GEORGE ROUS

Computer Communications and Networks

The **Computer Communications and Networks** series is a range of textbooks, monographs and handbooks. It sets out to provide students, researchers and non-specialists alike with a sure grounding in current knowledge, together with comprehensible access to the latest developments in computer communications and networking.

Emphasis is placed on clear and explanatory styles that support a tutorial approach, so that even the most complex of topics is presented in a lucid and intelligible manner.

Also in this series:

An Information Security Handbook
John M.D. Hunter
978-1-85233-180-1

Multimedia Internet Broadcasting: Quality, Technology and Interface
Andy Sloane and Dave Lawrence (Eds)
978-1-85233-283-9

UMTS: Origins, Architecture and the Standard
Pierre Lescuyer (Translation Editor: Frank Bott)
978-1-85233-676-9

Designing Software for the Mobile Context: A Practitioner?s Guide
Roman Longoria
978-1-85233-785-8

OSS for Telecom Networks
Kundan Misra
978-1-85233-808-4

From P2P to Web Services and Grids: Peers in a Client/Server World
Ian J. Taylor
978-1-85233-869-5

The Quintessential PIC? Microcontroller 2nd edition
Sid Katzen
978-1-85233-942-5

Ubiquitous and Pervasive Commerce
George Roussos (Ed.)
978-1-84628-035-1

Intelligent Spaces: The Application of Pervasive ICT
Alan Steventon and Steve Wright (Eds)
978-1-84628-002-3

Information Assurance: Security in the Information Environment 2nd edition
Andrew Blyth and Gerald L. Kovacich
978-1-84628-266-9

Peer-to-Peer Computing: Building Supercomputers with Web Technologies
Alfred W.-S. Loo
978-1-84628-381-9

George Roussos

Networked RFID

Systems, Software and Services

Springer

George Roussos, PhD
Birkbeck College, University of London, UK

Series Editor
Professor A.J. Sammes, BSc, MPhil, PhD, FBCS, CEng
CISM Group, Cranfield University,
RMCS, Shrivenham, Swindon SN6 8LA, UK

CCN Series ISSN 1617-7975
ISBN 978-1-84800-152-7 e-ISBN 978-1-84800-153-4
DOI 10.1007/978-1-00084800-153-4

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

Library of Congress Control Number: 2008924842

Printed on acid-free paper

9 8 7 6 5 4 3 2 1

springer.com

To my daughter, Katerina

Preface

The birth of RFID technology is credited to the 1948 research paper by H. Stockman on "Communication by Means of Reflected Power" [109]. After describing the main principle of communication by reflection and reporting his experiments, he concluded that

> "Evidently, considerable research and development work has to be done before the remaining basic problems in reflected-power communication are solved, and before the field of useful applications is explored."

It turns out that in fact this work required almost 60 years of science and engineering before it was mature enough to find its way into large-scale applications. This is surprising, as at first glance RFID is a simple technology that is straightforward to implement. And yet, its dull appearance hides great complexity: RFID involves several engineering disciplines, including systems, networking and software development, antenna design, radio propagation, integrated circuit techniques increasingly focused on printed electronics, receiver design, encryption and security protocols, and materials technology, to mention but a few.

In this book, I attempt to present the main ingredients for building complete network RFID systems, written for a knowledgable computing audience. I hope this book will help practitioners and researchers alike in establishing a robust framework for thinking about RFID. I also hope that the book will help anyone involved with RFID technology to make informed decisions that are appropriate for their particular problem at hand.

In many ways, I have tried to write the book I wish had been available when in 1999, together with a group of enthusiastic colleagues, we set out to work on the MyGrocer project, a second-generation pervasive retail system. At the time, information about RFID was fragmented and written primarily for a different audience altogether. Building large-scale open software systems for RFID had not been attempted to any significant extent, and experience was

sparse and mostly unpublished. Despite this, MyGrocer succeeded in demonstrating item-level tagging within a real supermarket environment in 2002. To my knowledge, this is the first demonstration of the technology outside the laboratory and thus marks a milestone for RFID.

Nevertheless, recent years there has been an explosion of interest in RFID, and many new technologies have made obsolete the state-of-the-art of 2000. The reason for this is twofold. First is the availability of very low-cost long-range passive RFID tags that require no battery to operate, and second is the wide availability of fixed and wireless communication networks that allow RFID deployments in the field to be linked with software and information services at the network core. These facilities have made possible large-scale commercial applications of RFID in the supply chain, ticketing, asset tracking, maintenance, retail, and personal identification.

Thanks to these applications, RFID today is one of the most popular computing platforms in use. According to some estimates, there are more than 3.7 billion RFID tags in use today, of which more than 1.6 billion were deployed during 2006—and this trend is accelerating. This popularity of RFID permits considerable cost reductions that have sparked further interest in the technology, as it offers unique advantages in instrumenting the physical environment.

Yet the same features that make RFID such a popular technology are also complicating its use. To provide battery-free operation and low cost, passive RFID tags have extremely limited capabilities, often being able to hold (and in fewer cases protect) only a simple entity identifier, which is employed as a means to automatically link physical entities and their stored information. This implies that the majority of system functionality must be supported by the network, for example by mapping an object ID to an entity description and attributes. Note that this intimate linking effected by RFID between real and virtual also creates considerable security and privacy risks that have to be managed so as to guarantee its safe use.

This book is written for readers with a firm grasp of computer fundamentals and experienced in software development. Unlike other books on RFID, it does not make any assumptions about knowledge of other aspects of engineering but only includes appropriate information so the reader can understand common trade-offs and the performance characteristics of different RFID flavors. My guiding principle in selecting material to include in this book is what I would have found useful to understand RFID when I first started working with the technology and exclude all the details that have not been of any practical use.

Contrary to custom, I begin with a quick introduction to RFID basics rather than applications. The reason is that a discussion of applications without a firm grasp of the basics quickly becomes unwieldy and obscures rather than clarifies the issues. My approach may appear somewhat abrupt at first but will pay dividends later in the book. After sorting out the basics, I proceed to discuss three RFID applications that I find of particular interest, especially

in the context of networked systems, namely e-passports, ticketing, and supply chain management. The second chapter concludes with a quick overview of other common applications and some discussion of the numerous standards for RFID.

Chapter 3 discusses in detail the different types of readers and tags and makes particular reference to common technologies used in real systems. It is highly likely that you have already used one or more of these technologies, even though it may not have been clear that this was so, and this chapter also works as a guide to spotting RFID in everyday situations. Chapter 4 takes a step backward and looks at radio waves, how they propagate in space, and how they can be captured by tags. The main reason for covering this material is that it justifies the very different performance characteristics of different RFID systems. Moreover, looking at the fundamentals is rather informative in understanding the design of RFID antennas, which have been considered the "black art" of RFID. This is certainly no longer so, and today the system architect has a wide variety of high-performance commodity designs to choose from.

Chapter 5 is by far the most tedious of this book, as I go into the details of different identifier systems, how they encode unique serial codes for tagged entities, and how codes are assigned to particular organizations. Although this is a rather esoteric issue, it is nonetheless necessary to have an in-depth understanding of such schemes when working with RFID, as this is the only way to interpret codes. To brighten up the discussion, I also include a bit of history on how we have arrived at the current situation.

Chapters 6 and 7 are intimately related. In the former, I present what I call the RFID stack; that is, a layered description of the structure and the functionality of modern RFID systems. Due to the fact that large-scale RFID is relatively recent, each system and vendor uses different terms and different subsystem boundaries to describe what are essentially identical components. The stack is my attempt to organize all such elements into a unified system architecture that provides a blueprint shared between all systems and platforms. I also outline how the systems of the major RFID vendors fit into this view. The next chapter complements the description of the stack by describing RFID middleware, which provides the glue that makes such unification possible. Note that middleware is without doubt the hottest area of RFID software, and for this reason I also include some material on forthcoming developments that may not currently be part of commercial systems or standards.

Chapter 8 discusses the different types of network services required for RFID support. In particular, I highlight the main elements of code resolution solutions and their limitations and the emerging class of repository services. I review several systems that provide solutions to this problem with a view toward providing common patterns of RFID systems and software that can be modified and reused as appropriate to fit the requirements of diverse domains and applications. In the short term, such services are unlikely to become as pervasive or ubiquitous as the common network services that form the

internet infrastructure. Before this happens on any significant scale, many data management and trusted operation issues would have to be resolved.

Chapter 9 deals with the issues that have given RFID its current notoriety, namely privacy and security. I use several cases where RFID has been used in a haphazard way to motivate the discussion of specific technologies and privacy protection approaches. Nevertheless, these issues have less to do with technology and everything to do with the choices we collectively make as a society and codify in law and culture, and RFID places considerable strain on both. Of course, our business is technology, but despite what many may claim, technology also shares the same context as the rest of society and has to be guided by ethics as much as curiosity about what is technically possible.

The final chapter of this book takes a look to the future. I start with a discussion of several improvements to RFID already in the pipeline that hold the promise of even lower-cost tags and extended capabilities. Many of these developments are due to mature in the relative short term and offer exciting enhancements to current RFID functionalities. Lowering the cost of the individual tag may also allow widespread item-level tagging of consumer products. This development will directly affect the end user—that is, the consumer—and it can potentially bring about fundamental changes in the shopping experience.

I discuss these changes and the potential opportunities and risks they may cause using the findings of the MyGrocer study, which outlined the main issues and explored specific consumer concerns. Moreover, I set the discussion of RFID in the wider context of pervasive computing, the next-generation computing paradigm that is gradually emerging from research, and link one to the other, identifying applications of particular promise. Pervasive computing is my own research specialization, and RFID provides a low-cost entry into a world where the physical and the digital are no longer separate.

Throughout the book, I have attempted to provide as many examples based on real case studies as possible and follow them up with detailed explanations and photographs that hopefully help to clarify the issues. My intention has been to provide narratives around my experiences with RFID and the experience that others have shared with me, rather than discuss the minutia of standards and specific programming interfaces. The emphasis is on developing an arsenal of techniques and designs that can be mixed and matched to fit the needs of new systems and applications.

Before concluding, I would like to express my gratitude to all those who have worked with me on RFID over the years as advisors, clients, colleagues or students. First, I thanks my companions in the MyGrocer experiment, in particular Panos Kourouthanassis and Juha Tuominen, both of whom played a critical role in the success of that project. Many thanks go to Roger Till and other staff at GS1 UK for valuable advice and guidance. Thanks are also due to the following persons for discussions, advice, or both: Alessandro Acquisti, Christof Bornhövd, Sastry Duri, Simson Garfinkel, Anatole Gershman, Vassilis Kostakos, Matthias Lampe, Olli Pitkänen, Tony Salvador,

Christian Tellkamp, and the participants of the Mobicom 2006 tutorial on network RFID.

Many thanks are also due to my students reading advanced information systems at Birkbeck College, who were a receptive audience for the early ideas that eventually became this book. Thanks also go to my research students Dikaios Papadogkonas, Jenson Taylor, Michael Zoumboulakis, and Dima Diall for various experiments and prototyping.

Last, but not least, my thanks go to my wife, Theano, and my daughter, Katerina, who gave up their time with me so that I could work on this book.

Bloomsbury and Lefkada
Summer–Autumn 2007

George Roussos
University of London

Contents

1

What Is RFID

Radio frequency identification, or RFID as it is commonly known, is an umbrella term that refers to several information and communication technologies that share the capability to automatically identify objects, locations, and individuals to computing systems without any need for manual intervention. The RF part in RFID in particular points to the fact that this capability is enabled by wireless communication between the computing system and the item identified. This modus operandi also implies that RFID operation is imperceptible to the human senses, a fact that has considerable implications for users of RFID technology in terms of security and privacy protection.

With RFID, automatic identification takes a very specific form: the object, location, or individual is marked with a unique identifier code contained within an RFID tag, which is in some way attached to or embedded in the target. In turn, a computing device that needs to identify the target employs an RFID reader to search for tags. When it receives an indication that a tag is present in its vicinity, it instructs the reader to request the code. The retrieved data are then recorded or otherwise processed in whatever way is suitable for the particular application.

There are many variations in the details of this process, which depend on the features of the specific flavor of RFID technology used. Notably, higher-capacity tags can often hold extra data, depending on the application, in addition to the identifier code. In Chapter 3, we will discuss some of the alternative technology choices and the trade-offs involved. Selecting the right type of RFID technology to match the requirements of a particular application is a critical decision for a successful implementation.

Nevertheless, all RFID systems have two features in common:

- Electricity required for the operation of the tag is transmitted by the reader wirelessly. In many cases, this would be the only power source for the tag.
- The tag employs a distinctive approach to communicate with the reader in that it modifies the reader's transmission to send out information rather than generate its own signal.

G. Roussos, *Networked RFID*, DOI: 10.1007/978-1-84800-153-4_1,

These two features set RFID apart from other wireless communication technologies and are responsible for its unique advantages but also for its significant limitations. In the remainder of this chapter, we take a closer look at the RFID auto-identification process and examine the main ingredients of its operation.

1.1 Automatic Identification with RFID

Deconstructing the mystery of RFID, automatic identification in practice can be the simple process of retrieving an identifier stored in a tag. Indeed, the basics of RFID are relatively simple and easy to understand. Reading a tag is a straightforward task in two steps:

1. transmit adequate energy to power up the tag, and
2. communicate with the tag to request and receive the identifier.

This simple process is depicted in Figure 1.1, which shows the main components of a basic RFID system and the succession of events as executed in time. Later in this chapter, we will look in turn at each of the individual steps of the process, discuss the different options for their implementation, and identify how distinct choices define the different flavors of RFID in common use. For now, we will keep the descriptions at a relatively high level, as the intention is to clarify the process and the parameters that influence its performance rather than provide minute detail on how each step is carried out. Details and complete examples will follow in Chapters 3 and 4.

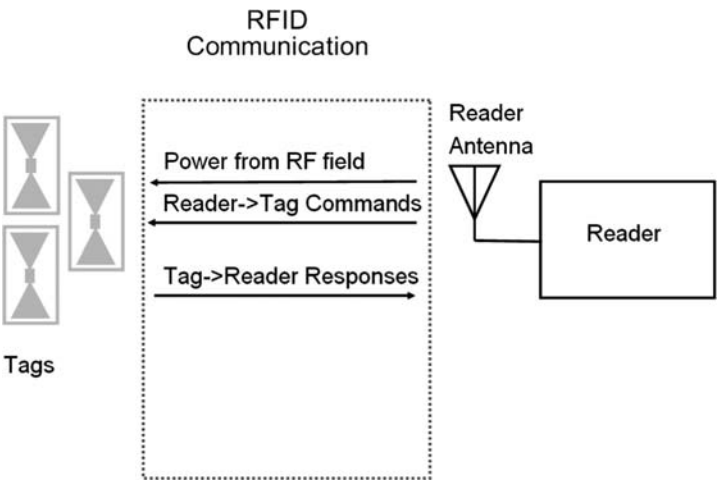

Fig. 1.1. A high-level view of the interaction between reader and tag in the RFID auto-identification process.

Compared with other wireless systems, RFID is characterized by the fact that communication is asymmetric with one peer, the *reader* or *interrogator*, taking on the role of the transmitter and the other, the *tag*, the role of the responder. Instead of creating its own transmission, the tag modulates or reflects the electromagnetic waves emitted by the reader to communicate. To some extent, this is the reason for the success of RFID: this technique allows a somewhat complex reader to be used with a very simple tag of small size, which can be built at low cost. A small number of fixed or mobile readers can be used with numerous tags to construct very large systems at relatively low cost. For example, in Chapter 2 we discuss the case of a metropolitan ticketing system that has deployed a few thousand readers to support several million tags.

Of course, the auto-identification task can be complicated by a challenging use context. For instance, dealing with high tag densities requires singulation and prioritization protocols to coordinate their accurate and consistent reading. In other cases, complexity may be due to elaborate schemes required for the encoding of identifier semantics as appropriate for specific applications, which may subsequently lead to dependence on associated network-based services that support code resolution and provide additional meta-data related to the tagged entity.

To be sure, building large-scale open systems with RFID is a challenging task that also requires support by dedicated network services and specialized system infrastructures and software. It is the task of this book to identify common patterns of use and robust designs to assist in this task. Nevertheless, for the time being we concentrate on the basics of RFID and leave such complications for subsequent chapters.

1.2 Energy Transmission

The first step of the identification process is the transfer of energy from the reader to the tag. This is achieved by means of capacitive or inductive coupling, whereby a change in current flow through the reader antenna induces current flow in the tag. Such coupling between electrical components and electronic circuits is a common phenomenon that is encountered in many everyday situations. For example, magnetic (or inductive) coupling is the principle of operation for electricity transformers and also the way electric toothbrushes are recharged. In this case, transfer of energy is through a shared magnetic field, and it is only the magnetic component of the electromagnetic wave transmitted by the reader that is involved.

Capacitive (or electrostatic) coupling, on the other hand, uses only the electric component of the electromagnetic field generated by the reader antenna. Such coupling is equally common in practical situations but is conspicuous primarily due to its negative effects, for example in causing interference in TV and radio receivers or between cables located adjacent to each other.

In either case, the source of the energy is the electromagnetic field created by the antenna of the reader when it transmits. When a tag is placed within this field, electric current is induced on the tag through its antenna and can be used to power up the chip. In many practical situations, the position of the tag within the electromagnetic field generated by the reader will change, and as a result the current induced on the tag antenna will vary. To deal with such fluctuations, most RFID tags use a capacitor to regulate the flow of current and provide a relatively uniform and thus usable supply to the chip.

Inductive coupling

Systems that employ magnetic coupling have special limitations. First, the magnetic field generated by the reader fades very quickly away from the antenna and, as a result, the range of a system developed with this technology is either very short or has to employ antennas of relatively large size to be capable of reading tags that are farther away. In practice, systems that use magnetic coupling would be used for applications that do not require an operating range of more than one meter or so, and in most cases the read range would be a few dozen centimeters.

The relatively small area that a reader can cover also means that if a tag is moving quickly, for example if it is carried in the pocket of a person walking quickly, it is again likely that it will not be readable, as there will not be enough time for the tag to harvest adequate power for its operation. Finally, due to the limited range of the system, it is not feasible to use it with a large number of tags, as the targets that carry them would quickly fill up the available space.

Systems that use inductive coupling can be easily recognized by the characteristic form of their antennas which are shaped in a coil (see Figure 1.2). We look at a system that uses magnetic coupling in more detail in Section 3.2.2.

Capacitive coupling

Systems that employ capacitive coupling are not similarly constrained but have other limitations. Due to practical and regulatory reasons, such tags operate at specific frequency ranges and as a result must be relatively large in size due to the magnitude of their antennas. Moreover, they are particularly sensitive to certain types of environments; for example, their operation is severely constrained by the presence of water in the environment.

Systems that use capacitive coupling can often be recognized by the form of the tag antennas, which are shaped in some form of a dipole (see Figure 1.3). Readers in such systems often employ patch antennas which are commonly placed within an enclosure, although other types are used for extended range. We look at a system that uses capacitive coupling in more detail in Section 3.2.4.

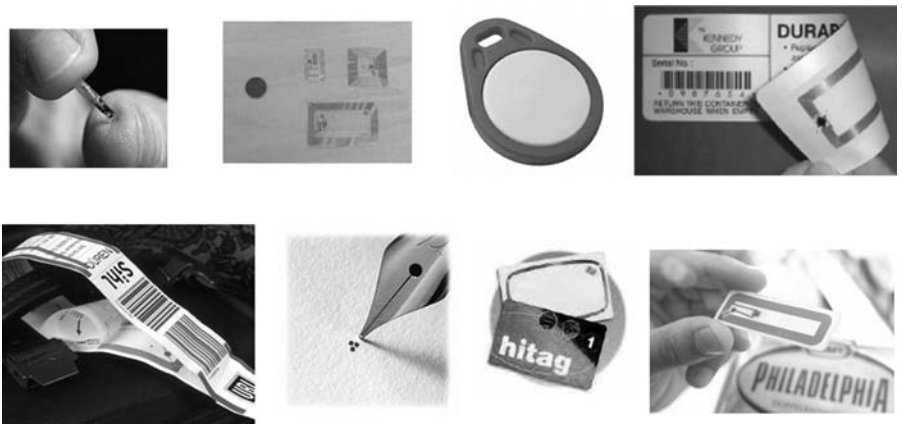

Fig. 1.2. A collection of RFID tags using inductive coupling in different form factors and packagings.

Passive and active tags

When tags carry no battery at all, the system is referred to as *passive* RFID and wholly depends on the reader for its energy supply. Other types of RFID employ a battery in the tag, in which case they are referred to as *active* or semi-active RFID tags. Semi-active tags only use their battery to provide power to the tag chip. This can be in addition to the power harvested through the antenna. Active tags use their battery also to transmit, which means that they provide a much more robust communication channel.

Active tags have considerable advantages over passive ones, which draw all their power from the reader signal, as they can transmit at higher power levels and thus have longer range and support more reliable communication. Moreover, active tags can operate in particularly challenging environments such as locations with significant radio frequency pollution caused by electric machinery. It is also relatively straightforward to augment active tags with additional sensing capability, for example temperature sensors, and they can initiate transmissions rather than simply respond to readers, but they stop operating when their battery expires.

Despite the advantages of active tags, the current interest in RFID is solely due to passive tags, which do not depend on batteries and thus do not require recharging or replacement. This is clearly a unique advantage, especially for maintenance of large-scale systems. In fact, active RFID is effectively just one of an increasing number of wireless local area communication technologies, and as such it is of limited interest for the applications considered here. In this book, we do not consider active or semi-active RFID further, although many of the discussions herein apply. From now on, we refer to passive RFID simply as RFID without further qualification.

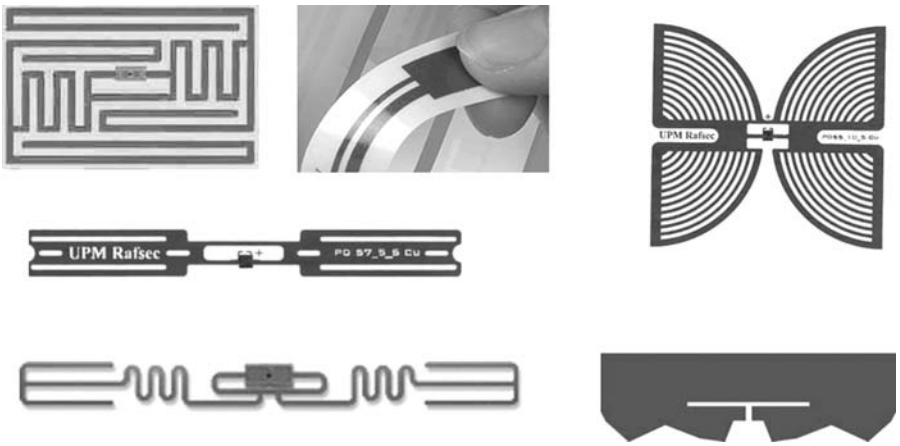

Fig. 1.3. A collection of RFID tags using capacitive coupling displaying a large variety of antenna designs.

1.3 Communication

At the end of the first step of the RFID process, the tag is charged and ready to communicate with the reader. As already noted, this communication is asymmetric in that the energy of the electromagnetic field transmitted by the reader is the sole source of power used for data transmission by either device. In both inductive and capacitive coupling systems, the tag modifies the electromagnetic field in such way that changes can be observed by the reader, which decodes and interprets them as data. However, the details of how this modification is carried out differ significantly between the two systems, which adopt distinct techniques.

Magnetically coupled tags communicate by changing the load resistance connected to their antenna in a process appropriately called load modulation (see Figure 1.4). By doing so, they increase or decrease energy consumption of the field generated by the reader antenna and effectively control the modulation of the radio signal. These changes can be detected by the reader by examining changes in the potential variation in its resistance and can be decoded. Because the magnetic field decays very rapidly with distance from the center of the reader antenna, the changes to be detected by the reader are tiny compared with its own transmission. For this reason, the tag modulates the radio signal in such a way that it responds in a slightly shifted frequency from that of the reader, which are often referred to as the sub-carrier frequencies.

