation

When the walking speed is 1 km/h, the moving object is hard to be detected because the pixel value deviation was under threshold value. Tracking detection function started at 6m and continued to the origin. On the other hand, when the walking speed is 7.2 Km, the moving object could be detected at 10 m but the ZTP camera could not be enough to track the moving object because of the pan and tilt speed limitation. On the other hand, when the walking speed is 3.6km/h and 5.4km/h which are average child and adult speed, sufficient motion tracking by PTZ camera could be attained.

6 Conclusions

In this paper, we introduced a new surveillance system which is combination of omni-directional camera and PTZ control camera. While the omni-directional camera can detect the moving object in wide area with 360 degree, the PTZ camera can automatically track to moving object with proper high resolution. Then we explained

Fig. 8. Automatic Tacking Image Function

Table 1. Experimental Result

Walking Speed	Tracking Start Position		Tracking limit Position
1	1km/h	6m	0m
2	3.6km/h	10m	1m
3	5.4km/h	12m	3m
4	7.2km/h	10m	9m

its system configuration, architecture, motion detection and extraction and tracking methods. In order to verify our suggested system, a prototype system was constructed and evaluated its function and performance. Through the performance evaluation, human motion with ordinal walking speed could be detected correctly. However, the motion detection for the human walking with slow speed could not be possible. This is because the improper threshold value to reduce the error due to the noise for omni-directional camera image. Thus, the the filtering function to reduce the error by omni-directional camera must be improved. Also the PTZ camera cannot track to the high speed moving object because of its response time limitation. This problem will be resolved by introducing multiple PTZ cameras and dividing the whole space into the sub domains to assigning those domains to the each camera.

On the other hand, the screen shot images of indoor using the proposed surveillance system are shown in Fig. 8 and 9. Normally the camera captures the image of entrance of the room and can detect the person coming into the room and track to the just bottom of the PTZ camera. Thus, the detection of the human movement could be possible when he is moving around the camera and the automatic tracking function could be effectively realized.

As future works, detection of multiple moving objects on the real street and motion tracking by multiple PTZ cameras will be developed.

References

[1] Miata, Y., Hashimoto, K., Shibata, Y.: A New TV Conference system with Flexible Middleware for Omni-directional Camera. In: 8th International Workshop on Network-Based Information System(NBiS 2005), pp. 84–88 (August 22, 2005)

[2] Hashimoto, K., Shibata, Y.: Design of A Middleware System for Flexible Intercommunication Environment. In: IEEE Proc. on Advanced Information Networking and Applications, March 2003, pp. 59–64 (2003)

[3] Kobayashi, R., Maita, Y., Hashimoto, K., Shibata, Y.: Remote Healthcare Education Support System by Combination of Different Video Stream. In: Proc. of the 68th IPSJ Annual Conference, 6T-10, March 2006, pp. 259–260 (2006)

[4] Kubota, Y.: Digital Video Dokuhon. Ohme Co. (1995)

[5] Canon Co., Ltd., Sharp Co., Ltd., Sony Co., Ltd., Japan Victor Co., Ltd., (in Japanese) (2003), http://web.canon.jp/pressrelease/2003/hdv.html
[6] Cutler, R., Rui, Y., Gupta, A., Cadiz, J.J.: Distributed Meetings: A Meeting Capture and Broadcasting System. In: ACM Multimedia 2002, December 2002, pp. 1–6 (2002)
[7] Yagi, Y., Yokoya, N.: Omni-directional Vision: Sensors. Journal of Information Processing Society

Author Index