Capacitive coupled RFID systems using the electric component of the carrier wave operate using a backscattering technique (see Figure 1.5) rather than load modulation. This process is very similar to the operation of radar in that the tag reflects back a small part of the electromagnetic wave emitted by the reader. The reflection can be used to transmit information by

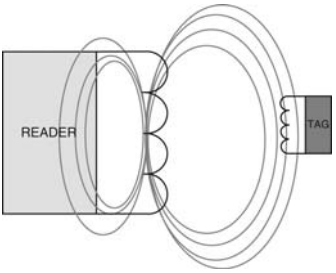

Fig. 1.4. In RFID systems that use inductive coupling, power transmission is via the magnetic component of the wave emitted by the reader antenna, and tag-to-reader communication is performed through load modulation.

examining the so-called reflection cross section (that is, the signature of the component of the wave that has been sent back to the reader) and comparing it with the original. In practice, data are encoded by the tag by turning on and off the load connected to its antenna and thus shifting the reflection cross section between two clearly identifiable characteristic signatures. Similarly to magnetically coupled RFID systems, communication by reflection leads to considerable loss of power and readers have to be highly sensitive to be able to receive messages from the tags.

1.4 A Very Brief History of RFID

Capacitive coupled tags use backscattering to communicate in a manner similar to radar. This is not coincidental, and indeed the beginnings of RFID technology can be traced back to work carried out during World War II on radar technology. In particular, an ancestor of modern RFID is the so-called Identify Friend or Foe (IFF) system, first introduced in WWII and still in use today to identify friendly or enemy aircraft. IFF, as used by the British Air Force at that time, was what we would describe today as a programmable tag that upon interrogation would return a code that identified the aircraft that carried it as friendly. There are several anecdotes regarding the early use of such systems and there is evidence that a similar scheme was in use by the Luftwaffe. Early IFF systems were further developed in the 1950s and nowadays are in common use in civil as well as military aviation.

But real progress with the technology in a form that we would recognize today as RFID was not made until the 1960s and 1970s, when the techniques became practical.[1] During this time, the majority of tags were very simple, often being able to encode a single bit of information to indicate presence

[1] For a fascinating discussion of RFID developments during this time, the reader should consult the works of Jeremy Landt, one of the pioneers of this era [73].

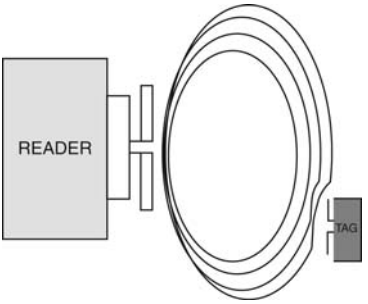

Fig. 1.5. In RFID systems that use capacitive coupling, power transmission is via the electric component of the wave emitted by the reader antenna, and tag to reader communication is through backscattering of the reader signal.

or absence. These simple devices that lacked any security provisions were primarily used for tracking nuclear materials and were the precursor of electronic article surveillance (EAS) systems, which are in common use today to prevent theft of various assets. This technology was primarily pioneered by Los Alamos National Laboratory and used by the US government.

Nevertheless, it was not until 1975 that the first truly passive tag using backscattering was developed, and soon after the technology was transferred to the commercial sector. At the time, the capabilities of systems were constrained by the size and simplicity of electronic components and an RFID tag circa 1980 would be dominated by the chip size, which could only hold a few bits of information. Such early technologies found applications in cattle tagging and railroad freight tracking, and active tags found applications in automated toll collection and keyless entry systems for cars and homes.

At the end of the 1980s, with the rapid miniaturization of electronics, which offered at the same time lower cost and higher performance and capacity, RFID technology became commonplace and found a variety of applications. One particular area of major growth has been the use of passive inductive tags to develop a variety of contactless smart cards which has found popular applications, especially in access control and ticketing.

At the beginning of the 2000s, RFID came into prominence due to its unique capability to automatically identify tagged entities at potentially very low cost. By this time, enterprise information systems had developed into the backbone of global trade, and supply chains that involve partnerships across the globe had reached relative maturity and could benefit considerably from the increased precision and lower cost compared with manual approaches of RFID. At the same time, the internet has been established as the primary infrastructure for the operational deployment of network services that could complement well the advantages of RFID.

Work in this spirit was led by the Auto-ID initiative, which more recently developed into the EPCglobal network, and initially aimed to provide simply

a more efficient means for data entry by extending the concept of the bar code to RFID. To a large extent, current interest in RFID technology is due to the evidence provided by this work of the feasibility of this technology to support large-scale open collaborative supply chains.

1.5 Summary

RFID is a wireless automatic identification technology that uses a tag embedded in the target entity to mark it with a unique code. Tags are read by a suitable reader, a device that provides power for the operation of the tag and for communication. There are many flavors of RFID, but a main distinction is between those that use the magnetic component and those that use the electric component of the wave emitted by the reader. Each alternative has particular characteristics with unique advantages and limitations. The technology has a history of over 70 years from the discovery of its principle of operation but has required many technological breakthroughs to become practical.

2

RFID Applications

To a large extent, the current notoriety of RFID is due to its rapidly increasing use in end-user applications and the unique features it they affords. To motivate our discussions of systems, software, and services in subsequent chapters, we begin our exploration of RFID technology by discussing three selected applications of special significance, namely e-passports, ticketing, and supply chain management. In all three cases, RFID has been deployed in large-scale systems supported by networked services of varying complexity and provides useful examples on which to model our discussions. We conclude this chapter by briefly outlining other applications of RFID as well as reporting on current standardization efforts, with particular reference to its role in application development.

2.1 ICAO e-Passports

In May 2004, the International Civil Aviation Organization (ICAO) approved the specification for the so-called machine-readable travel documents (MRTDs). MRTDs use standard RFID technology to store personal and biometric information on passports, visas, and travel cards. Their development is seen by ICAO and its member countries as a significant improvement over manual inspection of travel documents at border control points in terms of efficiency and data entry precision. Nevertheless, the rapid development of MRTD specifications and their quick adoption and implementation have been driven by the desire to increase the security of air travel, as evidenced by public statements by law enforcement officials on both sides of the Atlantic. By 2007, more than a hundred countries had issued or were in the process of issuing passports that conform to the MRTD specification. A smaller number have upgraded their border controls with scanning equipment for MRTDs.

G. Roussos, *Networked RFID*, DOI: 10.1007/978-1-84800-153-4_2,

RFID technology and MRTD

The ICAO provisions call for ISO 14443-compliant RFID tags (discussed in Section 3.2.2) embedded in travel documents to hold personal and biometric information on the traveler. RFID readers (see Figure 2.1) operated by immigration services interrogate and retrieve traveler information without the need for manual intervention. Due to the current capacity of tags, biometric data are restricted to photographs but the standard also provides specifications for iris scans and fingerprints for future use. Millions of e-passports are already in use, and thousands of MRTD-capable immigration control facilities have been deployed at disembarkation points in several countries. It is noteworthy that according to the specification MRTDs remain valid even if the embedded chip is damaged or unreadable for any reason.

To control access to the data stored in the MRTD, the standard recommends that information stored in the second line of the machine-readable zone (MRZ) of the document be used as the key for the reader to gain access to the RFID memory content. As a result, this key is made up of a combination of the passport number, its date of expiry and the date of birth of its holder, which is easily obtained. A long-term goal of ICAO seems to be the development of a supplementary public key infrastructure to allow MRTD inspecting authorities to verify the authenticity and integrity of the data stored in the document.

Fig. 2.1. A Greek e-passport and an associated reader. Note the mark at the bottom of the front cover of the passport indicating that it is MRTD compliant. The Greek MRTD system has been developed by Arcontia in collaboration with ACG.

The ICAO specifications also define the data that shall be stored in an MRTD:

- personal data that to a large extent duplicate what is currently displayed on the photo page of a typical passport already and

- biometric data that include a (low-resolution) photograph, textual descriptions of the characteristics of the holder, and provision for fingerprints and iris scans.

Though a lot of discussion has been devoted to the biometric information included in the e-passport, the reality is that current RFID technology employed in its implementation has insufficient capacity to hold this information. Tags available commercially in large quantities provide up to 4 Kbytes of storage which is not enough to hold even a single high-quality image that can provide adequate resolution for automatic face recognition. Fingerprints have similarly high storage requirements (even when compressed with sophisticated wavelet schemes), and this is also true for iris scans.

Also note that RFID in e-passports has further implications related to system maintenance. While passports are typically issued for a period of 5 to 10 years (for adults), RFID chips are under warranty for a mere 2-year period. This raises very considerable concerns for the long-term viability or indeed the value of this system and casts doubts on the prudence of the decision to invest considerable amounts on this technology.

MRTD security and privacy

While MRTDs certainly facilitate faster and more accurate data entry, they have justifiably received strong criticism regarding their security and privacy performance. Furthermore, upon closer examination, any claims that their deployment may improve the safety of air travel do not stand up to scrutiny and in some cases it may in fact reduce the effectiveness of border control due to the automation of the process.

Duplicating the RFID component of an MRTD (see Figure 2.2), including all the data held, is a fairly straightforward process that can be achieved at relatively low cost with commodity equipment. This was demonstrated at the Black Hat conference by Lukas Grunwald soon after e-passports became available, and today the passports of several countries have been attacked in this way. The simplicity of the duplication process is such that it makes copying the electronic component of an e-passport far easier than reproducing the printed elements, though modification of the passport data has not been demonstrated yet and it is certainly much harder, as their integrity is protected by the digital signature of the issuing authority.

Access to personal data stored in the chip is facilitated rather than hindered compared with paper only travel documents. Since MRTDs provide access to their data via a wireless channel, it can be easier to access such data as it is not necessary to gain physical access to the document itself and certainly not necessary to open it as required for visual inspection. It has been demonstrated that with specialized equipment it is possible to communicate with an e-passport over a distance of several meters and probe for access to the data. Even in cases where additional external shielding is used to protect

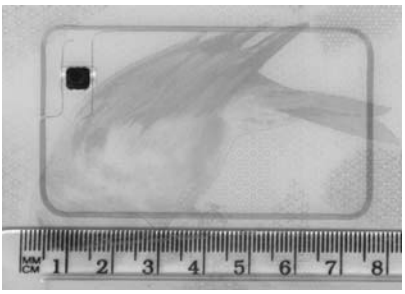

Fig. 2.2. RFID tag embedded in the UK e-passport.

against unauthorized remote access for example as implemented in the latest version of the USA e-passport, it has been shown that it is still possible to interrogate the RFID chip if the document is not firmly folded.

Finally, the ICAO specifications permit e-passports to only support what is referred to as passive authentication, which provides no access control mechanism to the RFID chip. Fortunately, most countries have opted for stronger protection under the provisions of the so-called basic access control (BAC) and active authentication mechanisms, although the implementation of either or both schemes is optional. Basic access control operates by establishing a secure channel between the reader that carries out the inspection and the e-passport and protects both the confidentiality and integrity of the transmission. To achieve this, BAC requires symmetric cryptography provisions, including the generation of encryption and authentication keys from passport information that is printed on the actual document (specifically passport code, expiration date, and birth date of the holder).

At border control points, inspecting officers would pass the MRTD presented by the traveler through an optical character recognition (OCR) system to retrieve the information used for the generation of the cryptographic keys. These are printed on the machine-readable strip according to a previous ICAO regulation. The actual cryptographic keys would be computed on the fly by an attached computer, which would subsequently communicate with the e-passport and retrieve the data, store it, and possibly display it on the operator screen. The whole process would last only a few seconds.

Yet recent work has highlighted the fact that due to regularities and practical constraints on the range of seed data, many if not all MRTD implementations would employ keys with low entropy and thus be relatively easy to break. In fact, practical attacks on e-passports on the general public, either via brute force or through eavesdropping in the exchanges directed by the inspection system, have been demonstrated using relatively inexpensive hardware in [17]. In addition to the obvious dangers associated with harvesting personal data in this manner, it is also feasible that the unique identifier established by the passport code could be used as the key for tracing applications, where

individuals could be followed from location to location and their movements recorded in some detail.

MRTDs and network infrastructure

E-passport applications have simple requirements both in terms of RFID tag access and supporting network services. At any one time, only a single tag is within the range of the reader and there is ample time to carry out the required data retrieval. Furthermore, the volume of collected data is relatively low since even at peak periods and for the largest operators of e-passport inspection systems, over a single day only a few hundred thousand recordings are made at most and normally data rates would be well below that. In any case, such inspection systems raise very restrictive interoperability requirements since they operate within a national scope and are commonly supported by proprietary database applications over secured private networks.

One area that could require some networking support is the implementation of active authentication and extended access control, although this is something to be considered in the future as far as the ICAO is concerned. Under these proposals, individual tags have unique keys that certify their authenticity and that can be checked against databases maintained by the issuing authorities. Moreover, rather than a cached copy of the public keys of all known issuers replicated at each station, a complete public key infrastructure (PKI) with key revocation provisions can be supported. Such a system would considerably improve security of the e-passport and privacy protection but at the same time would raise considerable additional issues related to its trusted operation.

Nevertheless, MRTDs cannot deliver their potential benefits—outside the narrow scope of data entry—unless they are supported by networked control points interlinked with law enforcement systems, as outlined by a recent report by the UK National Audit Office. In this report, NAO severely criticizes the decision to proceed with deployment of e-passports without provisions for enabling full network support, which they consider necessary to justify the investment in this technology.

2.2 Ticketing

One of the earliest and still fast growing applications of RFID has been in metropolitan public transport ticketing. RFID-based ticketing systems were operational as early as a decade ago and today have been deployed in numerous cities across the globe. In such applications, RFID tags would be embedded in credit-card-sized reusable tickets that store either a seasonal pass or credit that can be used against travel. Readers, positioned primarily at the entrance gates, validate the tickets or deduct fares for travel as appropriate (see Figure 2.3). Unlike older ticketing technologies, notably those employing a magnetic

strip to encode data, RFID-based tickets are capable of holding a code that uniquely identifies the passenger and the ticket and in some cases also store personal information about their holder and a record of the most recent trips.

RFID offers distinct advantages due to the superior durability of tickets, which are placed in a hard plastic enclosure and as a result can be used for much longer periods of time. Ticket longevity also benefits from the relative lack of wear and tear during automatic inspection since no mechanical components are involved in the process. Ticket inspection at the gates is also facilitated by the far higher read accuracy of RFID compared with magnetic, which helps maintain the steady flow of commuters in and out of the system, especially at peak times. Finally, RFID tickets can hold considerably more data, which allows the use of personalized unique identifers that can be used to virtually eliminate fare evasion.

Fig. 2.3. An RFID ticket validator at the gates of the London Underground network.

Many RFID ticketing systems employ the ISO 14443 standard, which provides specific facilities for transport applications. It is also common to use proprietary extensions to improve security through the cryptographic protection of communication between reader and tag as well as access to the tag memory (common choices include the Philips MIFARE protocols and the Sony FeLiCa system). ISO 14443 uses inductive coupling, has a relatively short range, and has several other useful features for ticketing which we discuss in more detail in Section 3.2.2. Its short read range in particular is used to the advantage of this application, as even in cases where readers are installed in relatively dense configurations, it is always clear which ticket corresponds to the tag presented by a specific passenger. Finally, data throughput requirements for ticket validation and inspection are very low compared with the available read performance, and the timing overhead as perceived by commuters is minimal.

The Oyster card

One of the largest RFID-based ticketing systems is the Oyster card [69] in London, UK, which supports more than 10 million active passengers and has deployed over 27,000 readers. The Oyster card can be used in all modes of public transport operated by Transport for London (TfL), including the Underground, the Docklands Light Railway, the Croydon Tramlink, and the London River Services. A measure of the complexity of this system is the fact that management of the London Underground services alone involves approximately 500 trains running at peak times, 253 stations owned (275 served) and in use for 19 hours per day, over 12,000 staff and numerous engineering assets. Moreover, the London Underground has been in continuous operation since 1863 and as such has inherited a variety of components and facilities that were created using legacy technologies but cannot be completely or concurrently overhauled due to the disruption this would cause to the operation of the service.

One area where efficiency improvements are critical is streamlined ticketing, which has a central role to play in reducing the time required to access trains. To this end, TfL has recently introduced improvements in self-service ticketing (for example through the operation of credit card processing facilities by passengers without the involvement of staff) and payment, faster entry and exit, and better transportation between the gates and the platform. Except for the latter issue, these improvements were the core objective of the Prestige project, which supported the implementation of RFID and included a number of additional aims, including the prevention of fare evasion.

Reduction of ticket counterfeiting, and "fare dodging" in particular, is estimated to cost 5 to 6 million British pounds on average each year. Prestige provided for major upgrades in all aspects of the entry and transfer subsystems over a five-year period and this has been estimated to cost over 10 billion British pounds. Of these, only about 1% is associated with the implementation of RFID technology. Clearly, the cost of the technology itself in this case can be easily justified, and indeed to a significant extent it can be directly recovered by the gains due to fraud reduction.

Interaction design, system affordances and performance

RFID was selected to replace magnetic strip ticketing because it can deliver improved performance in terms of higher throughput of passengers from the station entrance to the platform. RFID systems only require the card and the reader (which in this case is also capable of writing) to be in close proximity, thus eliminating the need for magnetic systems to insert the ticket into a reading unit for precise positioning. Instead, RFID readers can be placed in any convenient location on ticket vending machines and the gates, for example on the entry gate's top side and within easy reach for the majority of the passengers, as seen in Figure 2.4. Moreover, the card itself can be made out

of durable strong plastic, which reduces wear and tear on the card and makes reading errors virtually impossible. This fact results in considerably shorter times required to operate the turnstiles.

Fig. 2.4. Entering The Tube using the Oyster card. Note the placement of the Oyster on the special indicator, which ensures accurate reading and the displays embedded on top of the gates, which provide visual and sound feedback.

The colocation of both magnetic and RFID card readers on the same gating equipment (see Figure 2.3) and the fact that both magnetic and RFID cards have the same size gave rise initially to some interesting behavior. One of the early findings of user research was that commuters used to magnetic strip tickets frequently inserted their new Oyster cards into the slot of the magnetic strip reader, which compressed and damaged the card. Subsequently a provision was made that the plastic enclosure of the RFID chip be able to withstand at least 100 such events.

Although RFID cards can be operated at a distance and thus do not require contact between the reader and the tag antenna, in practice two considerations also had to be factored into an appropriate design and the selection of reading range. First, it must always be clear what card is presented to which gate by whom to ensure that correct charges are applied. This is so even in locations where gate density is high, resulting in a large number of individuals using the system concurrently and at close proximity to each other. To achieve this effect, the range of the system was restricted to only a few centimeters, a choice that provided the desired effect.

Second, cards must be read and updated with very high accuracy which implies good placement of the card antenna within the field created by the

reader. A well-known limitation of inductive RFID systems is that when the plane defined by the coils of the tag antenna is approximately perpendicular to the field created by the reader, then the coupling effect is weak and leads to a failure to charge the RFID chip. In early trials with the system, such erroneous use of the card was observed far more frequently than expected, likely due to the location of the Oyster card reader on the gate. As a result, usage guidance provided by TfL now suggests that the card be "touched in and out" with the card placed flat on the reader. Despite the fact that contact is not required, touching the card guarantees that the orientation of the antenna is appropriate and a successful read or write very likely.

A clear benefit of the new RFID-based approach is the reduction by approximately 3–4 seconds of the time required by commuters to get from the point of entry to the station until their arrival at the platform. This is primarily due to the fact that removing ticket validation errors, which disrupt the incoming flow, results into a smoother distribution of commuters along the system. This has been perhaps the biggest success of Oyster in terms of service improvement, together with its use in buses.

Using magnetic strip technology it was not practical to operate read-write equipment on buses, and as a consequence ticket issue and validation of passes was carried out manually by the driver. The small form factor and the lack of any mechanical parts in the RFID reader have made viable the use of the Oyster card in this context also and have accelerated the ticketing process, which is now conducted automatically. This has improved bus performance especially during the morning rush hour, which has benefited the most from the efficiency afforded by RFID ticketing. In the longer term, the introduction of Oyster is expected to have even more significant implications for buses in particular, as it allows the possibility of operating with cashless carriers, with additional security and safety gains.

From a user perspective, the biggest success of the Oyster card seems to be removing the "fumble factor". Users must take a magnetic ticket out of their wallet, put it into the reader, walk through the gate, and pick up the ticket on the other side, none of which are necessary with Oyster. So radically reducing the fumble factor was a major service offering and probably the biggest motivation for people to use it.

Another system design challenge has been how best to provide feedback of successful or unsuccessful operations to the passenger. Again, the emphasis in Oyster has been to support the highest possible throughput, which in this case implies that feedback should be simple and unambiguous. The Oyster system uses sound and simple red or green LEDs to provide confirmation of successful authorization as well as embedded displays in the gates and on ticketing machines for more complex interactions. In this way, access authorization is separated from general travel management and payment issues: the former is optimized for speed and is carried out using simple feedback, and the latter is designed with ease of use and adequate information provision in mind.

Organizational changes

Another lesson learned from the Oyster system is that while technology can be developed and deployed relatively speedily, organizational issues and ensuring stakeholder involvement require much longer time frames for developing successful solutions. To this end, a guiding principle that proved successful in Prestige was to set significant and realistic targets that required a limited number of changes to occur at any one time. When engaging the organization internally, it is necessary to identify all the roles and processes affected and allow the inevitable learning curve to catch up. In particular, staff training and internal communications are critical, but providing for these tasks increases in complexity with the size of the organization.

Common sense, if not due diligence, dictates that systems that have millions of users on a daily basis should be durable and easy to use and thus be designed accordingly and heavily tested. Nevertheless, even in cases where adequate provisions are made, it is still unlikely that the transition will be seamless. For this reason, it is necessary to take into account that even the simplest system when operated at large scale, will cause some degree of confusion that must be dealt with. Consequently, supportive measures should be in place beforehand and cater to the issues raised during the user's learning curve. This was clearly something that the Prestige project did especially well; for example, staff training started three years before the first cards were issued to the public, and planning for system re-engineering tasks started well in advance of that.

A longer term issue is the development of the Oyster card into a general purpose mobile electronic payment system. This approach has been adopted by other ticketing systems, notably the Octopus card in Hong Kong and the Suica card in Tokyo (see Figure 2.5), both used for micro-payments in a large number of retail outlets in each city. This extension of card use has been very successful and very popular with consumers in both cases mentioned above. But it also implies entry into a fundamentally different market, that of financial services, which is associated with significant risks and may not be desirable, especially in an organization that is publicly held.

Network support for the Oyster card

The operation of Oyster is supported by an extensive multi-tiered back-end information system that processes payments, records and validates transactions, identifies fare evaders and prohibits their future entry, and records transactions. Recent travel details are held on the RFID tag (last seven trips taken), in a local system at the station level, and at a centralized data warehouse. Updates to the data warehouse are carried out in nightly batches, where incoming data are cross-validated, cleaned, and updated. These supporting network services are proprietary applications delivered over a secured virtual private network, which despite its complexity and strict performance requirements,

Fig. 2.5. Using the Suica card as a general-purpose payment mechanism.

is nevertheless a fully controlled environment. In particular, the store-and-forward approach selected as a core architectural feature has been a very successful choice and has provided the basis for uninterrupted and robust performance.

Unique identifiers are used in combination with suitable cryptographic protocols to authenticate a card to the system and to provide protection of the communication between the card and the reader. As a result, ticket counterfeiting is very hard unless the cryptographic keys are cracked (a brute-force attack is unlikely to succeed in this case, as the wireless communication overhead incurs a considerable delay, which makes such attempts impractical). Moreover, the use of unique card identifiers ensures that specific cards can be pinpointed and associated with activities that violate the rules of acceptable use. Such cards can be subsequently blacklisted, thus preventing their holders from entering the system.

Management of blacklists is an interesting aspect of the system. Lists are downloaded to gates and bus ticketing equipment which locally carry out the verification of the access credentials presented by the cards. The lists are refreshed daily to include new card codes that have been found in violation and are purged monthly to remove cards that have appeared inactive for longer periods. This strategy does not completely guarantee the elimination of fare evasion by well-organized and determined commuters who are committed to performing complex usage patterns employing multiple cards. Nevertheless, it does provide a good compromise in the context of the overall risk management approach adopted by the system and fits well with the system design philosophy, which gives precedence to robustness and failure tolerance.

Indeed, system design driven by appropriate risk management considerations is a tactic employed consistently by the Prestige project. Notably, a core design trade-off relates to balancing system survivability against its ability to authenticate users accurately and provide access to transport services even when the card management back end becomes unavailable. In this case, the transport system as a whole will continue to operate for at least seven

days without any noticeable disruption to passengers at the minor additional risk of increased unauthorized access. This design decision has had great success up to now, and in practice the system has been operating without any high-profile or large-scale disruptions for several years.

The main ingredient that has guaranteed such undisrupted operation is the heavy use of caching and replication architectures at all system levels. The Oyster card itself employs its extra storage capacity—in addition to the unique identifiers and in some cases some personal identification data—to hold records of the most recent read-write events observed and the active value of the ticket. Even in cases when the information stored in the card cannot be verified in real time at the gate, passengers are still allowed into the transport system and the transaction is stored locally until network connectivity and communication to the central servers is restored. While this opens up a small window of opportunity for fare dodgers, the financial impact would be localized and still compares favorably against the magnetic strip system. The success of this approach is also highlighted by the fact that since the introduction of the Oyster, revenue under similar traffic conditions has been increased by 10%, while at the same time ticket inspectors carry out far fewer controls.

A complication that in smaller installations of RFID would be a mere annoyance or could be addressed effectively through longer user training (which is not an option here given the scale of this system) is that RFID cards, despite being successful in authenticating passengers who gain access to the system, frequently fail to confirm the successful update of the transaction history maintained on the card itself with the new trip details due to synchronization errors. This happens with some regularity and results in an increased degree of uncertainty regarding the status of the system. This problem has been addressed by differing the processing of incomplete transactions to the back end, which can do a better job of establishing the true status of the system, as it can view the totality of the sequence of recorded transactions and the local status of the card is discarded in subsequent writes.

2.3 Supply Chain Management

Supply chain management (SCM) deals with the movement of goods between organizations from raw material to finished product. Each supply chain is distinct and reflects the unique needs of the range of products that have to be processed, from supplying fresh food from the farm to the supermarket shelf to delivering uniforms from the manufacturer to the soldier in the desert. Nevertheless, all supply chains share a common goal: to keep the process simple, standard, speedy, and certain.

To achieve this goal, it is necessary that all participating organizations across the supply chain exchange accurate information at frequent intervals and that supply chain costs be unequivocally identifiable at all times. An

essential element of any solution that can meet these requirements is the use of open, worldwide data standards for globally unique product identifiers and product classification systems combined with internet-based information services that can be used to track and trace goods and services [14].

2.3.1 Creating Consumer Value

Among all retail sectors, grocery is the most competitive, as it operates with minimal profit margins. It is thus important that grocery retailers exploit any possible efficiency improvement opportunities offered by technology, and indeed over the past 50 years they have pursued this objective with considerable success. In particular, the supply chain of grocery products, also known as fast moving consumer goods (FMCG), has attained considerable operational gains through the implementation of information technologies, including bar codes, resource-planning software, and optimized logistics.

Efficient Consumer Response

The need to respond to consumer demand with greater efficiency has also produced Efficient Consumer Response (ECR), an initiative to raise performance levels across the entire retail sector [80]. ECR aims to achieve this through the re-examination of processes and procedures for the industry as a whole, recommending improvements, and overseeing the implementation of recommendations. ECR started in the United States but due to its clear business advantages has rapidly extended its reach to the rest of the world, with national and regional initiatives in action.

ECR has identified three priorities for the supply chain:

(i) to increase consumer value,
(ii) to remove costs that do not add consumer value, and
(iii) to maximize value while at the same time minimizing inefficiency throughout the supply chain.

In practice, these priorities are used to identify and fulfill specific goals, for example providing consumers with the products and services they want, reducing inventory, eliminating paper transactions, and streamlining product flow. To meet these goals, distributors and suppliers may need to make significant changes to their business processes that can only be applied through the implementation of novel information and communication systems.

And there is ample room for improvement. Decades after the introduction of information systems in production and logistics control, there are still significant inefficiencies in modern supply chains that adversely affect the cost of retail operations. Upstream supply chain inefficiencies affect the relationships of all trading partners and result in high out-of-stock conditions at the point of sale, a high returns rate, and long lead times. Inefficiencies in the downstream direction negatively affect the accuracy of demand forecasts, which

results in low on-shelf availability and thus loss of revenue despite the fact that products are available on-site. Moreover, information-sharing ineffectiveness between trading partners reduces the accuracy of demand forecasts and the scheduling of the replenishment process.

A direct consequence of low demand forecast accuracy is that trading partners have to maintain increased inventory levels to address unpredictable increases, which in turn result in increased logistics costs. Common practice today is to forecast consumer demand by processing historical point-of-sale data using decision support systems that utilize data warehousing and data mining techniques. However, using point-of-sale data to make forecasts results in lower accuracy because demand patterns chang rapidly and such fluctuations cannot be captured at the point of sale but have to be identified earlier in the consumption process. Moreover, historical forecasts cannot effectively take into account the influence of promotions and other marketing instruments, since the success rate of such mechanisms is generally hard to quantify beforehand.

The actual effects of this situation were quantified recently, with estimates that 53% of out-of-stock conditions are due to store replenishment inefficiencies. Even worse, a further 8% of on-the-floor out-of-stock conditions occur despite the fact that the necessary supplies are in storage on-site. To improve these results, it is necessary to record consumption data earlier in the replenishment process so as to allow for greater prediction accuracy, which leads to reduced inventories and optimized supply chains both upstream and downstream.

Vendor-managed inventory

One contribution toward the ECR goals is the so-called Vendor Managed Inventory (VMI), where the vendor rather than the customer specifies delivery quantities sent through the distribution channel. This reversal in the procurement process has become possible only through the deployment of Electronic Data Interchange (EDI) systems, a computer-to-computer exchange protocol for business data. VMI has succeeded in reducing stock-outs and inventory buffers in the supply chain. Common features of VMI include reduction in supply chain length, centralized forecasting, and frequent communication of inventory levels. From a fleet management perspective, delivery vehicles are loaded in a prioritized manner: items that are expected to stock out have top priority, then items that are furthest below the targeted stocking levels, then advance shipments of promotional items, and finally items that are least above targeted stocking levels.

In addition to EDI, a second technology critical for VMI is the standardization of bar codes for the automatic identification of products. This technology has played a central role in the automatic initiation and entry phases of the order cycle, which can be reduced by days. The two technologies together help develop collaborative relationships in which any combination of retailers,

wholesalers, brokers and manufacturers work together to seek out inefficiencies and reduce costs by looking at the net benefits for all participants in the chain.

Overall, VMI has been successful in significantly reducing inventory levels and the number of stock-outs. The latter issue is particularly important not only because of lost sales but also because shelf availability is central to supermarket strategy. Indeed, a significant proportion of supermarket profit margins are due to interest-free periods for products already available on the shelves. Thus, one of the main concerns of retailers implementing VMI has been the perception that reduced inventory will result in less product being available on the shelves and therefore loss of market share.

2.3.2 The Role of RFID in SCM

The traditional way to automatically capture product information in the supply chain, as employed by ECR and VMI, has been through the use of bar codes. Bar codes were invented for this reason and indeed have a long and interesting history [100]. Each printed bar code symbol represents a number that can be used to identify the labeled product and possibly link to stored information about it. For example, on the back cover of the citation just referenced, a bar code provides a representation of ISBN code 1846280354. ISBN (International Standard Book Number) is a numbering scheme specifically developed for books and provides a unique identifier for every book published. Using this number, it is possible to search one of several internet databases to retrieve further information, including its title, author, and publisher, and in some cases even retrieve a considerable proportion of the published material, for example on Google books.

Despite their great success and popularity, bar codes have several limitations:

1. Reading a bar code requires a line of sight between the label and the scanner. For instance, a truckload full of products will have to be unloaded and each individual item scanned to retrieve full information about its contents.
2. A label is typically printed and affixed to the product packaging and for this reason exposed and likely to be damaged. In this case, it is no longer possible to identify the marked product.
3. General-purpose bar code symbols can store only a small amount of data, which in the vast majority of cases identifies the particular product item as one of a certain type but cannot differentiate between to distinct items from the same product line.

Of course, products also have significant advantages, as they are a well-established, open, and robust technology and can be produced at minimal cost. In fact, in many cases bar codes are part of the design of a product's packaging and as such have almost no cost at all.

Similarly to bar codes, RFID can be used to store globally unique product identifiers, which moreover can provide item-level rather than class-level identification granularity. RFID tags can also provide the means for automatic capture and processing of this information with the added benefit that they do not require line of sight and that since they can be embedded inside a product, they are far less likely to be damaged. As a result, RFID is seen as a good candidate to replace bar codes, although this requires that the additional cost not exceed the potential gains from its implementation. Such gains would be due to a number of efficiency improvements that can be roughly outlined as follows.

- *Reduce inventory levels.* Highly accurate data allow all partners in a supply chain from the manufacturer to the retailer to maintain lower inventories and plan deliveries and shipments in a just-in-time manner. Currently, it is not possible to predict such patterns accurately, and as a result all partners need to maintain excess stock as a buffer against sudden surges.
- *Reduce out-of-stock.* Despite increased inventory levels, it is still common that supply chains run out of stock and as a result cannot respond to consumer needs. Of course, unavailability of stock is directly responsible for lost sales.
- *Reduce order and lead times.* By increasing the visibility of information held in the systems of supply chain partners, it is possible to implement aggressive ordering strategies driven by the vendor rather than the client, which can significantly shorten lead times for orders.
- *Reduce shrinkage.* By marking individual products, it is far easier to identify when and where items have been lost and as a result considerably reduce not only internal theft but also shrinkage due to products expiring.
- *Increase on-shelf availability.* Effective replenishment processes developed on the facilities outlined above would result in higher product availability at the retail outlet shelves. Lost sales due to products not being available on shelves, in some cases despite the fact that they are available in the store, account for a significant proportion of lost revenue.
- *Increase consumer service levels.* Detailed information about individual items can support higher service levels both at the point of sale and for after-sales services, especially safety through more effective management of product recalls. When all product items are tagged, it is also possible to provide fully automated quick checkouts that minimize waiting times at exit.

RFID is well-suited to provide the features required to collect exactly the kind of information that is needed in achieving these objectives. Such tracking and tracing of products can be achieved by establishing prominent control points at all levels of the supply chain (see Figure 2.6), so that items, cases, and pallets are quickly inspected and recorded. Of course, the collected data also highlight the role for integrated enterprise resource-planning systems. ERPs maintain and manage information related to the complete lifecycle of

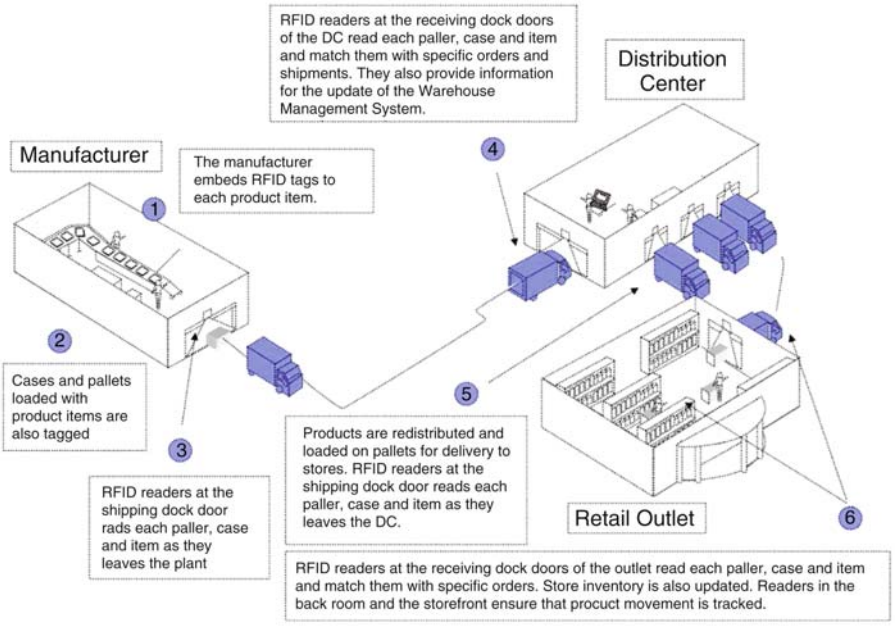

Fig. 2.6. RFID-enabled supply chain management.

business processes, and as a result a successful RFID deployment critically depends on the availability of such systems.

In any case, a typical example of a supply chain control point would be a dock door at a manufacturing or warehouse facility that is equipped with RFID readers that record the codes embedded in pallets loaded with incoming and outgoing products, and automatically update the company's ERP. Subsequently, RFIDs could be used for taking quick inventories using handheld devices [32]. We will often return to this example in subsequent chapters and will discuss in some detail how such a portal would operate as well as all the components required for its construction and operation.

2.3.3 A Brief History of RFID in the Supply Chain

The utility of RFID in this role was highlighted in the early 1990s in work carried out by the US Department of Defense (DoD). A feasibility study was conducted as a result of a new doctrine for land operations that was first employed during the First Gulf War, which dictated the rapid advance of mechanized forces at speeds unprecedented in military operations. This approach caused considerable problems in downstream information flows and upstream in the supply chain: intelligence about possible targets often had a lead time of over 12 hours and combat units often found themselves without

supplies due to the inability to replenish materials—in many cases despite the fact that the required resources were available in storage nearby.

This situation prompted the DoD to propose standard systems and protocols for the transmission of information in the so-called Network Centric Warfare model, but more relevant to this discussion, the management of the supply chain using RFID at the container level. At that time, the technology was not cost-efficient for commercial use, but it did highlight the advantages of a system for automatic identification and tracking of products.

The falling cost of RFID technology over the following decade meant that by the early 2000s tags had become cost-effective in a variety of situations. As a result, several pilot projects highlighted the possible benefits of the technology, and subsequently some of the largest retailers decided to deploy it: Wal-Mart in the United States, Tesco in the United Kingdom and Metro in Germany are all actively implementing RFID to track products across their supply chains. Although the vast majority of these deployments deal with product containers (that is, pallets and cases) rather than individual product items, the technology employed is being developed with a view towards supporting item-level tagging if and when it becomes cost-efficient.

These recent RFID developments have been driven to a certain extent by the Auto-ID Center, established with the support of several manufacturers, suppliers, and retailers of consumer goods for this reason [27]. The early work conducted at the Auto-ID Center has been taken over and further developed within EPCglobal, which operates as a subsidiary of GS1, a global standards organization of automated open supply chain systems previously known as EAN/UCC. GS1 was established through the merger of previously separate bar code standards organizations to ensure that a unique system will be used globally in the supply chains.

The main contributions of EPCglobal are being developed around the so-called Electronic Product Code (EPC) that has been formalized and provides an unambiguous numbering scheme to identify goods containers, services, companies, locations, and assets worldwide [27, 100]. This identifier scheme has evolved into a full system of specifications that define all aspects of network RFID, including tag operation and communication, identifier schemes for several types of entities, network resolution, and meta-data repository services. In subsequent chapters, we examine several of these specifications in detail, as they often represent the state-of-the-art in network RFID technology (see Figure 2.7 for an example of modern EPC-enabled RFID equipment).

2.3.4 Implementing RFID in SCM

Unlike ticketing and e-passport applications, supply chain management (SCM) is a far more complex and challenging environment for computing. One special requirement of SCM is the need to read large numbers of tags in a very short period of time while products are physically moved through warehouse portals or other supply chain control points. This task is made even more complex

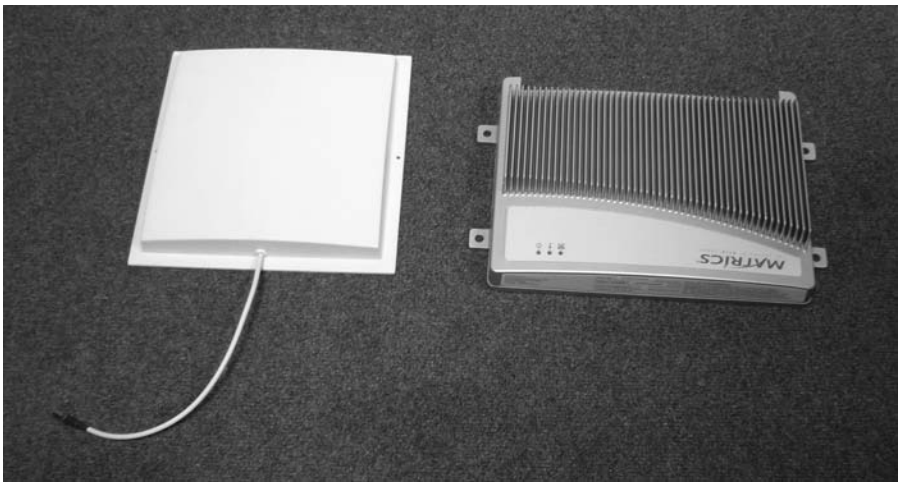

Fig. 2.7. EPCglobal has introduced its own specifications on tag memory layout and communication with readers. Several manufacturers have developed hardware specifically to support these standards, in this case a stand-alone networked reader that can support up to eight antennas.

by the fact that tags read at the same time may represent different types of objects and so they must be filtered and aggregated. Another complication is due to the fact that readers may need to be deployed in dense constellations causing their reading ranges to overlap and complicating communication with individual tags. Unreliable reads and writes at the media access layer cause cascading effects in applications and thus to reliably identify significant events additional smoothing of the data has to be performed.

Finally, despite the fact that early SCM applications focused on container-level tagging, it is becoming increasingly common to extend RFID to the item level, with several retailers currently implementing the technology in commercial applications [69]. This fact can potentially have the greatest impact, as it creates a situation where large collections of objects are directly available to computing systems for auto-identification and can be used to develop end-user applications. A back of the envelope estimation gives a sense of the size of the problem at hand: there are roughly 500 million pallet shipments per year across the globe and across industrial sectors, which correspond to 100 billion cases of products and approximately 2 trillion items. Each item would be scanned and recorded several dozen times as it moved through the supply chain and to its final destination. Without doubt, this will produce a massive volume of data that must be processed and used effectively by applications.

2.4 Other Applications

In the previous sections we touched upon three of the numerous applications that employ some form of RFID technology. Since an exhaustive review of applications would require a book in itself, in the closing sections of this chapter we will only summarize a few of the more popular areas that highlight the capabilities of RFID.

2.4.1 Asset Management

Asset tracking, monitoring, or management is an integral task of doing business, especially for large organizations. Similar to SCM, asset management can potentially offer increased efficiencies and lower operational costs. For example, deploying and re-deploying equipment in quick successive cycles to meet the needs of business operations requires a detailed inventory of what assets are available, their condition, and where they are located. Tagging with RFID has the potential to lower the overhead of tracking and tracing and provide more up-to-date information thus improving flexibility.

One task for which RFID can offer significant improvements is prescriptive maintenance, whereby equipment use is constantly monitored and maintenance tasks are communicated proactively to staff that can carry out repairs before actual failure happens. Such facilities improve overall field service efficiency, reduce outages, and offer savings in parts costs. Of course, as in all other applications we have considered, RFID alone will not achieve this, but it requires the support of network-based information systems that manage and communicate tasks.

There are many different situations where asset management would be involved that offer widely diverse profiles and requirements, including document management, aircraft maintenance, management of building materials for the construction industry, baggage handling, and cleaning equipment auditing. Although each of these applications employs RFID as a core component of its solution, systems as a whole are fundamentally different and require specific domain expertise. For instance, Figure 2.8 shows the TrakSens software used to monitor the auditing of robotic cleaners.

Finally, RFID is becoming increasingly popular for tracking assets in healthcare applications including anti-counterfeiting for drugs and the control of bio-samples (see Figure 2.9). As regards the latter application, a particularly interesting flavor of RFID has been developed with reader antennas using coil-on-chip technology and tags embedded in test tubes, which offer an interesting alternative for the automation of laboratory records.

2.4.2 Electronic Payment

Over the past decade, electronic payments carried out either through the use of a physical token, for example a credit or debit card or a mobile phone, or over

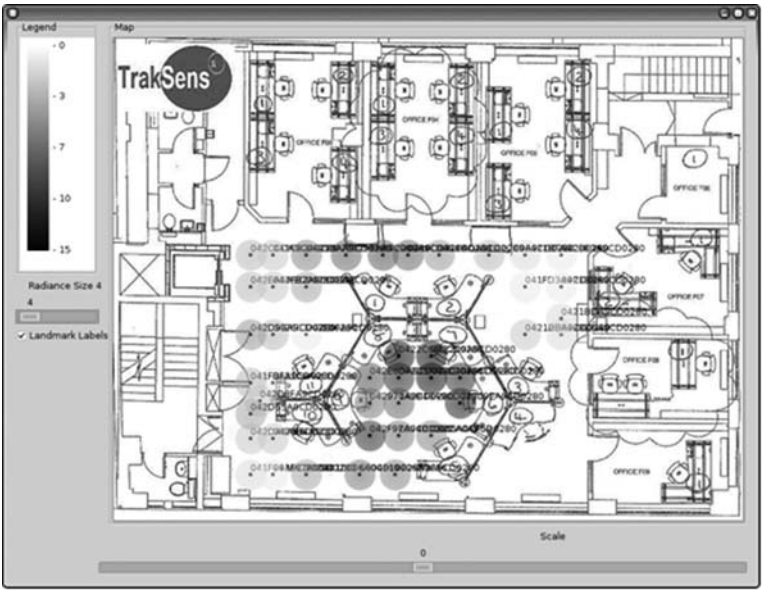

Fig. 2.8. Auditing cleaning assets with RFID. Image by Dikaios Papadogkonas.

the internet, have been steadily increasing in popularity. In Europe and North America in particular, some form of electronic payment is used for a large proportion of all transactions involving relatively large amounts. Electronic transactions are not as popular when smaller amounts are involved, roughly the size that would typically be made in coins. Several reasons are quoted as possible justification for this including the "fumble factor" also involved in ticketing (a signature or a PIN number is also required in most cases) and the relatively high per transaction cost, which makes such micropayment systems expensive to operate.

Implementing RFID-based micropayment is an approach that has attracted considerable interest in recent years, and there are already several solutions available from most major credit card operators. Such cards, often called contactless smart cards, hold electronic cash on their RFID chip and communicate with a vendor reader to transfer small amounts without the need for further authentication by their holder. For example, Figure 2.5 shows the Suica card used for ticketing in the Tokyo metro system, which can also be used for such small payments.

Despite their potential benefits, such contactless smart cards have only been deployed in controlled trials due to their potential for abuse. Although the amounts involved may be relatively small, collectively they represent a significant financial risk, and a staggered approach is being developed for their adoption by the general public. Instead, several card operators are considering

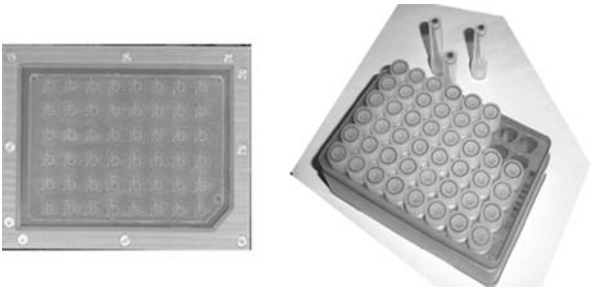

Fig. 2.9. Maxell Seiki coil-on-chip technology and RFID-enabled test tubes for asset management of bio-samples.

multi-cards which combine a magnetic credit card, a (contact) smart card for cash payments, and an RFID ticket.

2.4.3 Animal and Human Tagging

One of the first applications of RFID has been in tagging cattle for automated accounting in farms. Such tags facilitate accurate record-keeping, including recording the locations where particular animals have been kept, immunizations, and feeding patterns. In this context, RFID has been used as a more effective solution than bar codes, which would often be destroyed due to soiling and the generally difficult conditions under which cattle live. Tags of this type are typically molded as cattle ear rings and use LF frequencies.

Another use of RFID with animals is in tagging pets. This type of tagging can take two distinct forms:

- Collar tags are used to operate pet doors that allow tagged animals to move freely in and out of the house but remain locked at all other times.
- Embedded glass-enclosed tags are used in some cases to provide identification for pets, so that if lost and later recovered, they can be returned to their rightful owner.

This type of RFID also commonly uses LF frequencies and has fairly simplistic operation with minimal security controls.

There are also uses of RFID for tagging humans. A common application concerns the identification of hospitalized patients, who can be issued an RFID bracelet that verifies their particulars. This approach has been developed in response to increasing numbers of errors that lead to either unnecessary surgery or the administration of the wrong medication, with detrimental health effects. There are already several hospitals that have experimented with this technology, but currently uptake remains restricted to specific wards, often those involving extensive surgery and lengthy recuperation periods.

A variant of the embedded glass tag for pets has recently been certified for use with humans. The so-called Verichip is easily implanted using a syringe but requires a minor surgical procedure for its extraction as it binds to

human flesh. It has been used in a small number of applications (for more details see Chapter 9) notably to establish the VIP status and provide payment facilities to patrons of a nightclub chain in Europe, which has attracted extensive publicity. Other applications are less benign, and there have already been proposals for its use to tag military personnel and migrant workers in the United States. Unsurprisingly, the availability of this tag has raised very considerable concerns and highlights the ethical issues related to RFID use.

2.5 RFID Standards

Looking at the different applications of RFID, it immediately becomes evident that each specific domain has its own priorities, norms, and preferences. This is reflected in the standardization process, which has followed distinct and often incompatible paths in establishing the operating parameters of RFID systems as specified for a particular domain of application. This is despite the fact that many of these systems share many common features and could potentially benefit from lessons learned in a different domain. This is not the case, though, and in practice RFID standards are numerous, often incompatible, or mutually exclusive. Yet support for standards is a core ingredient, especially for open systems.

The most prolific producer of RFID standards over the years has been the International Organization for Standardization (ISO), which has issued almost 50 different specifications, several of which have many parts. Navigating the numerous ISO standards is a complex exercise, and we will not attempt to do so in this book. Instead, we note that especially in the context of the supply chain and UHF tags specifically, EPCglobal represents a considerable challenge to ISO, with several of its systems already established as the de facto standards.

2.5.1 EPCglobal

To be sure, EPCglobal provides the de facto standards for RFID tags in the UHF range. Since its founding as the Auto-ID Lab, EPCglobal has provided the critical mass of technology providers and FMCG sector support to develop and evolve a variety of standards for the supply chain. Without a doubt, this work has played a central role in extending the popularity of RFID. Unlike those of the ISO, EPC standards are tightly controlled by specific industrial interests, and it is doubtful that they can represent the concerns of all involved in the multifarious applications of RFID.

This point is of particular significance, especially in the context of general purpose computing, as the declared objective of the work of EPCglobal is to construct the Internet of Things in the sense of establishing a common ground for the development of open and shared infrastructures for auto-identifiable networked objects. However, this is where the similarities between EPCglobal

and the internet standardization process stop: the structures and aims of the two systems are fundamentally different.

EPC and related standards have been developed in competition with the ISO which has been working on a comparable system, albeit at a much slower pace. After several years of conflict, it seems that there are now encouraging signs of collaboration with the EPC Class 1 Generation 2 tags (further discussed in Section 3.2.4).

2.5.2 ISO 14443

ISO 14443 specifies a class of RFID proximity tags that are particularly well suited for ticketing applications. This standard comes in different parts and is relatively complex as it provides for payment and as such support for relatively complex exchanges between the card and the reader. Such cards operate in the 13.56 MHz band and use the magnetic field created by a reader coil antenna, and they typically have a range of a few dozen centimeters. We will take a closer look at the structure of the tag and its supported protocols in Section 3.2.2.

2.5.3 ISO 15693

ISO 15693 specifies a class of RFID vicinity tags similar to those of ISO 14443 in that they also operate in the 13.56 MHz band and use the magnetic field created by a reader coil antenna. However, ISO 15693 tags have a far greater operating range which can be between 1 and 1.5 meters, but support only relatively simple exchanges, primarily the transmission of a unique identifier code.

2.5.4 ISO 15459

ISO 15459 defines a class of unique identifiers for transport units, including supply chain items and containers, returnable assets, and product groupings. It also outlines registration and code address space management processes that re-use existing ISO standards and procedures. In this role, it is roughly equivalent with the specifications of the different serialized electronic product codes developed by EPCglobal and similarly relates to pure identifiers that can be subsequently represented in multiple forms, including bar codes and RFID.

2.5.5 ISO 18000

ISO 18000 is an all-encompassing specification for RFID air interfaces and aims to provide a comprehensive reference for all frequencies and types of tags, including LF, HF, UHF, and microwave active and passive tags. Each section

of the standard describes the physical layer specifications for communications between reader and tag, the protocol and the commands, and specific anti-collision and singulation methods (see Chapter 4).

ISO 18000 was first published in 2004 and has been in direct conflict with the EPC Gen2 specifications developed in parallel. The 2006 revision of the standard offers certain modifications (discussed in Chapter 5) that cater to better interoperability between the ISO and EPCglobal systems at the air interface layer. However, standards at higher levels of the protocol stack remain incompatible, and both systems follow their individual divergent paths. EPCglobal has distinct advantages, especially in terms of vendor support, but ISO compliant serialized identifier systems, for example ISO 15459, also have advantages related to intellectual property and subscription costs.

2.6 Summary

In this chapter, we outlined three applications of RFID that highlight how the technology is used in many practical situations. Yet RFID applications have very different characteristics and as such very different requirements. The result is that there is no single solution or system blueprint that would be appropriate for every application domain and every deployment. Instead, in this book we aim to describe a collection of different techniques, designs and solutions that system designers can mix and match and tailor to their specific requirements. One factor that plays a central role in this is related to system size and parameters. To help with the decision process, we summarize the different aspects of the three application domains discussed so far in Table 2.1 as a guide to similarities with the designs that we will discuss in subsequent chapters with reference to these applications.

Table 2.1. Characteristics of three large-scale industrial RFID applications.

	Ticketing	*e-Passport*	*Supply Chain*
Tag Density	Low	Low	High
Tag Range	Short	Short	Long
Tag Lifetime	Medium	High	Medium
Tag Complexity	High	High	Low
Tag Security	Strong Crypto	Crypto (known key)	Password (32-bit)
Network Support	Delay Tolerant	Local	High—real time

3

Readers and Tags

In this chapter, we begin the detailed discussion of RFID technologies starting with the core components of every RFID system; that is, the reader and the tag. The two devices have an asymmetric relationship in that the tag is simple and offers few facilities besides holding and transmitting the code, while the reader takes the leading role at the cost of higher complexity.

3.1 Readers

There are two ingredients in the operation of an RFID reader: the specification of a scanning plan that the reader executes to collect tag observations, and the processing of the incoming stream of RFID readings that are produced as a result of the execution of this plan and of course depend on the details of the interaction with the tags. A simple plan specification would require that the reader continuously scan for any tag within its vicinity, retrieve its unique identifier, and reports its findings. Such a plan might appear desirable at first sight, but it can have unexpected side effects; for instance, even if only one tag is within the interrogation range of the reader and remains at that fixed location, its code would be retrieved several times per second, resulting in a constant stream of potentially redundant information that should nevertheless be processed. Of course, there may be cases where this is desirable; for example, as a means to confirm that a specific object has not moved away from an observed location.

In any case, readers have to provide components that allow them to receive, interpret, and execute a plan by transmitting instructions to tags, receiving responses, and finally either processing the retrieved codes or forwarding them to other devices for additional processing. Again, there are several alternatives for the implementation of the specifics of the process. For example, a reader can be a simple stand-alone device that receives instructions over a serial interface from a host computer or it can be a network device that receives information over a web service protocol that can carry out advanced

G. Roussos, *Networked RFID*, DOI: 10.1007/978-1-84800-153-4_3,

processing of observations and return reports using a variety of distributed middleware interfaces. Whatever their particular flavor, all readers consist of three principal components:

- One or more antennas, which may be integrated or external.
- The radio interface, which is responsible for modulation, demodulation, transmission, and reception. Due to the high-sensitivity requirement, RFID readers often have separate pathways to receive and transmit.
- The control system, which consists of a micro-controller and in some cases additional task and application-specific modules (for example, digital signal or cryptographic co-processors) and one or more networking interfaces. The role of the control system is to direct communication with the tag and interact with applications.

A basic RFID reader is shown in Figure 3.1, which also highlights its main components. This is a typical example of such a device, and the particular one shown is manufactured by ACG in Germany and is designed for use with a computer that provides power and control over the USB port. This type of reader often supports a simple set of instructions that are used by the host to specify the details of the scanning plan. In return, the reader would produce a stream of data reporting the results back to the host program in a fairly raw format. In the following section, we will consider some examples of how this works in practice.

Nevertheless, readers are increasingly becoming complete network computing devices (akin to routers) that provide advance processing of RFID observation streams and wired or wireless connectivity to the internet. Such devices are far more complex and powerful than the type of reader discussed

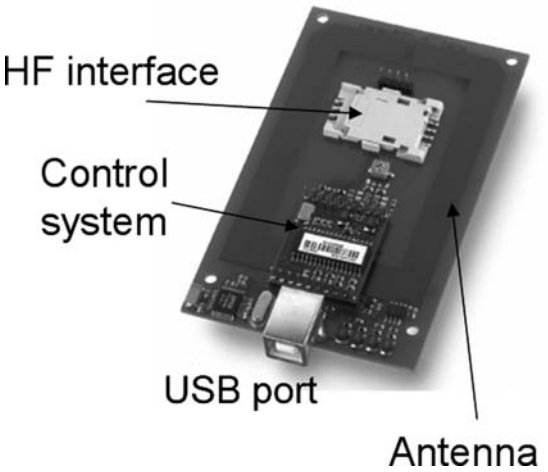

Fig. 3.1. Simple RFID reader, highlighting its main components.

above and may use multiple components for the implementation of each of the core areas of functionality. For example, Figure 3.2 shows such a high-end reader, which provides ethernet connectivity, support for multiple frequencies, and its own power supply. Such readers would also receive a scanning plan from a driving application or other middleware, which they would implement by issuing instructions to the tags within their range. A reader of this type offers considerably higher processing power and can deal with very large numbers of tags at the same time. It can also conduct additional processing on the retrieved codes, for example filtering and aggregation according to complex rules, and also encode the data stream that results from this into more complex data structures such as SOAP envelopes. Finally, readers of this type offer management facilities, checking and reporting their state and also coordinating with other network devices to ensure good operation. The details of the different protocols employed in this case are discussed in Chapters 7 and 8.

At a high level of abstraction, when dealing with high tag densities and much more complex operations on tags the overall process remains the same but in practice its details can become quite complex. Especially in the case where a large number of tags are within range or when two or more readers overlap there are situations that can cause conflicts that lead to erroneous readings and failures. In such cases, additional collision-avoidance techniques must be implemented to ensure that communication is organized in a

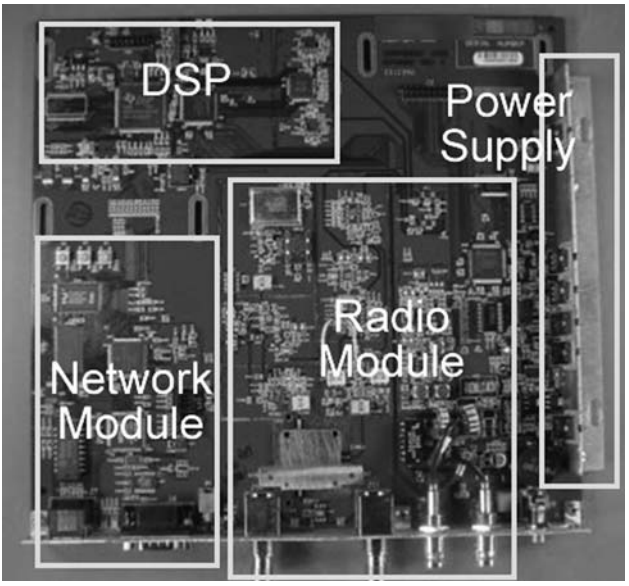

Fig. 3.2. High-spec network RFID reader highlighting its main components.

structured way so as to allow all tags to participate in this process. We will look at some of these problems later in this chapter.

Finally, although the name reader seems to indicate that this is the only capability of the device, in fact the vast majority of readers can also write to tags. Having said that, not all tags are writable (often this is solely due to cost considerations), and even those tags that are may have restrictions on which data are writable or implement access control mechanisms that may prevent writing altogether.

3.1.1 A Simple Reader Session

The reader in Figure 3.1 is a typical example of an HF RFID reader and supports a number of different protocols operating at 13.56 MHz. The particular device was developed by ACG and incorporates TNX technology (a Phillips spin-off) to support the standard ISO 14443 and ISO 15693 protocols and the proprietary Phillips MIFARE protocols. Communication between the reader and the host computer is via a (character or binary) serial interface over which instructions can be sent to the reader and results obtained. In general, each manufacturer maintains its own protocols for this communication link, and there is little compatibility, if any, between products at this level.

In this section, we will use the ACG-specific protocols to communicate with the reader, and although these would not be usable with other reader makes, they are nevertheless typical examples, and to a large extent all such protocols operate in a similar manner.[1] Note that ACG readers come in different form factors and are easily adaptable for use in a variety of applications.

For this example, we assume that the reader is connected to a computer and connect to it over a serial connection and a terminal emulation program set for character communication, as we will be using the character protocol to interact with the reader. Table 3.1 shows the commands that are entered into the terminal and the responses received from the reader. The scenario we implement is one where the reader is set to continuous scanning mode so that as soon as a tag is within range it is discovered and its memory content partially read. This would correspond to the typical operation of a ticketing application as described in Chapter 2, whereby the reader waits until it discovers a ticket within range and then accesses its associated information to check that passengers have a valid ticket and operates the gates or denies access accordingly. After the cycle is completed, the reader returns to scanning mode and waits for the next passenger.

There are a few points to make about this scenario. First, the way that the reader identifies that a tag is within range is by receiving an identification

[1] There is another reason why this particular choice is of interest: two popular RFID open source projects support the ACG reader and provide good reading material for the developer, namely RFID Dump by Lukas Grunwald and RFIDiot by Adam Laurie.

Table 3.1. Reading Mifare tags with an ACG reader using its character protocol.

Command	Response	Explanation
c	no response	The reader is set to continuous read mode.
card enters range	D6520F88	Reader reports the UID of the tag discovered.
.	S	Continuous read mode terminates successfully.
s	D6520F88	Confirm successful selection of card with UID D6520F88.
l01AAa1b1c1d1e1f1g1	L	Successful login to memory sector 1 using cryptographic key A with code A1B1C1D1E1F1G1.
rb04	E4265AFC2345BAC 85273465839247281	Successful read of memory block 04 returns 32 bytes of data.
c	no response	The reader is set to continuous read mode.

code, which in Table 3.1 is called the UID. This number is transmitted by the tag soon after it powers up and establishes communication with the reader, and it is a unique identifier similar to the MAC address of networking protocols for example ethernet addresses. Note that this is not the code that identifies the tagged entity, for instance the passenger number in a ticketing application, but rather an address that is used to identify beyond doubt the peers during a communication session. This UID does not necessarily need to be fixed for the particular tag as is the case for MIFARE and ISO 14443A tags, but can change for every session; for example, it can be a pseudo-random number generated by the tag.

It is often the case that in addition to the two identifiers already discussed, a tag would have a third unique identifier stored in its memory. In the case of the MIFARE chip, this would be the Card Serial Number (CSN), which is hard-coded by the manufacturer for authenticity reasons. This may be confusing when discussed at this level of abstraction but is not so in practice, as the individual details of each reader protocol are often encapsulated within an application programming interface (API) that deals with these complexities transparently. For the application developer, the only important code is the entity identifier, which would be available through appropriate interfaces.

A second important fact about this scenario is that to gain access to the contents of the memory of the tag, it is necessary to provide a password (in this case represented by a hexadecimal string). This password should be known in both reader and tag, which implies a considerable management problem in keeping passwords secret. Furthermore, if the password becomes common

knowledge, then the whole security of the system is compromised and all tags will have to be reprogrammed or replaced. Looking back again to the ticketing application, this clearly represents a very considerable potential problem, and such a security compromise would have significant cost implications.

A final comment is that after the reader is granted access to the tag memory, it reads specific areas from which it retrieves binary data that must be further processed and interpreted by the application. In general, there is no information about the semantics of the memory contents, although in this case the extended MIFARE protocols can be used to gain more information. In the ISO family of standards, there are additional specifications that allow the reader to retrieve specific information about the use of memory areas by applications through the transmission of the so-called Application Data Unit (APDU). In the session of Table 3.1 this is not used.

Having looked at this scenario from the reader perspective, in Section 3.2.2 we will revisit this process as seen from the perspective of the tag.

3.1.2 An Advanced Reader Session

In the previous section, we looked at a simple scenario where a host computer directs the operation of a reader that it controls through a direct serial connection used to issue commands and retrieve information from the tags. In this case, the host executes all the application logic that deals with the interpretation of the incoming data and micro-manages the operation of the reader. This may be useful in some cases, but in many applications it would be more appropriate to issue complete instructions to the reader regarding tags and operations of interest and wait for the result. This is particularly so in cases where a large number of readers are used for similar roles, as in the case of recording incoming shipments at warehouse dock doors as discussed in Chapter 2. Note that such higher-spec readers typically have several antennas, often located at relatively long distances from the device itself, each of which can be operated and identified individually.

Moreover, we have already noted that many of the more powerful modern readers are fully networked devices with advanced computational and communications capabilities. Many of these readers follow the EPC specifications, which include the so-called Reader Protocol (RP), which exactly defines the interactions between hosts and readers. With the RP, the process has two stages: first, it is necessary to issue instructions that define a so-called Data Selector (DS) object, which specifies the rules for scanning tags for example the sources of readings and the patterns of interest. The DS is used by the reader to guide data acquisition; that is, to specify which tag observations will be retained and reported and which will be discarded. These operations are carried out independently and the results buffered by the reader without any further involvement by the host. Note that the definition of a DS does not need to be specified by the particular host that wishes to subscribe to it,

but as soon as the object is created at the reader, any authorized host can subscribe and receive information.

In any case, the host application will retrieve data by opening a communications channel to the reader and registering its interest for the particular DS. It will also specify the manner in which it prefers to receive the data; for example, if it will poll the connection for incoming data or if it prefers to receive event notifications and reports asynchronously whenever data become available. RP requires that such a session have three distinct stages, each consisting of one or more messages, often encoded in XML. Such XML messages can be transferred between reader and host over most common protocols, including TCP and HTTP.

The communication sequence begins with the exchange of handshake messages that specify the communication details between host and reader. For example, the host may send the handshake message

```
RPS111X1X1AR0000END1
```

which specifies that both sides of the communication will be encoded in XML (characters 8–9 and 10–11 for host and reader respectively, both set to X1) and for each message sent, an explicit acknowledgment of receipt is required (characters 12–13 set to AR).

The next step is the transmission of the XML message that specifies the data to be collected through subscription to a DS object. It also defines the ports and other channel details as appropriate for the specific protocol employed. An important aspect of this is the receiving endpoint, which is the address where the data collected should be sent. The process ends when the host sends a goodbye message terminating the communication and de-registering its interest in the particular DS.

Note that the details of this relatively low-level process are not aimed at the application developer but are most often encapsulated in middleware that provide a higher-level programming interface. In most cases, applications are written to these specifications although there is nothing actually preventing one from accessing these interfaces, too. Having said that, the EPC system provides an additional protocol, the Low Level Reader Protocol (LLRP), which provides full access to the complete capabilities of a reader and allows the configuration of every parameter of communication with tags. We will look at middleware for RFID in Chapter 7.

3.2 Tags

Whatever the complexity of the reader, its role is to provide energy to and communicate with the tag, most often to retrieve its code. Readers can also have other interactions with tags, which may be instructed to remain silent for example or may be write rather than read targets. In any case, the tag is a far simpler device and consists of:

- the antenna,
- the chip, which in most cases implements a simple state machine and holds the object identifier, and
- a protective paper or polymer enclosure, which guards against rupturing the antenna, which would result in the immediate expiration of the tag.

In most cases, the chip also incorporates a capacitor, which is used to store the energy harvested by the antenna and regulate it so as to provide the smooth voltage required. Otherwise, minute fluctuations in the strength of the coupling would lead to variations in the voltage supporting the operation of the chip and thus cause unreliable results.

There is little that a tag can do other than respond to instructions from the reader. As soon as the tag has adequate power for its operation, it initializes itself by running self-check routines and sets itself in a condition where it is ready to receive instructions. Then, in response to instructions received, it will move from state to state and provide responses depending on the type of system and the application.

There is a lot of variety in the capabilities of tags in terms of holding capacity, which can be as little as a few bits and up to several kilobytes; in terms of range, which can be from millimeters up to several meters; and communication data rate and density. Moreover, some of the higher-specification tags can offer advanced authentication and cryptographic protection of their contents, though the majority provide only such basic features such as password protection, which is of limited use to closed and tightly controlled systems. Some of these characteristics are due to the particular RFID technology employed and others have to do with format, size, packaging, and cost considerations that affect the design trade-offs.

One of the main distinctions between RFID technologies is related to the frequency range used, which to a large extent defines the manner of coupling between reader and tag that results in energy transfer and also the method used for communication. Modern RFID commonly employs one of two bands, either the Ultra High Frequency (UHF) band (that is, the range of electromagnetic waves whose frequency is between 300 MHz and 3 GHz) or the High Frequency (HF) band (which extends to frequencies between 3 and 30 MHz). UHF tags typically use the electric component of the wave, while HF tags use the magnetic component both for energy transfer and communication. Each technology has distinct advantages, but it is worth noting that HF tags have a longer history of successful implementation, especially for contactless smart card applications, but UHF tags have recently grown rapidly in popularity, primarily due to their use in supply chain applications. A comparison of the characteristics of the two main tag types is shown in Table 3.2.

It is worth noting that many older RFID tags use the LF band (30 kHz to 300 kHz), but due to various limitations of this technology, such tags are used today primarily in specialized applications, notably for implantable tags, due to their superior operation near water. LF tags are of limited importance in

Table 3.2. Comparison of HF versus UHF RFID technologies.

	HF (near field)	UHF (far field)
Frequency	$\approx$ 13 MHz	$\approx$ 900 MHz
Spectrum Allocation	Uniform	Fragmented
Cost (per tag)	$<$ 15 cents	$<$ 15 cents
Range	$<$ 30 cm (1 m max)	$<$ 4 m (10 m max)
External Interference	No	Cellular phones
Memory Capacity	4 Kbits	256 bits

the context of networked RFID and we will not discuss them specifically in this book.

3.2.1 Tags that Use Magnetic Coupling

Power transmission from the reader to the tag is by magnetic induction (the principle employed by power converters), and for this reason near-field readers and tags have a characteristic antenna design that also makes them easily identifiable: their antenna is a simple coil (see Figure 3.3). The effectiveness of this process depends on the strength of the near field at the tag location, which in turn depends on the distance between the center of the reader and the center of the tag antennas (and the particular frequency used). In any case, at frequency f, the near field ends at a distance proportional to $\frac{1}{2\pi f}$ from the reader antenna. For example, at 13.56 MHz, the frequency used by the popular ISO 14443 standard, the near field extends to about 3.5 meters from the reader. However, due to constraints on the spatial layout of ISO 14443 systems, in many cases the practical range of such a system would be approximately 8–12 cm using a medium-sized antenna on the reader and credit-card-sized tags (note that the range can vary considerably depending on several parameters discussed in more detail in Chapter 4).

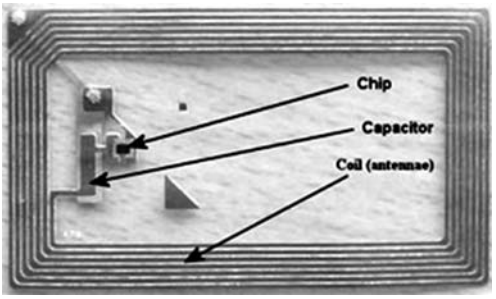

Fig. 3.3. The anatomy of an HF RFID tag.

One of the advantages of the 13.56 MHz frequency that makes it so popular is the fact that this section of the wireless spectrum is assigned worldwide to smart cards and labels and hence is globally available to the vast majority of RFID applications. Other frequencies commonly used by near-field RFID are within the 120–136 kHz range, but these are rapidly loosing in popularity, as they can only be employed for very short-range communications. Their short range makes them unattractive for applications as in most practical situations they necessitate contact between the card and the reader (but not the electronics directly).

Tags communicate by changing the load of the tag antenna in such a way that they control the modulation of the radio signal in a process appropriately called load modulation. These changes can be detected by the reader and decoded by examining changes in the potential variation in its resistance. Because the magnetic field decays very rapidly with distance from the center of the reader antenna (inverse cube ratio), the changes to be detected by the reader are tiny compared with its own transmission. For this reason, the tag modulates the radio signal in such a way that it responds in a frequency slightly shifted from that of the reader (what is often referred to as the subcarrier frequencies).

3.2.2 ISO 14443 Tags

In this section, we will look in more detail at how a typical example of tags that use magnetic coupling operates, specifically looking at the case of ISO 14443A. When such a card enters the interrogation field of the reader and sufficient voltage becomes available, the first step is to carry out a number of initialization and configuration routines so that the tag is ready for operation and enters the IDLE mode. The next step is to build a communication relationship with the reader, taking into account the possibility that other tags may also be active and may already be transmitting. For this reason, while the tag is in IDLE mode, it will not respond to communication until it receives a request command REQA from the reader. At that time, the reader changes its state from IDLE to READY (see Figure 3.4) and responds with an ATQA message.

At this point, the reader is now aware that at least one tag is within its range, and as a result it initiates the process of taking stock of all tags available. This is done by issuing a SELECT command which instructs all tags to respond with their UID serial number (see Section 3.1.1). Since more than one tag can be within range and may attempt to communicate with the reader at the same time, it is necessary at this point to conduct what is called an anti-collision process (more on this in Chapter 4), which ensures that eventually each tag will report its UID to the reader correctly. When a full serial number is received by the reader, it sends out a SELECT command followed by the full UID to specify the particular tag with which it wishes to establish communication. The tag recognizes its own UID and confirms the success of

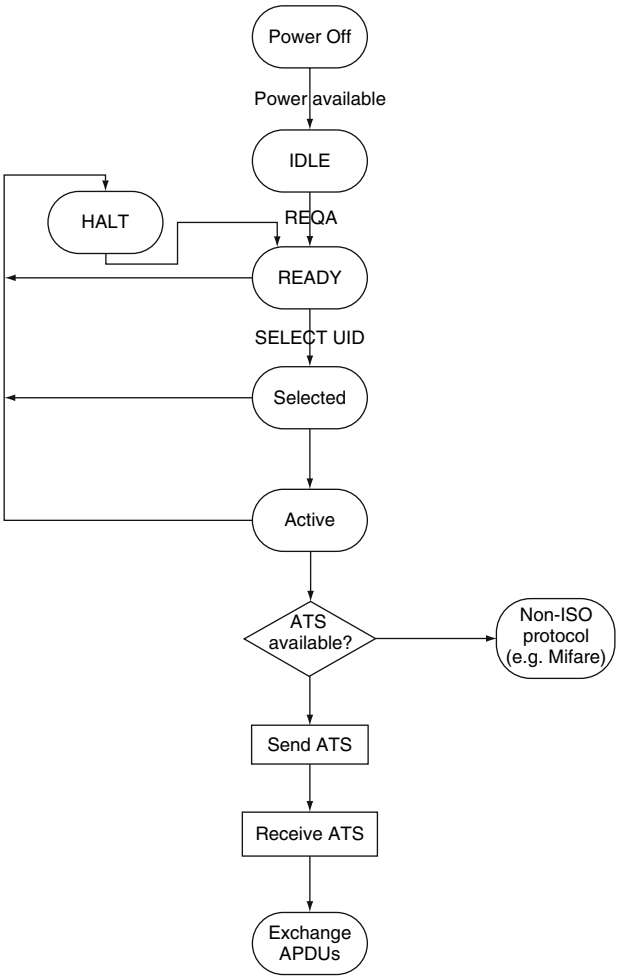

Fig. 3.4. State-transition diagram for ISO 14443A tags.

the command by responding with SAK (select acknowledge) and moves into the active state and is ready to communicate to the reader information stored in its memory or receive instructions to write to its memory.

In the example discussed in Section 3.1.1, the tag was using the MIFARE specification, a fact that is reported by the tag to the reader as part of the SAK response, so that communication can proceed. An alternative is that the transmission of information is compliant with ISO 14443-4. In this case, following the receipt of the acknowledgment to select, the reader issues a request for answer to select (RATS) command, and the tag responds with

an ATS which contains information required for effective communication; for example, possible baud rates and the maximum size of data blocks. This is followed by the actual exchange of application data via APDUs.

3.2.3 Tags that Use Capacitive Coupling

RFID systems using the far field of the carrier wave operate using a technique called backscattering rather than load modulation. This process is very similar to the operation of radar in that the tag reflects back a small part of the electromagnetic wave emitted by the reader. The reflection can be used to transmit information by examining the so-called reflection cross section, which is the signature of the component of the wave that has been sent back to the reader, and comparing it with the original. In practice, data are encoded by the tag by turning on and off the load connected to its antenna and thus shifting the reflection cross section between two clearly identifiable characteristic signatures. Similar to near-field RFID, also in this case there is a very considerable loss of power during the reflection process, and readers have to be sensitive to less than a microwatt in most cases.

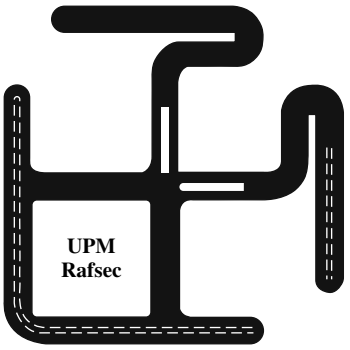

Fig. 3.5. A Rafsec EPC Gen2 tag operating at UHF frequencies.

Because of the involvement of the far field, the tag antenna is a dipole (see Figure 3.5). This fact can again be used to identify far-field tags via simple visual inspection. Far-field RFID commonly operates in the UHF band between 865 and 956 MHz, but note that the complete range is not available to applications globally (and there are also radically different signal power output limitations, especially between Europe and America). Instead, common far-field tags are able to respond in the complete range and it is the responsibility of the reader to select frequencies that are allowed within a particular regulatory region (typically 865–869 MHz in Europe, 902-928 MHz in the North America, and 950–956 MHz in Japan). Far-field systems allow for

longer-range communication and it is common to achieve distances between 3 and 4 meters using a patch antenna of approximately 30cm (more on patch antennas in Chapter 4). Using different antenna designs and power amplification, the range of such a system can reach up to 10 meters. More detailed descriptions of far-field RFID performance can be found in [22].

3.2.4 EPC Gen2 Tags

A typical example of a modern tag is the EPC Class 1 Gen 2 [22, Chapter 4], which operates at UHF frequencies. The chip has a relatively complex non-volatile memory structure divided into four distinct areas (see Figure 3.6). The reserved memory bank holds two 32-bit passwords, the "access" password for gaining access to the contents of the tag and the "kill" password, which when presented permanently disables the tag. The EPC memory bank contains the Electronic Product Code assigned to the object, location, or the asset on which the tag is attached and optionally related metadata. The Tag Identification bank contains information about the type and the manufacturer of the tag, including a unique serial number that identifies the tag itself. The user bank is optional and can be used freely by applications.

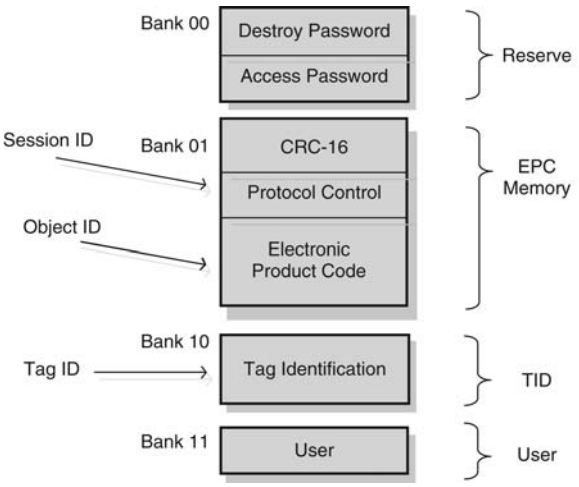

Fig. 3.6. Memory layout of an EPC Gen 2 tag.

Note that although we talk about radio frequency identification, a single tag holds several identifiers or codes that correspond to different functions and have distinct roles and semantics, including a fixed tag ID and a writable object ID. Tags often use a third identifier, the so-called session ID (in the case of Gen 2 tags, this is a pseudo-random number generated by the Protocol Control section), which is used by the reader to address the tag during a

particular session. The session ID is roughly equivalent to the MAC address of a typical wireless networking physical layer protocol, but in the case of Gen 2 it is only locally unique. Alternatively, the session ID may be fixed and stored in the tag memory, as is the case for ISO 14443 Type A tags. Note that tags that employ this approach can be easily traced using the session ID as a handler, a fact that raises very considerable privacy and security issues which we discuss in more detail in Chapter 9. For this reason, most recent tag protocols have implemented a randomization process whereby tags use a pseudo-random number each time they are interrogated by a reader so as to avoid easy tracing.

Compared with ISO 14443 tags, which we considered in Section 3.2.2, Gen2 tags have a much richer state space and command language. One way to look at this is that Gen2 tags are the RISC processors of RFID since they opt for a larger number of simple command primitives. This approach allows for simpler response logic on the tag, which in the case of Gen2 has particular importance due to cost implications but also because Gen2 is expected to be used in situations where a very high number of tags and readers will operate within the same environment. Due to this higher complexity of Gen2, we do not give complete details of its operation here but provide a simplified view that retains the flavor of this technology.

There are three families of Gen2 commands that the reader uses to manage the tag population. Note the emphasis on tag population management rather than communication with individual tags. The SELECT commands determine which groups of tags will respond to reader commands during a particular session; INVENTORY commands identify individual tags within a group and single them out for communication; and ACCESS commands relate to instructions sent to an individual tag to carry out specific operations, for example to read its memory contents.

A tag can be in one of seven different states but rather than go into the details of each and describe all possible pathways between states, we will only consider a simple scenario from the tag perspective (see Figure 3.7). As soon as the tag has adequate power for its operation it enters the READY state.[2] In this state it waits silently until it received a SELECT command which identifies its session as the current one but does not respond and returns to the READY state. Then the reader will issue a QUERY command, in which case the tag will check if its slot counter is 0[3] in which case it responds by transmitting

[2] There is an exception to this in the case where the tag has been issued the KILL command during a previous session, in which case it has entered the KILLED state. This state is non-reversible in that the tag is required to never exit, and it effectively means that the tag will not respond to any further commands. Note that this is implemented in most cases by a software mechanism rather than in hardware.

[3] The initial value of the slot counter is a random number selected by the tag following specific rules, and is decreased by one every time a new QUERY (or QUERYADJUST or QUERYREP) command is received.

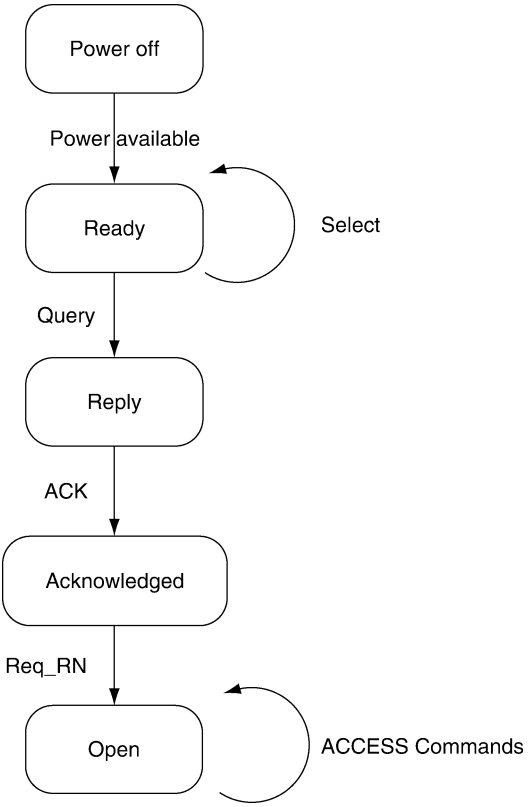

Fig. 3.7. Simplified partial Gen 2 tag state transition diagram.

a 16-bit pseudo-random number RN16. It also moves from the READY to the REPLY state.

The reader acknowledges the tag by returning a ACK command and the same RN16. On receipt of the acknowledgment, the tag responds with its PC and EPC data and moves into the ACKNOWLEDGED state. The reader confirms the receipt of the data with a *Req_RN* instruction which asks for a new pseudo random number from the tag that will be used as the handle for subsequent communications. The tag responds accordingly and moves into the OPEN (or SECURED) state, where it is now possible to access and modify its contents using one or more of the ACCESS family of commands.

3.3 Summary

In this chapter, we took a close look at the two core components of any RFID system, the read and the tag. We examined typical examples of readers and tag designs for the two principal classes of RFID, with particular reference to

common standards specifically ISO 14443 at the HF frequency band, which uses inductive coupling, and EPC Gen2 at the UHF band, which uses capacitive coupling.

4

Physics and Lower Layers

While this book is primarily about networked systems and is aimed at readers with a background in computing, it is nevertheless necessary to gain a concrete grasp of the implications of specific RFID technology choices and their implications for system design by looking "under the hood" into lower layers of the system. Such discussions are hardly ever an issue in modern computing, but this is not the case for RFID. Its significant performance limitations imply that a system designer cannot completely ignore low-level details and focus on high-level abstractions only. In fact, in most cases it would be necessary to experiment and measure the behavior of a system to tailor it to the particular conditions of the specific deployment environment. And of course real deployments often also imply evolving situations that require redesign or other system modifications.

In this chapter, we discuss the lower level issues that are important for the system designer. Readers with a background in physics or electrical engineering will almost certainly find these discussions elementary, but for most computing professionals they would be outside their domain of expertise. Specifically, we will look briefly at the characteristics that influence the propagation of electromagnetic waves and how these can be used to enable coupling and thus energy transfer, modulation techniques appropriate for RFID, antenna characteristics, anti-collision methods, and last but not least how all these affect the performance of readers. In all cases, we will take a practical approach and make specific reference to systems that employ these techniques.

4.1 Radio Frequency: Characteristics and Communication

RF, or radio frequency, refers to a specific range of frequencies (rates of oscillation) that correspond to alternating current electricity used to produce and detect radio waves. This so-called radio spectrum corresponds to frequencies

G. Roussos, *Networked RFID*, DOI: 10.1007/978-1-84800-153-4_4,

between 3 Hz and 300 GHz. Radio waves or radio signals are electromagnetic waves that have two components that depend on each other for their generation: a magnetic component often called the H-Field, and an electric component, the E-Field (see Figure 4.1). Radio waves are typically generated by an electric current alternating at a radio frequency flowing through a special type of conductor called an antenna.

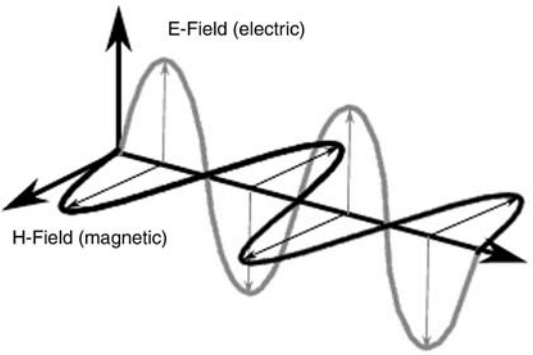

Fig. 4.1. Propagation of an electromagnetic wave.

Inductive systems

Looking at an antenna constructed as a conductor loop (see Figure 4.2), the effect of attaching a source of alternating current at its endpoints is that first a magnetic field is generated. Propagation of this field causes the generation of an electric field by induction. As a result, the field generated by this antenna, which was initially magnetic only, is transformed into an electromagnetic field. When the field reaches a distance of $\lambda/2\pi$, it starts to separate from the antenna and propagate into space as an electromagnetic wave.[1] The area around the antenna and up to the point where the free electromagnetic field is generated is called the *near field* of the antenna. Beyond this point, the electromagnetic wave has fully formed and separated from the antenna and the corresponding area is called the *far field* of the antenna.

As soon as an electromagnetic wave enters the far field, it can no longer affect the loop antenna. This implies that RFID systems that employ inductive coupling have an absolute range limit at this distance after which energy

[1] The distance $\lambda/2\pi$ is the wavelength of the electromagnetic wave λ divided by twice the transcendental number $\pi = 3.14159....$ The wavelength of a wave is the distance between two consecutive peaks in Figure 4.1.

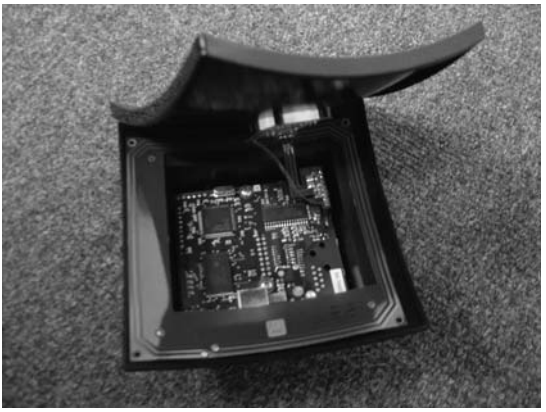

Fig. 4.2. A conductor loop antenna used by the Philips Pagoda reader, which employs the magnetic component of the radio wave for energy transfer and communication with tags.

transmission is no longer possible. For example, at the operating frequency of the ISO 14443 standard, which is 13.56 MHz the near field extends to 3.6 meters. At higher frequencies, for example in the UHF band employed by the EPC Gen2 tag, the near field is only 6 cm.

Another limitation of inductive coupling is the fact that the strength of the field generated by the antenna in the near field decays in proportion to $1/d^3$ with the distance d from the center of the loop. This rapid loss of strength also affects the transmission from the tag to the reader, resulting in a huge loss of power overall. Due to this fact, only a small proportion of the near field is actually usable, especially when taking into account regulations that restrict the power output of electronic devices, especially for consumer applications.

As a consequence, if the tag attempted to transmit at exactly the same frequency as the reader, the magnetic field it would be able to generate would be dwarfed by the signal of the reader, and thus it would be impossible to communicate. Instead, tags need to devise an alternative way of communicating with readers, and they achieve this by using a technique called *load modulation*.

With load modulation, the tag uses its integrated load resistor, which it switches on and off quickly, say at frequency f_s. The result of this is the transmission at two frequencies left and right of the frequency used by the reader (that is at $f_r \pm f_s$), called *sub-carriers* (see Figure 4.3 for the case of RFID systems using 13.56 MHz). This effect can be used to transmit information which is received by the reader by detecting small fluctuations in current at its own antenna.

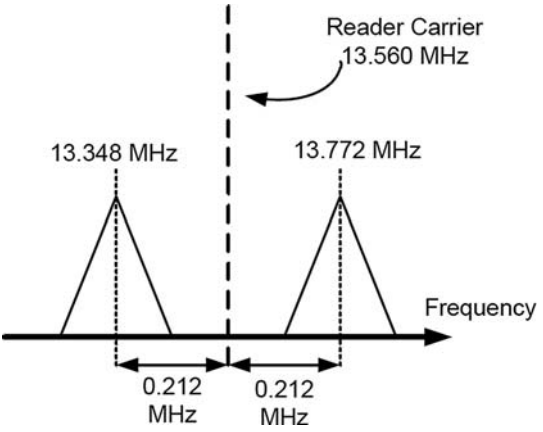

Fig. 4.3. Communication via load modulation using the sub-carriers of the reader transmission.

Capacitive systems

Using the near field of an electromagnetic wave to develop an RFID system is certainly feasible, but it has certain limitations, including a restricted read range and relatively low data transmission rates due to the rapid drop of the magnetic field. If longer distances are required, or indeed if a system should be capable of dealing with a large number of tags concurrently and exchanging more information, then alternatives must and have been explored. Clearly such alternatives should exploit the advantages of the far field to overcome these limitations.

Consider again the electromagnetic field generated by an antenna, this time shaped in the form of a *dipole*, depicted in Figure 4.4. As noted previously, after a short distance, the electromagnetic field separates from the antenna and propagates spherically into space, also transporting energy. During its movement, the wave encounters different objects, which interact with the wave. A proportion of the transmitted energy is absorbed by the object, and the rest is scattered toward different directions. Note that the electromagnetic field decays in proportion to $1/d^2$ with the distance from the dipole antenna.

The antenna of a tag can be constructed using two electrodes lying on a flat substrate. The electrodes have precise dimensions to match the particular frequency used, and in this way they can absorb most of the energy that reaches them. When the antenna comes in contact with the electromagnetic waves propagating from the dipole antenna of the reader, then an electric voltage is generated between the electrodes that can be used to power up and operate the chip.

A small part of the reflected energy will arrive back at the antenna of the reader, where it can be used to estimate the distance and the position of the

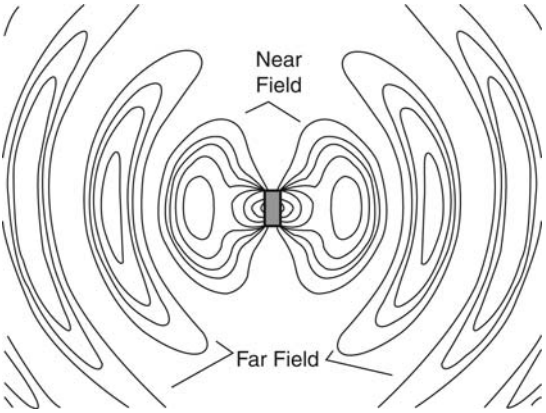

Fig. 4.4. Propagation of electromagnetic waves in the far field.

object. Moreover, the tag can modify its *reflection cross-section* to encode and transmit information. It achieves this by modulating the impedance of its antenna between two distinct values that can be clearly identified by the reader, thus encoding binary digits. One way that the reader can discover such backscattered waves is by comparing the returned wave against a reference wave that is identical to the one it originally transmitted.

The reflection cross-section is a measure of the ability of the object to reflect electromagnetic waves, and it depends on various parameters, including the object's size and shape, its material, and its surface, as well as the wavelength of the reader transmission. Tag antennas are carefully designed so that they correspond to specific cross-section profiles that can be operated robustly in specific modes. The energy loss of the reflected signal is also proportional to $1/d^2$ from the tag antenna, so when a tag transmission is received by the reader, only $1/d^4$ of the initial power employed by the reader remains. As a result, a typical RFID system using the far field would have a practical range of approximately up to 6 meters.

Finally, irrespective of the means by which energy is transmitted from the reader to the tag, the power available to the chip is extremely limited, typically of the order of 10 μW to 1 mW. It is not surprising then that chips used in RFID tags have extremely restricted functionality: compare this against the minimum power consumed by a typical embedded processor at standby, which is of the order of 500 mW.

4.2 Data Encoding and Modulation

In the previous section, we discussed how the electromagnetic wave transmitted by the reader can be modified by a tag to communicate information. In this section, we look at how the reader signal and the tag response can be

modulated to encode data. There are a great variety of techniques and alternative encodings that can and have been used to do so in the case of RFID, and we will consider some practical examples used for Gen2 tags.

Reader-to-tag communication

As noted early on in this book, a characteristic feature of RFID is that it is asymmetric. As a result, communication from the reader to the tag can differ considerably from the techniques used for communication from the tag to the reader. For example, in Gen2 systems, the reader uses Amplitude Shift Keying (ASK) modulation and pulse interval for the data encoding (PIE) to achieve speeds between 26 and 128 Kbits/s. Both techniques are very simple and are selected so that tags are not overburdened by the decoding of reader instructions.

The ASK scheme in this case is simply the use of variations in the amplitude of the electromagnetic wave emitted by the reader to represent high and low values corresponding to binary digits (see Figure 4.5, bottom) encoded using PIE (see Figure 4.5, top). PIE encodes binary data using timed slots, with a zero taking up two slots (high followed by low) and one lasting twice as long (three high slots followed by one low slot). For each message sent, PIE also requires an initial preamble, and all subsequent transmissions must be synchronized. It is not necessary to mark the end of transmission with an end-of-file code.

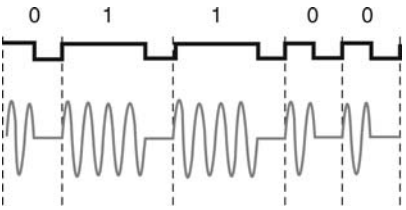

Fig. 4.5. Data encoding and modulation for communication from reader to tag using Pulse Interval Encoding (top) and Amplitude Shift Keying modulation (bottom).

Tag-to-reader communication

Due to the fact that tags use load modulation or backscattering to communicate with the reader, they are particularly constrained in their choice of data encoding and modulation schemes, which have to be especially lightweight. For example, in Gen2 systems, the actual modulation scheme employed by a particular type of tag is selected by the manufacturer at production time and

can be either ASK or Phase Shift Keying (PSK). All tags must support two data-encoding methods, namely FM0 Baseband and Miller Sub-carrier (also known as delay encoding), since at the beginning of any session, the reader may instruct the tags to use either method. The data transfer speeds that can be achieved by each alternative approach are 40 to 640 Kbits/s and 5 to 320 Kbits/s respectively.

Similar to ASK (see Figure 4.6, middle), PSK is a basic modulation scheme that modifies the phase of the electromagnetic wave emitted by the reader to represent high and low values (see Figure 4.6, bottom). Using the Miller sub-carrier encoding binary digits are represented using the following rules (see Figure 4.6, top):

1. A zero digit does not affect the signal level unless it is followed by another zero. In that case, at the end of the current time slot, a transition to the opposite level takes place.
2. A unit digit causes a transition to the opposite level at the middle of the current time slot.

Miller encoding is often used in wireless communication because the encoded signal requires less energy than other schemes, including the simpler non-return-to-zero (NRZ). It achieves this by not adding extra bits (for padding or error control) to the original data stream.

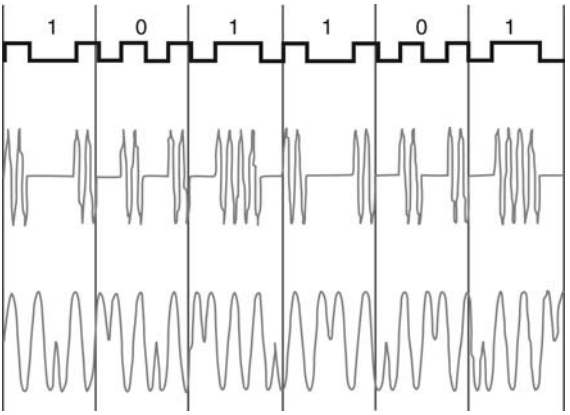

Fig. 4.6. Data encoding and modulation for communication from tag to reader.

4.3 Antenna Performance

The antennas of the reader and tag have a critical role to play in any RFID system, as they are not only responsible for communication between the two

devices but must also efficiently radiate and harvest energy, respectively. Taking into account the massive loss of energy during transmission and the very low levels of power that are available to the tag chip anyway, good antenna design is necessary to achieve robust and error-free system operation. To this end, most RFID antennas are manufactured using a highly conductive material such as copper or conductive ink, which are sensitive to electrical and magnetic currents found in radio waves. Moreover, an RFID antenna is almost always tuned to operate within a certain frequency band, the so-called *resonant frequency*, and all aspects, including its construction materials, its length, and its structure, are precisely chosen to optimize efficiency.

As already noted, tags that operate in the near field of the reader antenna make use of the magnetic component; for example, those compliant with the ISO 14443 standard, which operate at 13.56 MHz. This target frequency directly influences the design and construction of the tag antenna which in this case is a coil (see Figure 2.2) similar to the reader antenna (see Figure 4.2). To improve the capability of an inductive RFID, there are two alternatives that can actually be combined:

- increase the length of the reader and the tag antennas (while ensuring that the antennas are still tuned to the resonant frequency) and
- use tag antennas that employ several loops, in their coils, the more loops the more prominent the coupling effect.

On the other hand, RFID tags that employ the far field carry antennas that are usually manufactured as straight lines or dipoles, so that they can best utilize the electric component of the electromagnetic carrier wave. The most basic type of far-field antenna is the *half-wavelength dipole antenna*, which consists of two components, each a quarter of the wavelength of the resonant frequency. For instance, a half-wavelength dipole antenna for Gen2 tags operating in the 915 MHz band would require each dipole to be 8.2 centimeters long (that is, one quarter of the 32.8 centimeter wavelength). To increase the tag's capability to harvest power in the far field, one common approach is to increase the surface of the antenna, which will be able to retrieve more energy.

Readers operating in the far field could also use a similar dipole antenna, but in most cases they employ a more efficient *patch antenna*. A typical half-wavelength patch antenna would consist of a metal patch of length half the wavelength of the resonant frequency, suspended over a ground plane, with the complete assembly contained in a plastic radome for protection (see Figure 4.7).

Last but not least, tag performance also depends on the alignment of tag and reader antennas. Both near- and far-field systems completely fail if tags are placed perpendicular to the reader antenna. When the angle of orientation is less than 90 degrees, there is still loss of performance, the magnitude of which depends on the particular angle. This problem can be addressed to some extent by using multiple reader and tag antennas in complementary

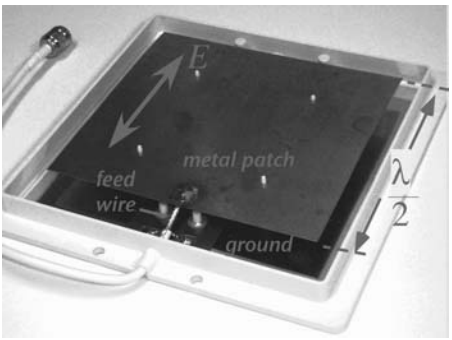

Fig. 4.7. Patch antenna for far-field readers.

orientations and alignments so that the likelihood of a tag placed completely perpendicular to all antennas is minimized.

Fig. 4.8. A UHF tag with a "wiggle" antenna.

Tag antennas often have an unusual appearance because they attempt to be offset at abrupt angles that allow the tag to present some part of its antenna to the radio wave at an angle most conducive to coupling, whatever its orientation. The typical half-dipole antenna is very efficient with the right orientation, but its performance drops rapidly when it is oriented unfavorably. Several of the tags displayed in Figure 1.3 and Figure 4.8 have exactly such usual antenna designs, so that they can be operated well irrespective of their orientation. Another tag design has been developed with a significantly larger conductive area that ensures that it is capable of harvesting higher amounts of energy. Its characteristic antenna shape has given this type of tag its nickname as the so-called DogBone tag (see Figure 4.9).

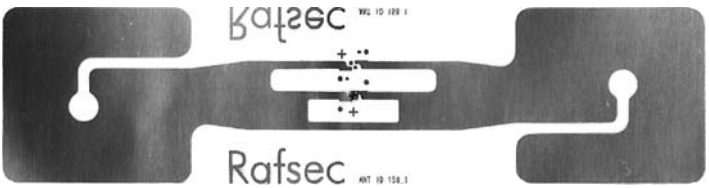

Fig. 4.9. A DogBone UHF tag.

4.4 Anti-collision and Singulation Techniques

Finding efficient ways to share the air medium is a particularly acute problem for modern RFID, as larger systems imply higher tag and reader densities and also the need for faster data transmission. Since more and more readers and tags will have to share the medium, the likelihood of errors due to collisions also becomes higher. It is thus necessary to provide more effective coordination protocols that allow readers in particular to avoid mixing their transmissions, to be able to pinpoint and isolate their communication peer, and to maintain several concurrent channels while guaranteeing high performance. Clearly this is an issue solely for tags that use the far-field as the short range of inductively coupled RFID systems does not allow for high tag or reader densities or require high data rates.

One technique employed between readers to avoid each other is Listen Before Talk (LBT). In LBT, before a reader transmits, it checks on the channel it intends to use for other transmissions. If it detects a signal, then it is obliged to switch to another channel and repeat the process until it discovers a free channel. Moreover, when a free channel is discovered and used for communication, the reader is still obliged to turn its transmitter off for 0.1 seconds every 4 seconds so that other readers can also gain access to that channel.

Nevertheless, it is still possible to have many more readers colocated and interfering with each other compared with available channels. In this situation, LBT is not enough and more complex mechanisms must be employed so that all readers can carry out their designated tasks. There are two main techniques to achieve this so-called Dense Reader Mode (DRM):

- *Time synchronization*: under this scheme all readers transmit simultaneously and then listen for tag responses while their carrier wave transmissions remain active.
- *Frequency separation*: under this scheme readers transmit on even-numbered channels, while tags respond on odd-numbered channels. Adopting this mode of operation implies that the strong reader signals do not mask the weak backscatter from the tags. Tags do not select their frequency of operation but respond to the strongest reader signal they receive.

Both LBT and DRM make use of specific channels to organize communication, and so it is interesting to look at how the spectrum available to RFID is organized. Table 4.1 shows the frequency bands available to EPC Gen2 tags in various regions or countries, the allowed transmission power, and the number of allocated channels. This table indicates the great flexibility that Gen2 tags must show: European regulations allow for the use of 10 channels of 200 kHz each at 2W effective radiated power (ERP)[2] power output while North American regulations permit frequency hopping between 52 channels of 500

[2] This is equivalent to 3.2W effective isotropically radiated power (EIRP). EIRP is in common use outside Europe.

kHz each and at 4W EIRP output. This implies that the data rate of communication between readers and tags will be much less in Europe (and also India and Japan), by as much as 30% of what can be achieved in North America (500 versus 1500 reads per second for EPC Gen2 tags). For applications that depend on immediate reading of very large numbers of tags in motion, as in the case of pallet loads passing through dock doors for instance, this would result in a limitation of the speed with which they can be transported.

Table 4.1. Available operating frequencies and channels for UHF RFID in different regions of the globe.

	North America	Europe	Japan	Australia	India
Band (MHz)	902–928	866–868	952–954	918–928	865–867
Power	4W EIRP	2W ERP	4W EIRP	4W EIRP	4W EIRP
Channels	50	10	TBD	16	10

Avoiding collisions between reader transmissions solves only part of the congestion problem. It is also necessary to devise mechanisms by which transmission between tags is also avoided. In this task, RFID has a major advantage, as unlike in other wireless communication systems, the reader can orchestrate the order in which tags respond. The majority of modern RFID systems address this requirement by employing some variant of the slotted ALOHA protocol to support anti-collision and singulation techniques that allow tags to be accessed in an ordered way and to isolate specific tags that can be addressed in isolation [31].

Gen2 tags diverge from this format and provide an interesting alternative algorithm called the Query Slot Protocol (QSP), which we briefly discuss here. Note that in designing the QSP a core requirement was that the reader never transmits the unique identifier of a particular tag during a communication session. The reason for doing so is that this mode of operation reduces the number od possible opportunities for attack on privacy, as otherwise the strong reader signal would propagate farther and effectively broadcast the code in cleartext.

Every Gen2 tag must provide several facilities, especially to support singulation, including an internal slot counter, a flag to indicate when it is in `inventoried` state, and a 16-bit random number (RN16) generator. The first step in the process requires the reader to transmit the parameter Q, which upon receipt causes all tags to clear their inventoried flag and generate a random Q-bit number,[3] which they load into their slot counter. If the slot counter of a tag is zero, then it backscatters an RN16. The reader would acknowledge receipt (at that time, the tag would set the inventoried tag and go to sleep mode) and initiate another round, whereby all remaining tags will decrement

[3] A number between 0 and $2^Q - 1$.

their slot counter by one and check if their slot counters are zero. If after $2^Q - 1$ rounds all tags have been inventoried, then the process completes. Otherwise, the reader will detect a collision and respond by transmitting a new parameter Q and resetting the process.

Clearly the efficiency of this process depends on estimating a good choice for Q—if it is too big then there will be many slots empty, and if it is too small, there will be many collisions. Q is estimated experimentally by the reader, which modifies its value until appropriate performance is achieved. After each tag has been inventoried, the reader can instruct specific individuals to communicate with it in an orderly manner.

4.5 Sources of RFID Read Errors

There are two types of common read errors in RFID systems:

- *negative reads* are situations where a particular tag is located within the vicinity of the reader, but the reader is unable to communicate with it, and
- *positive* or *ghost reads* are cases where a reader appears to retrieve a tag identifier although no such tag is within its vicinity.

The source of these errors varies but all systems must be able to deal with them. In particular, in Section 4.3, we identified several cases when a tag may not be positioned in such a way that it can be interrogated. For example, when a tag is moving at a relatively quick pace through an interrogation region of a particular reader, it may be positioned in a favorable orientation only for a short period of time during transit. As a result, to ensure that this tag will be observed, it is necessary that the reader continuously scan with a view toward discovering new tags. If during this time a tag appears only on a few occasions, then the reader must estimate the likelihood that the tag is actually located within its vicinity or not.

In the recent past, a large proportion of ghost reads were the artifacts of inefficient singulation protocols, a source of errors that today has almost been eliminated. Nevertheless, it is not uncommon that due to peculiarities of a particular environment, tags that are located farther away than the distance specified by the operating parameters of the reader can still be read, presenting their codes when they should not.

Figure 4.10 shows how negative RFID reads can occur due to interference and tag-response collisions. The former situation is due to the operation of an external RF source (for example, a second reader or other wireless network) or metallic material in the vicinity, and the latter can be the result of collisions between two tags attempting to respond to the same command. Such problems are relatively straightforward to deal with, as they lead to complete failure.

Writable RFID systems also face the additional problem of indeterminate situations. These occur when a write command has been issued but not confirmed and the tag is no longer in range. As a result, it is not possible to verify

the success or failure with a subsequent read, and thus it is not possible for the reader to determine the final state of the tag.

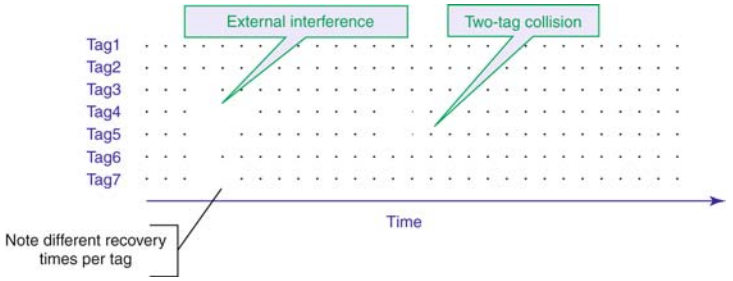

Fig. 4.10. Negative tag reads due to interference and RF collisions.

Another significant source of errors in the case of far-field systems is the presence of specific materials that cause reflections and thus detuning of the reader signal. For example, when metals are placed between the reader and tag, in most cases they would completely absorb the signal. A similar result would be caused when metal artifacts with size approximately equal to half the wavelength of the resonant frequency are present nearby, as they would effectively act like antennas and absorb a considerable proportion of the signal. Errors can also be due to reflections and scattering of the reader signal on surfaces placed within its range.

Different operating frequencies and signaling technologies have different sensitivities to different materials: near-field systems are mostly unaffected by dielectric or insulator materials (for example paper, plastics, masonry, and ceramics), but metals weaken the field (depending on how ferrous they are) and may also detune tags if they work at a resonant frequency. Far-field systems can penetrate dielectric materials, but water molecules absorb signal energy and metals reflect or scatter the signal to such a degree that they can potentially completely cloak tags. Moreover, at UHF frequencies, tag-on-tag effects (like those depicted in Figure 4.10) are particularly strong due to the higher tag densities and increased reader range.

4.6 Summary

In this chapter, we discussed issues related to the operation of lower layers of RFID from the perspective of computing. Our objective has been to understand the trade-offs involved in selecting and implementing different alternatives, especially the central role of good antenna design for robust system operation. We also identified the different types and common sources of read errors to set the context for subsequent chapters, where we discuss data cleaning and filtering techniques.

5

Identifier Systems

RFID has been used for the identification of natural and manufactured objects, locations, humans and other species, and more recently services. As a result of the great diversity of applications of this technology, there is no single coding scheme that provides a universal system of identifiers but rather each domain has developed its own approach, often addressing the specific requirements of its particular context of use. In some cases, these identifier systems have a relatively long history that predates RFID and have been used in alternative realizations, commonly as bar codes. This chapter starts with a brief review of legacy identifier schemes used within specific application domains and proceeds to discuss modern systems of universal identification that are appropriate for open network RFID systems.

Note that in this chapter the focus is on so-called *pure* identifier systems that is, on the code structure itself rather than on how the identifier is encoded using a particular RFID technology. This distinction between representation and encoding was introduced by GS1 in the 1980s in the context of bar codes (discussed in more detail in Section 5.2) and has allowed the consistent evolution of identifier schemes despite changes in implementation technologies.

5.1 Application-Specific Identifier Schemes

Application-specific identifier schemes are used today in the majority of RFID system implementations. These schemes are often proprietary and operated by a single organization or exclusive in that only a small number of selected participants have access to the system. The code structure in this case is designed to facilitate the particular application, and it is uncommon to make provisions for its use outside its selected domain. For example, a metropolitan transportation system that employs RFID would encode individual ticket serial numbers on the tag that are meaningful and valid only within the boundaries of the system but worthless outside this use context. The specific ticket serialization specification would be internal to the issuing organization and would be

G. Roussos, *Networked RFID*, DOI: 10.1007/978-1-84800-153-4_5,

tightly linked to associated billing and access-management applications, which are more often than not also specifically designed for the particular context of operation for example, local data processing regulations.

Moreover, such application-specific numbering schemes are in most cases inaccessible to either users or third parties, as they are cryptographically protected with some type of symmetric scheme, and with good reason. In any case, even if access to the internals of the coding scheme is possible, since this is fully controlled by the issuer, it would still have very limited use in external applications as there are no guarantees about either the permanence of identifiers or their consistent use.

There are numerous examples of application-specific schemes, notably those used in closed supply chains and for e-passports. Closed supply chains are operated specifically for the benefit of a particular retailer, who often has exclusive relationships with suppliers. In this case, product and container numbering can be done in isolation, with retailers able to choose the coding scheme and all supporting technologies in a way that best fits their requirements. Of course, this would most likely preclude the system from inter-operating with open supply chains in the future, but in many cases this is not desirable anyway.

In fact, proprietary systems can offer a significant cost advantage compared with standards-based approaches. This is due to the use of less powerful technology which is often adequate for the generally simpler requirements of a single application. For instance, it may not be necessary to fully implement standards that offer features that are complex but not appropriate for the specific retail use. Significant cost savings can also be realized in this case by removing the need for licensing or subscription fees related to the acquisition of a dedicated section of the identifier address space for use by the system.

A compromise in the proprietary versus open system question is offered by certain standards-based but application-specific schemes that offer a useful trade-off. For instance, a common use of RFID is in animal tracking, where ISO 11784/1 specifies the operation and the code structure of standard identifier codes. Table 5.1 shows the structure of such a 64-bit ISO identifier, although the actual semantics of particular codes are not fully defined (bits 27–64). Clearly such codes have little application outside animal tracking but within this sector provide a simple but convenient way to organize identifier code allocation.

Another implication of the flexible definition of this coding scheme is that simply acquiring the identifier would not provide the full information about the animal tagged, as it is not possible to differentiate two animals that correspond to the same general categories as classified by the nation identifier segment. Since this segment also differs across countries, the codes are hard to interpret when retrieved outside their country of origin and can hardly be thought of as universal identifiers.

Indeed, although application-specific codes offer distinct advantages, it is also true that they are generally unsuitable for open network environments.

Table 5.1. Code structure for animal tracking applications following ISO 11784.

Flag (1 bit)	Specifies use for animals or other application.
Reserved (14 bits)	Reserved for future use.
Data block (1 bit)	Specifies that additional data will be transmitted.
Country code (10 bits)	According to ISO 3166.
National identifier (38 bits) (Country specific)	Animal type, breed, region, breeder. (exact combination depends on national norms)

Due to their focus on specific application requirements and optimal performance in that context, they often do not scale. In most cases, this would be either because the code structure restricts the use of the scheme to relatively low tag numbers or because it cannot support effective management of the code space as there is no mechanism for the allocation of complete sections of codes to particular users in a manner similar to the way internet addresses are assigned to autonomous systems. This fact has a severe impact on the ability of organizations to independently manage tagging of objects. To be sure, such administrative facilities are a core feature of any successful universal identifier code scheme.

Finally, codes employed in exclusive or proprietary systems have semantics that cannot be easily interpreted following a simple set of rules. This implies that automated processing of tag data is difficult and in general cannot be carried out outside the system boundary. Since common semantics cannot be established, the exchange of data is also inefficient and often restricted by the need for manual intervention. This lack of coordination often leads to subsegments of the address space being developed independently toward incompatible directions and prevents future interoperability and convergence, as highlighted by the code structure of Table 5.1.

5.2 Pre-RFID Universal Identifier Systems

The discussion of application-specific codes highlights the role of two elements that any general-purpose identifier system should provide:

- open code semantics, standardized rules, and automated mechanisms for identifier code translation, which can support comprehensive use across application and authority domains, and
- supporting administrative structures that facilitate efficient code allocation, in particular the delegation of responsibility for the independent management of specific code sub-domains to particular users and the capability of assigning serials to particular items locally.

While the former is primarily a technical matter, the latter depends on the complex network of international standards organizations and the adoption

of common processes. As such, in the development of such a system it is reasonable to attempt to exploit existing administrative structures, and to use past experience in managing automatic identification technologies of global scope whenever possible.

5.2.1 Universal Identification with GS1 Bar Codes

The most common and most comprehensive open automatic identification system was developed around bar codes. Although early versions of the GS1 system (One Global System) developed in the 1970s were tightly linked to printing and scanning bar codes, subsequent developments disassociated this task from the numbering scheme proper (for a brief history of bar code systems, see [100, Chapter 2]). Today, GS1 provides an abstract universal identifier system that can be implemented on a variety of carriers, including of course bar codes and RFID. By way of introduction to universal identification, this section examines the structure and operation of codes developed under GS1, the rationale for the adoption of this particular approach, and last but not least, its limitations with regard to its potential use as a general-purpose system of identifiers. Note that until recently GS1 was also known as the EAN/UCC system, and parts of the standard still retain this name.

Consider a typical bar code, for example the one displayed in Figure 5.1(a), which follows the EAN-13 specification.[1] This bar code is simply a symbol that encodes a string of 13 decimal digits in a visual representation. The number encoded by this symbol is a unique product identifier constructed following the rules of the 13-digit version of the Global Trade Item Number (GTIN-13) specification. The GTIN-13 standard defines the code and the EAN-13 standard how the code is laid out in print. The EAN-13 symbol can be read into a computer system using a (portable or fixed) low-power laser scanner, which translates the sequence of white and black bars into the corresponding digits.

The GTIN-13 code follows a scheme designed to ensure that each number assigned to a product line is unique and also includes a unique code that identifies its manufacturer:

- the first seven digits specify the GS1 *company prefix* that identifies the manufacturer, the importer or wholesaler, or the retailer of the product,
- the next five digits represent the *product code*, which identifies a product line, and finally
- the last digit is a *checksum* used by acquiring computer systems to confirm that the code has been retrieved correctly.

The company prefix, which is also known as the manufacturer code, is assigned to the particular business by GS1, while the digits corresponding to

[1] The EAN specifications define several similar identifier schemes of different lengths, with the 13-digit version being the most common in current use.

(a) EAN-13 bar code of identifier 978-1846280351.

(b) ISBN-10 bar code of identifier 1-846-280-354.

Fig. 5.1. Two bar codes encoding numbers that refer to the same object under two different identifier systems. The object identified is the book of reference [100].

the product code are selected by the manufacturer. There are two exceptions to this rule: when a retailer sells products under their own brand (which in fact may be manufactured by a third party on behalf of the retailer) and when the product does not have a bar code assigned at the time of manufacture, for example if it is imported from a country outside the GS1 system.

Note that GTIN identifiers do not contain product classification information in the code: information about the industrial sector, the country or the region where the product was manufactured, or the type of product (for example clothing, food, electronic device, and so forth) cannot be retrieved from the GTIN number. GTIN is a simple unique identifier akin to a key in database parlance, and to obtain information associated with the product, it is necessary to query a related product information repository. Moreover, the unique identifier characterizes the product line; for example, code 5038862382601 refers to any 1-liter carton of pomegranate, blueberry, and acai smoothie produced by Innocent Drinks rather than a particular product item; for example, the specific carton of smoothie that was produced at 12:15:01 on January 1, 2007 at their Fruit Towers facility.

5.2.2 Beyond Product Identification

To be sure, there are many bar code symbologies, several of which operate outside the GS1 system. Some varieties carry information in addition to a simple identifier; for example, the expiration date of the product. GS1 also provides a wider collection of numbering standards, in addition to GTIN, designed to deal with specific environments, including pallets, locations, and returnable assets. For example, the Global Location Number (GLN) specifies a numbering scheme very similar to GTIN in code structure, which can be used to identify locations. Codes similar to GTIN codes are assigned to locations by a specific manager (identified in the company prefix) and printed as EAN-13 bar codes.

Resolving GTINs and GLNs to product descriptions or locations respectively requires a lookup in a global repository that maps a code to its representation. In the case of the GS1 system this resolution service is provided by the Global Data Synchronization Network (GDSN), which is also specified with the GS1 system. GS1 provides not only the standards and specifications for the participation in this service but also oversees and maintains the service infrastructure overlaid on the public internet.

In addition to GS1 identifiers, particular sectors have developed and operate long-running and very successful schemes that are used almost exclusively within the sector; for example, all prescription and over-the-counter drugs in the United States must be labeled with a bar code displaying their National Drug Code (NDC) number. The vast majority of books are identified using a 10-digit International Standard Book Number (ISBN) rather than a GTIN code, which is also guaranteed to be unique within the ISBN system. In this particular case, ISBN and GTIN have recently come to an arrangement to merge their systems by prefixing ISBN codes with a special three digit code and integrating them within the GTIN address space.

5.2.3 Limitations of GS1 Codes for Item-Level Tagging

Despite its popularity, the GS1 system has some limitations that restrict its utility as a general-purpose universal identifier scheme. First, it carries the legacy of its primary context of use, which gives priority to the requirements of supply chain management and retailing, and has limited provisions for object registration outside the scope of such applications. For example, a central part of the specification is developed around the role of the manufacturer to control the numbering scheme to the exclusion of subsequent custodians of the object, and there are no facilities to support the use of the scheme with natural objects or with hand made or custom-designed product items.

Furthermore, the standardization process is carried outside recognized international standard-making bodies through the GS1 organization, which was created and is controlled by commercial companies with a vested interest in supply chain applications (mostly for consumer goods). This orientation of GS1 toward the private interests of these entities raises considerable questions of fairness and ownership, and thus it has limited credibility as the guardian of a scheme of such importance. Notably, GS1 has only a limited relationship with ISO, for example, and indeed its most basic bar coding schemes did not become part of ISO until more than 20 years after their development. Nevertheless, it is not possible to discount the importance of the GS1 system since the popularity and the extensive availability of codes on the majority of manufactured products creates a de facto standard that dominates the market.

Finally, when examined from a unique object-identification perspective, the majority of bar code specifications share one fundamental limitation in that they are unable to distinguish between one instance of a particular product line and another. As noted earlier, all cartons of a particular type of

smoothie drink of a specific size and packaging for example, are assigned the same code and hence are indistinguishable from one another. Clearly this places severe limitations on the capability to identify specific objects, as each item cannot be differentiated from any other but rather product families are treated as single units. This shortcoming has been addressed by recent GS1 specifications that cater to item-level tagging, and it has now become possible to uniquely identify individual objects and items that are assigned their own so-called Electronic Product Codes.

5.3 Electronic Product Code

The most successful RFID identifier scheme in terms of industrial adoption is beyond doubt the Electronic Product Code. EPC has been developed as an independent activity within the GS1 system under the EPCglobal brand, as in some ways it breaks with the GS1 tradition. EPC focuses on RFID specifically, but unlike other existing RFID standards that concentrate on the how, EPC offers an extended set of standards that define both how and what data will be stored on the tag, as well as stipulating the complete life cycle of code production, capture and processing. Separate specifications describe the tag memory layout and communication with readers (see Section 3.2.4), and the composition and layout of the EPC universal unique identifier scheme (discussed in this section), which employs only some features of existing GS1 schemes.

5.3.1 Serialized Global Trade Identification Number

The main type of EPC identifier code is the Serialized Global Trade Identification Number (SGTIN), which is an extension of the GTIN code with a serial number that pinpoints a particular item within a product line. This addition removes the main limitation of the GS1 scheme for universal unique identification discussed earlier and capitalizes on the higher holding capacity of RFID and some of the more recent bar code symbologies.

SGTINs come in two versions that have similar structure but support two distinct identifier code lengths of 96 and 198 bits respectively. SGTIN-96 codes are the most common and while they may not offer an address space of adequate size to tag objects at a truly global scale, they have been introduced as an interim solution due to their considerable cost benefits. SGTIN-96 codes consist of six parts, namely:

- the header, which identifies the tag as an SGTIN-96 (8 bits),
- the filter value, which allows the pre-selection of the object type (3 bits),
- the partition, which indicates the split of the last 82 bits between the remaining three fields (3 bits),

HEX	30700048440663802E185523
Binary	00110000011100000000000001001000010001000000001100 11001000000000000101110000110000101010100100011
URN	urn:epc:tag:sgtin-96:3.0037000.06542.773346595

Filter	Company Prefix	Item reference	Serial Number
3	0037000	06542	773346595
Shipping Unit	P&G	Bounty Paper Towels (15 pack)	Item UID

Fig. 5.2. Example of an EPC SGTIN-96 tag and its decoding. The top table shows the actual forms of the EPC in different stages of the encoding process and the bottom shows the interpretation of the SGTIN-96 identifier in particular.

- the company prefix, which contains the GS1 company prefix (20–40 bits, variable depending on partition code),
- the item reference, which contains the GTIN reference number and identifies the product line (4–24 bits, variable depending on partition code),
- the serial number, which is the unique identifier of the specific tagged item (38 bits).

Following common practice within GS1, the header, filter, partition, and company prefix sections of the EPC are provided by GS1 so that their use and assignment are coordinated and guaranteed to be uniquely defined, but the item reference and serial number are assigned by the manager or else the manufacturer of the product. Table 5.2 shows an example of an EPC encoding an SGTIN-96 identifier, in binary, hexadecimal, and as a Uniform Resource Name (URN), and its interpretation. Note that the company prefix segment of the EPC is often also referred to as manager code.

5.3.2 Other Types of EPC Identifier Codes

In addition to product items, EPC also provides schemes for tagging other types of resources. The Serialized Shipment Container Code (SSCC) is used to identify shipping containers, for example boxes, pallets and other stock-keeping units (SKUs), and the Global Returnable Asset Identifier (GRAI) is used to tag returnable assets. These would typically be returnable packaging or transport equipment. The SSCC is particularly significant for supply chain applications and follows the common structure with the notable exception that its serial number segment is defined by the standard GS1 systems. A similar structure is also followed by the final two types of identifiers, called Global Returnable Asset Identifier and Global Individual Asset Identifier (GIAI). Finally, EPC provides for two additional types, which are defined outside the

GS1 system, namely the resource codes defined by the Department of Defense specification for military supply chains and a general-purpose type predictably called General Identifier (GID-96), which is a catchall for other uses of the EPC tag specifications.

Tagging of locations is supported with the Serialized Global Location Number (SGLN), which is a serialized form of the Global Location Number (GLN) defined within the standard GS1 system. GLN provides serial numbers allocated by the company that issues the code that correspond to internal company locations and are not necessarily available to external parties. SGLNs follow a structure very similar to SGTINs with header, filter, partition and company prefix. The last past of the GLN is the location reference, which is a number the semantics of which are at the discretion of the manager. Since these numbers cannot be interpreted without access to their definitions, it is necessary for a company to publish the appropriate correspondence in appropriate repositories, which often take the form of a public internet service, either the so-called EPC Information Service discussed in Chapter 8 or the GDSN.

One point that sets GLNs apart from other symbolic absolute location systems is that they define a rather extended concept of location in addition to physical places, which in the context of the supply chain would often be stores, warehouses, manufacturing plants, warehouse gates, loading docks, or vending machines. GLN also includes within its scope legal (for example, companies, subsidiaries, or divisions) and functional entities (in most cases these would be departments within the company, for example accounting or fulfillment). In any case, this unique identifier can be encoded in an RFID tag that can be automatically read by interrogators within its vicinity, which can subsequently resolve this information through the GDSN and thus discover its location.

5.3.3 Allocation of EPC Codes

As noted earlier, the second ingredient for the effective operation of an identifier system is the administration of the code allocation process. In the case of EPC, this is managed solely by GS1, through a separate membership scheme offered under the EPCglobal activity. GS1 maintains a worldwide network of national organizations that coordinate both the promotion of its standards and the allocation of EPC manager codes.

The GS1 organization has a relatively long history of intimate involvement in the development of bar code standards, business messaging, and other information technology solutions for the supply chain. Until recently, GS1 consisted of a highly heterogeneous collection of regional industrial associations, including UPC in North America, EAN in Europe, and JAN in Japan, which maintained divergent and incompatible versions of bar coding standards in particular. To address this situation and the increasing pressures of global supply chains, these organizations merged in 2005 to form GS1, which now

Table 5.2. ISO/IEC 15459 universally unique identifier example.

Data Identifier	Issuing Agency Code	Company	Serial Number
25S	LE:EDIFICE	E999	C204060897294374

offers a fully global, although still decentralized, registry that maintains ownership of the EPC scheme, among others. Any user or supplier that wishes to acquire EPC codes has to seek permission via membership inEPCglobal, which implies incurring quite significant costs.

5.4 ISO Standards

A direct competitor to EPC is the ISO 15459 specification on unique identifiers, with provisions on registration (Part 2), common addressing rules (Part 3), transport unit address provisions (Part 1), and item-level tagging for the supply chain (Part 4). Similar to EPC, under this scheme a universally unique identifier is associated with an object by its manufacturer at production time. Rather confusingly, ISO codes also build on relevant standards developed originally for bar codes, in some cases the same GS1 standards ratified under ISO provisions. Nevertheless, ISO has adopted a distinct approach that sets it in contrast to EPC in that it opts to integrate the majority of existing standards and identifier systems under its provisions rather than break with the past.

Following other related ISO standards, each part of the code holds alphanumeric digits rather than numbers as is the case in EPC. ISO 15459 codes have four parts (see Table 5.2 for an example):

- a data identifier (DI) header, which specifies how the contents of the identifier code are constructed and their meanings,
- an issuing agency code (this does not exist in EPC),
- a company ID, which is equivalent to the company prefix of EPC, and
- a serialized item code, which incorporates the item reference, for example the code of the product line, and the item serial number.

The DI in particular is roughly equivalent to the header and partition sectors of the EPC. However, in keeping with ISO practice—and contrary to EPC which introduces new rules—it reuses the ISO 15418 specification for the interpretation of codes and the data encodings provided by ANSI MH 10.8.2. For example, setting the DI to 25S specifies that the object ID is a globally unique serial object number, while a DI set to 2L specifies that the object ID is a location specified in a format defined in a subsequent field, for example a postal code.

This approach allows for continuity and a smooth shift from bar codes to RFID, far greater flexibility and the integration of a wide variety of coding schemes within a single system that cater to distinct needs and contexts. ISO

14223/2 codes, for instance, which are specific for use in animal tracking and include information on the species and the premises where they are held, can be integrated within the ISO 15459 provisions. The way to achieve this is by simply setting DI to 8N. This facility also allows improved interoperability with other competing or emerging numbering schemes which can be incorporated under particular DIs as well as offering greater flexibility for future extensions.

Due to the pervasiveness of ISO standards, EPC has recently modified some of their proposals to also cater to better interoperability between the two systems. In particular, Gen 2 tag specifications have been extended with the provision of a parity bit as a toggle to indicate the type of identifier stored in the EPC memory bank (see Bank 01 in Figure 3.6; the Numbering System Identifier is part of the Protocol Control section) so that other identifier codes can be stored in addition to EPCs.

5.4.1 Allocation of ISO 15459 Codes

The rules for the allocation of ISO codes are defined in Part 2 of the 15459 standard documents. Similar to the ISO registration practice for bar codes, ISO 15459/2 identifies the Netherlands Normalization Institute (NNI) as the root registration authority. NNI is the only organization that can authorize the assignment of issuing agency codes (IAC). In the case of the ISO 15459 code displayed in Table 5.2, the IAC is EDIFICE, the European B2B Forum for the Electronics Industry, an association of electronics suppliers, which has registered with NNI with a view toward providing its members with individual unique company identification prefixes.

There are a relatively large number of organizations with IAC registrations some of which are multi-national commercial companies (for example, IBM maintains its individual registration), and others are associations of regional or sectoral scope. In the case of EDIFICE, each member assigned with an individual company ID can subsequently define how to structure the object serial numbers that correspond to it. A common approach of course is to follow GTIN and separate the number into two parts, the first identifying the type of the object (often referred to as product class) and the second identifying the particular item within this class (often referred to as item serial number), although this is not mandatory.

5.5 Universal ID

A third system of general-purpose universal identification for RFID is the so-called Universal ID (uID) specification, maintained by the Ubiquitous Networking Laboratory (UNL) at the University of Tokyo. uID has a structure similar to that of EPC and ISO identifiers but offers a fixed code length of 128

Table 5.3. uID identifier example, the object ID stored on the user memory of the RFID tag would be 33123372312323CFE2456A789B245E57892 (in hexadecimal notation).

Version	TLD	Class	Domain	Identification Code
0x3	0x3123	0x3	0x723123	0x23CFE2456A789B245E57892

bits, which represents numeric rather than alphanumeric values. The code is subdivided into five sections namely (see Table 5.3)

- the version, which is roughly equivalent to the EPC header (fixed length of 4 bits),
- the top-level domain (TLD) code (fixed length of 16 bits),
- the class code (CC), which is similar to the EPC partition code in that it specifies the boundary between the remaining two fields of uID (fixed length of 4 bits),
- the domain code (DC), which is similar to the ISO data identifier in that it specifies the type of the particular identification code (variable length of 4-100 bits), and
- the identification code (IC), which is the serial number of the tagged object (variable length of 4–100 bits).

The DC is the mechanism by which other identification schemes can be incorporated into uID, and there are already provisions for the most common types, including JAN (the Japanese standard for bar codes), ISBN (for books), and last but not least EPCs. The IC is the serialized part of the complete object uID and corresponds to the concatenation of the CIN and the serial number in the ISO standard. Although a code allocation system is operated on an ad hoc basis by the UNL, it is expected that at a later stage of development of the uID system this will be conducted through a membership-based system similar to the EPC.

5.6 URI-Based Identifiers

In our discussion of EPC codes, we have described several representations of *pure* EPC identifiers, including binary and hexadecimal, but also as Universal Resource Identifiers (URI) following the web convention. This is certainly the form in which most software developers will encounter EPCs, as lower-level details are mostly encapsulated within middleware abstractions. Naturally, one could ask what the advantage of a unique identification code is when a URI could be directly encoded on a tag. The answer to that has to do mostly with the current state of the RFID technology and in particular with the limited storage capacity of UHF tags. Indeed, at their current capability,

UHF RFID can hold a few bytes of data, which would be insufficient for storing a URI of medium length.

Of course, HF tags do not have this limitation, and indeed they are capable of storing a complete URI pointing to a location on the web where additional information about the tag can be retrieved. This approach clearly has considerable advantages for general-purpose tagging and has been adopted for location and service identification under the Near Field Communication (NFC) scheme [87]. NFC was seen initially as a technology that connects two sophisticated mobile devices, but it has recently been extended to include interactions with RFID tags and readers, and currently supports application-specific provisions for ticketing and mobile electronic payments [20].

5.6.1 URLs in Near Field Communication

NFC-enabled mobile devices can interact directly with ISO 14443 RFID tags and retrieve data content (recall that this type of RFID tag currently provides up to 4 Kbytes of storage). NFC smart phones in particular are capable of acting as both reader and tag and at the same time provide internet connectivity over a cellular or other wireless network. As a result, it is natural to use a Uniform Resource Locators[2] (URLs) embedded in tags as identifiers for the associated resource or object.

Clearly this would also require additional application layer protocols to supplement the functionality of ISO 14443 to support the transfer of URL data across NFC devices. This is facilitated via the NFC URI RTD specification, which defines how URIs can be stored in a tag using the NFC Record Type Definition (RTD) format, assembled and exchanged between devices using the NFC Data Exchange Format (NDEF) technical specification. By adopting this approach to identification, tags become physical hyperlinks that can be further associated with actions. This physical hyperlinking feature is greatly enhanced by the increased resource availability on most NFC phones, for example their support for embedded Java virtual machines,[3] which allows the automatic launch and execution of applications relating to the interpretation of the collected URL. For example, in [3, 89, 95] this facility of NFC is used to transform smart phones into a tangible interface for pervasive computing applications that automatically identify and retrieve information relating to particular objects or locations.

[2] Names and locators, URNs and URLs respectively, are the two different ways to identify network resources supported by RFC 1630 and primarily intended for use on the web.

[3] A common feature of such systems is that they provide access to the functionality of NDEF via Java classes that are available to applications, in this case the DiscoveryManager and TargetListener classes included in the JSR-257 API specifications.

5.6.2 URLs in Mobile RFID

The selection of near field RFID technology as the basis for NFC is primarily due to the fact that the extended memory capacity of HF over UHF tags facilitates longer URIs. Nonetheless, despite its limitations, the spirit of the NFC approach is employed by the Mobile RFID protocol, which supports ISO 18000-6C and EPC Gen 2 tags [74]. Mobile RFID also proposes a numbering system for identifying services and content called mCode and its compact version called micro-mCode. mCodes come in various lengths, ranging from 48 to 128 bits and follow a hierarchical structure similar to the ones already discussed in this chapter and compatible with both EPC and ISO 15693 systems.

Since it is not possible to store complete URIs in a single UHF tag, in Mobile RFID mCodes are automatically prefixed with the HTTP protocol and the service location of a local Object Description Service (ODS). ODS was specifically developed for this use, and in response to a Mobile RFID query it provides a full URI that contains complete service locations or names that identify the particular tagged object or location.

5.7 Summary

There are several options in choosing a universal identifier code system. From a technical perspective—with the possible exception of the potential performance implications of A particular code length size—there is very little that differentiates them, as they follow roughly similar structures and organize their address space in approximately equivalent ways. From an administrative perspective, each scheme requires registration with an issuing authority and some care in structuring the internal code structure, though outside the cost of licensing there is also little difference between the three systems in this area.

Each of the systems has a particular orientation, with EPC focused on the supply chain, and ISO on a variety of industrial automatic identification applications including the supply chain, while uID appears to have a broader scope and aims to address the more extensive requirements of general-purpose pervasive computing. On the other hand, EPC clearly has attracted considerably greater interest due to its exclusive supply chain focus and the fact that it provides a complete set of specifications for middleware, resolution, discovery, and repository services (see Chapters 7 and 8). Moreover, several vendors have already integrated these specifications into their products, and as a result EPC standards are well supported in commercial information technology products and have become significantly more mature than any of the other systems. Yet these benefits are not necessarily valuable in all applications.

6

System Architectures for RFID

In previous chapters, we have discussed the basics of RFID, looking at each individual element in turn. In this chapter we turn our attention to how these primitives are combined with other computing and communications technologies to build complete systems. There are two main ingredients required to achieve this, a way to transform raw observations into higher level application events, and a supporting collection of infrastructure services that are required for the coordination of individual information systems across authority domains and the construction of universal infrastructures for the resolution of data. In this context, a special role is reserved for one system component that bridges the RFID domain and the IP-based network by translating and routing data and instructions from the one to the other.

6.1 A Motivating Example

One question about RFID relates to its influence on current network and information system architectures, if any. To be sure, ever-increasing volumes of data captured through a variety of applications are typical of modern computing, as is the trend for more distributed and higher-performance systems. Within this context, many see RFID as just another application domain to follow this pattern and that can be accommodated within existing technologies and architectures. Yet there are aspects of RFID that are quite unique and addressing their requirements calls for new network devices, software components, and information services.

These requirements are best identified within a concrete case study and without a doubt supply chain management stands out among all RFID applications due to its unique complexity, the high number of potential participant organizations involved in a fully operational system, and its open and federated architecture. In Chapter 2, we presented a high-level view of how this application domain can benefit from RFID, and in this chapter we look at

G. Roussos, *Networked RFID*, DOI: 10.1007/978-1-84800-153-4_6,

the details by tracing a product item over its lifetime from production until purchased by a consumer.

Let us consider a carton of orange juice produced by Innocent Fruit Ltd, a subscriber to one of the numbering schemes discussed in Chapter 5, with its dedicated manager code. At production time, the carton is tagged and assigned its individual serial identifier, which uniquely identifies it among all other cartons of orange juice produced by the company. Together with many other cartons of the same type, it is placed in cases, each of which is assigned its own container identifier. In turn, many cases of juice are further packaged together and loaded on a pallet for transport—of course, each pallet is also assigned its individual container identifier. As orders arrive, pallets are subsequently loaded on a cooler truck for delivery to a client. The complete shipment is scanned pallet by pallet as each one exits the dock doors of the production facility of Innocent Fruit Ltd, and the records of all items and containers are recorded and stored in the enterprise resource-planning system of the company. The data collected from a particular shipment are cross-checked against the corresponding order to confirm that what was requested is shipped.

In this case, the client to receive this shipment is a supermarket chain, and the delivery is to be made at one of its regional distribution centers. From there, products are forwarded to individual retail outlets on demand and on a daily basis. Upon arrival, each pallet is unloaded from the truck and enters the warehouse through its dock doors, where again it is scanned and checked against orders. Note that even for a retailer of modest size, at any time there would be many trucks delivering products from different manufacturers to the same warehouse. Each shipment has to be matched with its supplier, accordingly checked for accuracy, and the relevant warehouse management system updated.

A typical distribution center holds several thousand product types, and during an average day hundreds of thousands of product items will enter and leave its facilities. Depending on the type of product, a pallet can easily carry 1,000 cases or 20,000 product items, which have to be scanned, accurately identified, matched with product descriptions and orders, and recorded in the few seconds when the pallet passes through the dock doors. While shipping information is likely to be provided by the supplier directly, most likely through the use of Electronic Data Interchange (EDI), individual identifiers would need to be checked via appropriate online product information repositories and supporting code resolution services that map codes to product descriptions and their related information. Furthermore, although it is quite feasible to use information only about the kind of product that is delivered, to take full advantage of the use of RFID, warehouse management and resource-planning systems should all be able to deal with item-level information. Until recently, the majority of such systems were incapable of dealing with serialized data, which had to be truncated, and though there has been progress toward this, such systems do not yet make full use of item-level information.

If the products assembled into pallets arrive correctly as ordered and accepted by the distribution center, they are split into cases and stored in different sections of the warehouse as appropriate for their category. Later, individual cases or items are repackaged for delivery to particular retail outlets—typically a pallet with a particular shop as its destination will contain a large variety of products, not solely cartons of juice, as required for the replenishment of shop supplies. As previously, these pallets would also be scanned and the shipment recorded on exit from the warehouse dock doors.

Upon arrival at the store, pallets again pass through dock doors, where they are scanned, cross-checked, and then inventoried. A network of readers in the back room and the storefront ensure that product movement is tracked at various control points so as to improve shelf availability and promotion management, reduce theft, and so forth. Hopefully, after a matter of hours or days, our carton of juice will arrive at a retail outlet, be placed in a display refrigerator, and be purchased by a customer. Its RFID tag could be used to speed up the checkout process, and unless it deactivates at the point-of-sale, it will be carried away by the consumer together with the product.

In this chapter, we examine this scenario step-by-step and identify all the individual system components and facilities required for its implementation. In the process, we highlight how the different elements fit together to produce an architecture unique for RFID and how the different features are reflected in software and systems specifically developed for RFID.

6.2 RFID Processing Stages

The scenario discussed in the previous section has two distinct features: RFID tags are scanned at high densities at specific control points while passing through the supply chain. In most cases, these control points would be located at dock doors instrumented with RFID antennas. Due to the limitations of RFID technology, it is necessary that scanning of tags be carried out in relatively controlled environments and under conditions that ensure a high read rate. At the same time it is necessary to select locations such that all products have to pass through them to guarantee that the complete stock has been taken. Docking doors are particularly suitable for this role, and furthermore their specifications have been standardized, a fact that can be of considerable help in developing RFID equipment that can be used under different circumstances with small modifications.

The second characteristic feature of this scenario is that a considerable proportion of the functionality is shifted from the network core to the edge. By network edge we are referring to two different aspects of the system. First, the term implies the edge of the corporate network, which is traditionally seen as being developed around the data center, which is often referred to as the core. Data centers are usually isolated and well protected from external

threats, and tightly controlled. Instead, RFID processing is carried out in satellite sites, for example warehouse facilities.

Edge is also a reference to the fact that RFID operates close to the ends of the IP network, a location usually reserved for clients with secondary and supporting roles. This is not the case for RFID, where the processing of raw observations has to be carried out near the point of capture of tag data, with communication between reader and tag not employing IP protocols. This may be an obvious observation following the discussion of RFID technology in the previous chapters, but it has an important implication in that it is clearly not possible to address a particular tag directly from the network. On the other hand sightings of a particular tag can be reported to the network, and in fact this is the normal mode of operation of readers. This asymmetry in addressing tags, and as a consequence in locating specific entities and initiating communication, is a direct consequence of the asymmetric relationship between reader and tag and firmly places RFID outside the internet realm. As such, a core component of RFID systems must be to provide the means to bridge the two systems and provide facilities for their interoperation.

Taking a closer look at the tasks that have to be carried out at each of the control points of the RFID system as discussed in the context of the supply chain scenario, different stages of RFID processing can be identified. Moving consecutively from the lower level, where observations are acquired by a reader, towards application level processing, where high-level or business rules are affected, RFID data are processed along the following pipeline

Collection of observations. At the lowest level, RFID readers interrogate their vicinity for the presence of tags and collect raw data as a result of this. For example, after tags have been inventoried and cataloged, readers would request and retrieve identifier codes and potentially additional data stored in the chip memory—in some cases, this would require a further intermediate authentication step to allow access to this information. Depending on the application, the duration of the interrogation cycle can vary considerably.

For example, in e-passport applications, a read cycle could last up to a minute and will include the visual scanning and recognition of information printed on the document, the calculation of access keys, the interrogation of the passport content, and the retrieval of biometric information. On the other hand, in supply chain applications, several hundred tags would be read per second during their transit through the warehouse dock doors, but only their identifier code would be retrieved. In other cases, the read phase would be followed by a further write cycle, as is the case in ticketing applications, where information about the current trip would be added to the ticket.

Additional sensors and actuators may be activated at this stage to collect additional information. For example, temperature sensors embedded in the reader infrastructure could be used to record the environmental conditions in which a particular object has been observed. In other cases, as objects pass through specific locations and their codes are retrieved, LED displays could be triggered to indicate information related to the state of the object; for

example a red light could point to products that are beyond their expiration date.

Smoothing of observation data. Raw RFID observation data (that is, the stream of codes as retrieved directly from the tags) can be erroneous or incomplete or contain spurious identifiers. As a result of various read errors, these data may be misleading or unsuitable for use in specific applications and have to be postprocessed to become useful.

Smoothing observations is the process of cleaning raw collected data to remove different sources of errors, including:

- Incomplete codes that are unusable and must be discarded; this may happen for example when the tag moves beyond the interrogation zone of the reader before transmitting its complete identifier.
- Codes that belong to objects that are transient and thus irrelevant to the application and must also be removed; for example, identifiers that are affixed to garments that belong to passersby.
- Indeterminate reads must be resolved; for example, using authoritative records from local persistent storage, as is the case in ticketing systems where the tag moves outside the reader range before a successful update can be confirmed.
- Tags that have not been read and must be rescanned.

The last point is particularly important due to the fact that for various reasons tags are not read correctly every time a reader is within range, and as a result, to ensure that all tags have been read, the process is repeated several times. This in turn raises other questions regarding the sampling process and as a result smoothing can be a relatively complex process. We will discuss this in more detail in the following chapter.

Translation of observations into events. Following smoothing, observation data are still not useful to applications, which are interested in higher-level events. For example, in a supply chain application, it is not relevant to the business logic layer if a tag has been read by a particular reader but rather the fact that a specific pallet containing particular product items has entered the warehouse through a specific portal. This transformation of lower-level observations into higher-level application events is typically achieved via filtering and aggregation.

Filtering refers to the task of scanning the input stream for specific tags or tag groups. For example, a filtering rule may define that all item-level identifiers are discarded and only cases recorded or that only specific ranges of products are retained, specified according to the class section of their code. This allows an application to process only those identifiers that it considers relevant to its task, thus separating the stream of incoming data.

Aggregation on the other hand refers to the process of grouping together identifiers that share a common pattern and keeping only composite measures,

for example the total number of tags observed. For example, a shipment-checking application operating at the exit of a manufacturer may be required only to check that the correct quantities of products are shipped rather than the full details of the pallet contents. Aggregation would allow quick operation, as only the essential information is used by the application. That is not to say that the complete dataset is discarded; it may be retained by a different subsystem for further processing by a less time-critical process.

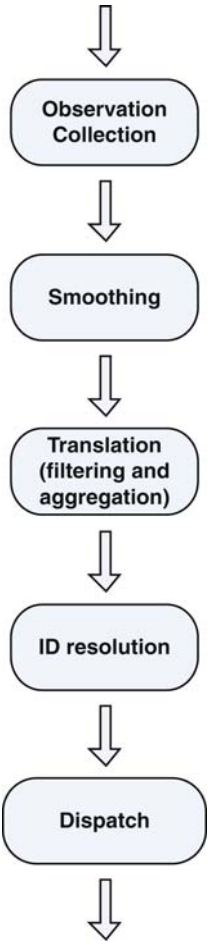

Fig. 6.1. The RFID processing pipeline.

Identifier code resolution and context retrieval. At the end of the previous stage, identifier codes have been retrieved, smoothed, and those that are of interest to the application separated for further processing. However, the

specific identifiers that have been recorded must be associated with descriptions of the entity they represent and associated data retrieved. This is a two-stage process and requires access to network services to

1. map identifier codes to network service locations that can be further queried about related data and historical details and
2. respond to specific queries related to the current condition, the properties, and the history of the object.

The first step can be seen as a general-purpose code-resolution process at the end of which the application has acquired information that specifies an access point and the manner in which it should be accessed to obtain information held about the code. This is a relatively straightforward task akin to internet domain name resolution, and indeed in some cases the DNS can be used for this task.

The second step of the process is far more complex and can take different forms, depending on the results of the query of step one. The details of the protocol, for example, that should be used to access related data as well as the type and availability of the data itself will define the complexity of this task. Protocols and services that can be used to this end are discussed in more detail in Chapter 8.

Dispatch and process event data. With both the list of entities and their full information at hand, at this point it is possible to conduct application-level processing, where decisions are made on the basis of higher-level strategies, for example business rules in the case of supply chain applications. For instance, when the system identifies that a specific pallet has entered a particular retail outlet, it would trigger updates of inventory records to include the items contained in this shipment and notify operators to replenish out-of-stock products that have now been delivered.

Note that this process works bi-directionally; that is, applications control data flow by defining events of interest and by declaring their interest to the RFID infrastructure. In fact, information flows from applications towards the observation level with instructions of what to look for and when, and tag-related data flow from the raw RFID level toward applications signaling specific events.

6.3 The RFID Stack

Having identified the different stages of processing RFID observations, we now take the point of view of how best to abstract system components that can carry out these tasks in an unambiguous and effective manner. A suitable representation of system architecture that achieves this goal can be encapsulated in the so-called RFID stack displayed in Figure 6.2. In addition to providing a blueprint for RFID systems, the stack also helps identify the location of each component within the network [19]:

- Observations are collected at the reader level outside the IP network.
- Observation processing and event translation occurs at the network edge.
- Application-domain-specific logic operates at the network core (or data center) level.

The operation of the stack is assisted by two cross-cutting service planes represented as vertical bars in Figure 6.2 [21]:

(a) network services used to resolve identifier codes captured in observations to entity descriptions and further information services that can be queried for additional meta and context data, and
(b) infrastructure management (that is, maintaining configuration and status information related to the operating condition of RFID readers and other sensor elements).

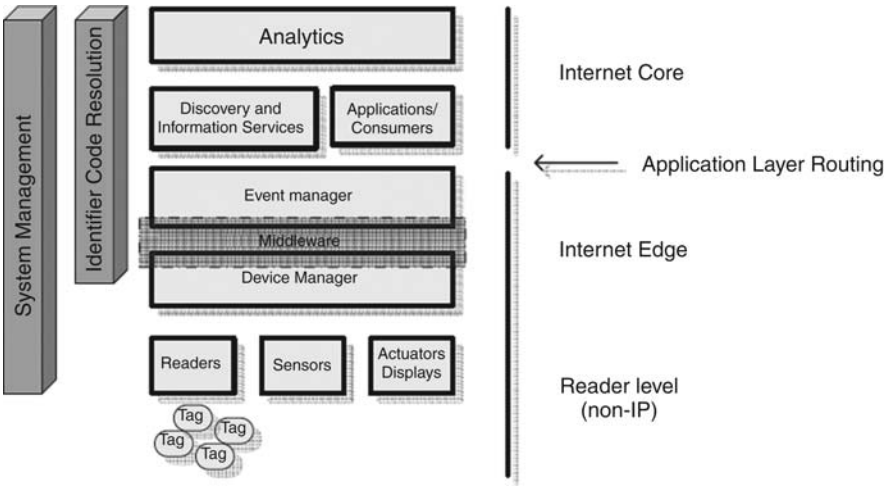

Fig. 6.2. The RFID stack.

Reader layer

At the bottom of the RFID stack is the non-IP segment of the system, where tags, sensor, and displays communicate with readers to receive instructions or return observations. At this layer, communication is through simple often device- and manufacturer-specific protocols that are not affected by the complexities of higher layers.

Readers are responsible for the translation of such lower-layer details into useful data representations that can be consumed by higher layers of the protocol stack. Often this process involves augmentation of the raw data received by

the end device (for instance a tag), with additional data known to the reader (for example, the readers own identity, its location, and the time when the specific observation was collected). Readers can also record and process other details related to the specific communication session, for example whether an access control mechanism was necessary to gain access to the tag memory or if writing in addition to reading was carried out.

Readers control communications at this layer and provide the bridge in translating incoming instructions from higher layers that use relatively complex interfaces and IP-based transport and deconstruct them into step-by-step processes using the specific physical layer protocol as appropriate. A consequence of this fact is that individual tags, for example, are not addressable from the internet directly. Instead hosts that are interested in communicating with a specific tag for whatever reason have to use a reader as an intermediary.

Edge layer

The edge layer relates to observation processing near the point of observation capture. The need for the introduction of this layer stems form the fact that at the point of capture, observations are numerous and redundant to a large extent. As a result, it is neither efficient nor useful to transport these data further up in the RFID stack, as it would cause congestion, increase the operational costs of the system, and overload computing resources that are more effectively occupied otherwise.

For this reason, pre-processing is located (physically) closer to the point of capture, with a view to providing smoothing, filtering and aggregation facilities. This layer receives the raw stream of data captured by the reader layer and transforms it into a stream of events that make sense in a higher context. For example, the incoming stream of observations may indicate the times, codes, and reader antennas through which several cartons of Innocent smoothie have been observed, and the outgoing stream would report that three cases of this product have entered the warehouse. The latter information is useful for stock-keeping, and the former is superfluous and unusable in that form.

The functionality required for the implementation of the edge layer may be shared between several hardware components and software services. For example, advanced readers that have recently become available would be able to do some pre-processing before passing on the data for more complex processing. At the very least, such readers would be able to cache observations and deliver them in batches in a synchronous or asynchronous manner. Other devices in the local network may apply further transformations to the observation stream and be capable of forwarding the data to a recipient at a remote location. For performance reasons, it would also be desirable to support different communication patterns so as to meet the requirements of specific systems and applications.

The coordination of the activities carried out at the edge is the responsibility of the event manager (EM), a conceptual structure that acts as a data and context transformation filter. In the following section, we will consider the operation of the EM in details. At this point, note that the EM represents the principal innovation required for the operation of an RFID system at the local level.

Network layer

The network layer has a twofold role:

- to connect the network edge with the system core and
- to provide access to the resolution and repository services.

Access to the network layer is provided by the EM and takes the form of event streams communicated over the IP network following one of several alternative modes. In any case, the EM also has the role of updating the current state of all connected systems through specific interfaces. In most cases, these would be using the specifications of one of several information services discussed in Chapter 8. The EM may also provide local persistence of captured streams for redundancy or so that a historical record can be maintained for future use.

More than representing specific devices and functionalities, the network layer in an RFID system is the glue that brings together the edge and the system core. Both endpoints of this communication are (to different degrees) secured and controlled systems, but the segment that connects them is an unreliable public network and traffic can flow in and out if unchecked. As a result, the network layer must implement measures for the protection of both communication peers while at the same time preserving connectivity and adapting to meet changing conditions in the public network.

For example, one critical role of this layer is to ensure that identifier resolution can be carried out under any circumstances and that disruption is kept to a minimum. This is particularly critical because the identifiers stored in item- or container-level tags are not self-descriptive and it is necessary to use a code resolution service on the network to retrieve a description of the type of entity they represent and other required information.

Network core

In a large-scale RFID system, the role of coordinating all network activities carried out in satellite locations would be provided by the network core. By core, we imply a network location that provides redundant system-wide services and has total awareness of the state of the complete system. Note that the network core is a logical entity and does not not necessarily exist at a single physical location. In fact, to increase redundancy and performance, it is most likely that such universally accessible service points would be replicated

or shared between several data centers or other trusted service aggregation points.

The exact details and structure of the core depend primarily on the detail of the particular application as well as the size of the system. In any case, the level of abstraction at this layer is the highest, and in terms of traditional multi-tier system parlance would correspond to business logic. For example, the core can be implemented as a data warehouse or as an enterprise resource-planning system, and process the data coming into the system from all the satellite sites. This component can also have a proactive role in directing information by specifying desired behaviors and other types of captured data, which will be translated by the lower level to specific event management and observation planning tasks.

Finally, another role of the core element of the system is to carry out data analytics to identify and record longer-term trends and identify patterns from overall system use. For example, ticketing systems would employ a data warehousing approach to estimate station and train usage trends and also to identify ticket holders that try to evade the fare charging system.

Resolution and repository services

Network services represent the first of the two vertical planes working in parallel with the RFID stack. We will discuss the role, structure and operation of these services in more detail in Chapter 8, but for the time being it is enough to note that they play the following roles:

- When a tag identifier is presented to them, they return a service description endpoint where additional information about the tagged entity can be retrieved. The service description point also includes information about the different ways that it is possible to access and query the service.
- When a service point that corresponds to a particular code is queried with the specific identifier, it returns information held about the tagged entity. Such information may include the type and other details of the item, observation and events it has been involved with, and other historical information captured about it.

As a result, common tasks that must be carried out at the end and network layers as well as the core of the system all require access to both types of information to operate. For example, in the standard warehousing receiving scenario, the EM needs to match codes observed at the docking portal to product descriptions so that it can confirm the complete receipt of a shipment and match it to its corresponding order.

System management

The final ingredient of the RFID stack is the second vertical plane, which relates to the management of software and system elements. Of course, network device management can be carried out using some version of SNMP and

related management applications, and this functionality would typically be employed in an RFID system. However, it is also necessary to have specific controls related to the operation of particular RFID components.

For example, to manage the operation of the reader infrastructure it is necessary to provide specific monitors of the operation of individual reader antennas so that it is possible to diagnose problems with their operation. This is a common issue with RFID since its transparent operation implies that it is often unclear just by observing collected data to recognize an actual failure against a situation where readings are not collected because there are no tags in the vicinity of the reader.

Other common tasks would involve the management of the level of support for particular types of tags and communication protocols provided by the infrastructure as well as their update. Such tasks can be represented in an abstract model with bindings for general-purpose protocols, including SNMP. Note that due to the general lack of commonly accepted standards in this area, and their relative immaturity where they exist, each vendor of network RFID equipment appears to have developed its own solutions, and for the short term at least the management task appears to represent a major bottleneck for the deployment of systems.

6.4 The Event Manager

As noted earlier, one of the most significant effects of RFID in term of systems architecture is that it pushes a substantial proportion of system functionality to the network edge. To ensure appropriate levels of service, for example, efficient processing of observations and persistence, computing at the edge requires support through the provision of a specific set of features and capabilities, which we refer to as the RFID *event manager*. Event managers can be implemented in a variety of ways in software or hardware or a combination thereof, but in all cases they fulfill the same tasks to

- bridge the IP and RFID networks by translating RFID observations into higher-level events via filtering and aggregation,
- manage the RFID reader infrastructure and related sensor and actuator devices, and
- offer a single service point and interface to applications.

From a system architecture perspective, event managers are similar in scope and function to peerage structures common in overlay networks including peer-to-peer [88] and pervasive computing overlays [105], as well as dynamic proxy servers in application-level active networks [13, 40].

Seen this way, the RFID architecture consists of a network of event management peers that are transparently accessible through a single interface and coordinate their operation via super-peers at the network core. This "architecture-less" infrastructure is multiplexed with traditional directory and

state repository services to deliver complete system functionality. Another way to look at the event manager is that it provides an RFID-specific implementation of a system element common to pervasive computing applications (see Chapter 10 for more on pervasive computing), namely it supports the translation of lower-level data to higher-level context information [66].

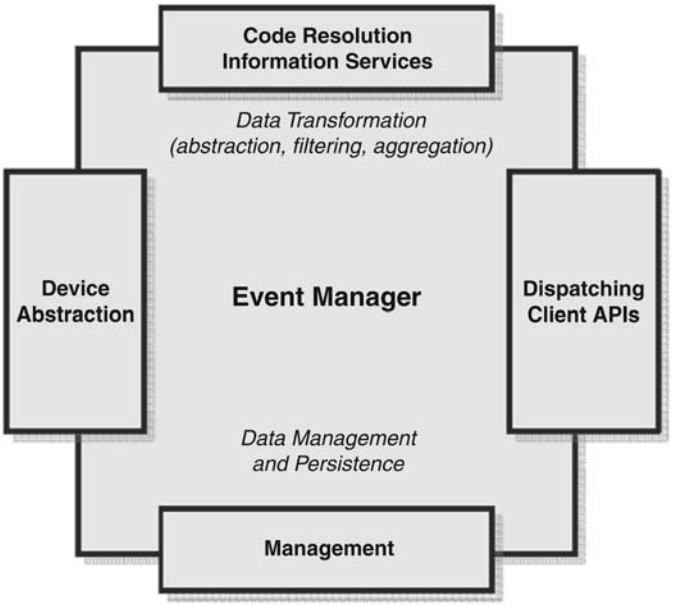

Fig. 6.3. The structure of the RFID event manager.

The structure of the event manager (EM) is depicted in more detail in Figure 6.3. Its role is twofold:

1. To carry out data transformation. Incoming observations collected through interactions with readers and other devices must be filtered, smoothed, and aggregated and otherwise processed to extract application-level events. The EM transforms incoming observation streams by applying statistical techniques and relational rules that provide local context into events that make sense within the higher-level context of specific applications.
2. To provide observation and/or event persistence. Since the EM operates at the network edge, it is almost certain that in most cases it will not be efficient to transmit the complete record of collected observations to the network core. Moreover, events forwarded to applications at the higher stack layers may be lost due to network or application unavailability. As a result, the EM may opt to store transaction records locally to provide persistence in case of such failures.

To fulfill this role, the EM supports four communication channels. The first channel is between the EM and RFID readers and other devices that provide a source of observations as well as control of actuators and displays. Communication over this channel would typically be over a low-level interface, which still provides a level of abstraction in that it hides individual differences of particular platforms and technologies. A common example of this would be the EPC Low Level Reader Protocol (LLRP), which provides a collection of primitives that offer abstracted reader and other sensor actions and can be used to control and manage such devices. The EM would also provide several data ports supporting a variety of protocols for incoming observation streams.

The EM would also provide support for communication with applications that consume event data, which may include target-specific applications such as warehouse management systems or entity-related information services, for example the EPC IS repository service, which is detailed in Chapter 8. Communication would be bi-directional: applications would employ one of the APIs supported by the EM implementation to specify the types of events they are interested in and the regularity with which they wish to be notified that new data have become available. One of the challenges related to the data transformation task carried out by the EM is the preparation of a processing plan that will produce all the different types of events requested by applications in an efficient and effective manner.

Another way to look at the EM is as a filter that converts the incoming stream of raw data into higher-level abstractions. This context translation process can happen at different levels, leading to progressively higher-level abstractions until it meets the application model. However, due to its complexity and its dependence on the context provided by the specific application domain, this task has not been developed, even within relatively mature areas (for example, supply chain management for grocery products).

6.5 Platforms

In this chapter, we have discussed the different elements of an RFID system and how they fit together to form the RFID stack. A core component of this architecture is the event manager, which plays the central role in orchestrating processing and dispatching of observation streams. In this section, we briefly discuss how different platforms implement the EM and present a number of systems that highlight the different approaches.

6.5.1 Oracle Sensor Edge Server

Oracle implements the EM as part of its 10g application server under the Sensor Edge Server (SES) product line. The SES design corresponds well with the EM elements outlined in Figure 6.3, with support for the EPCglobal standards and common J2EE messaging facilities. From a practical point of

view, operating the SES at a satellite site would involve the in situ installation of at least one hardware device with full network and internet connectivity to run the SES software.

There are three features of SES that are worth noting:

- the use of Oracle Streams for communication to the core layer,
- the provision of a high-performance local repository at the EM for maintaining the complete historical record of collected observations, which also supports extensions to SQL triggers for notifications occurring in the absence rather than in the presence of specific events, and
- the implementation of the Sensor Edge Mobile (SEM) component, which provides off-line data collection and batch processing at the EM.

Streaming data is one of the strong points of the Oracle platform, and by using their standard stream engine, SES provides a bi-directional communication channel and a secure, scalable, and reliable transportation method. Of course, advanced database technology is also a strong point for Oracle and can be used to good effect in supporting the persistence of observation histories.

SEM is in essence a mobile database client integrated with a reader that can be used to scan tags and collect information beyond the reach of the network. Its built-in synchronization facilities imply that when SEM reconnects to the network, the collected data are uploaded and processed automatically by SES.

Finally, one feature that presents some interest is the use of negation in triggers. Event-condition-action rules have been designed to produce results in response to declared events and conditions, but in the case of RFID it is equally important to take action when an expected event does not occur. Such facilities have been added to both database and stream processing, and it is now possible to trigger events in the absence of data.

6.5.2 IBM Premises Server

IBM's implementation of the event manager is in the form of the so-called Premises Server (PS), based on the WebSphere platform. Similar to SES, the PS is a J2EE application server with extensions to provide specific RFID-related facilities and features. The PS also requires that the IBM Device Infrastructure (DI) be available on all the devices that act as sources of observations. IBM DI is an OSGi-compliant component that handles the delivery of data to the EM for further processing. Also similar to SES, PS would also require a local installation on the premises of the satellite site to host the software.

IBM PS is notable for its role and its support in the development of the ALE middleware discussed in detail in the following chapter, being one of the principal contributors to this specification. A second interesting fact about this system is its tight integration with the full business infrastructure supported by IBM. This is not a standalone component, but thought and effort have been

invested in its role within the software engineering lifecycle, which greatly facilitates the development of applications.

6.5.3 Cisco Application Oriented Networking

Cisco supports the implementation of EMs and RFID processing in general by taking a rather different route from the application server-based approach adopted by Oracle and IBM. Cisco EMs are network devices that also carry out all the usual data-routing tasks and are not necessarily dedicated to RFID processing. They carry out observation-processing tasks by examining data streams or other messages flowing through the network and applying on-the-fly data transformation rules to produce higher-level events.

The main facility of networking equipment exploited in this approach, dubbed Application Oriented Networking (AON) by Cisco, is the ability of network processors to extract the content of packets they inspect. This is a major diversion from the role of network equipment to operate at the lower layers of the protocol stack, and it has emerged through the work of the networking community on active networks outlined earlier in this chapter. Note that Cisco provides a complete suite of software development environments specifically designed to support the design and deployment of such data filters.

Compared with application servers, there are two potential advantages of this approach are two:

- *Traffic shaping.* Due to the fact that with the growing popularity of RFID in the SCM, and especially with the potential introduction of item-level tagging, observation and event streams are expected to represent a considerable proportion of all network traffic, AON can provide much greater control in managing and prioritizing the flow of this information. Unlike server-based approaches, in-network processing of data can support far greater flexibility and forecasting of needs as well as a staggered strategy for developing additional network capacity.
- *Systems management.* AON devices are managed within the integrated management facilities provided by Cisco for all network components, carried out from a single centralized location. Moreover, the relatively advanced remote management facilities provided can lower the costs as they can reduce the need for engineering staff to be present at the satellite sites. Finally, this approach also offers distinct advantages in network capacity planning and the management of security features and policies via a single control point.

In practical terms, instrumenting a site for RFID requires that a variety of readers and antennas be installed within an environment that has not been designed for this, and can be hostile to wiring and digital communications in general. As a result, readers will have to be networked using a combination of wired and wireless approaches to fit the needs of the particular installation, and in this context AON would have an advantage in that it would integrate

seamlessly with the rest of the infrastructure as simply another network component that can be managed through the same processes and utilities.

Finally, AON can be easily integrated with other platforms and solutions to offer a partial data transformation facility rather than provide the complete solution. In this way, a particular system can choose to carry out those parts of the observation stream processing in-network that would most benefit from this approach and combine this with additional elements that are best carried out elsewhere; for example, historical data can be maintained within a separate data repository.

6.5.4 Reva Tag Acquisition Processor

Reva Systems is a company that was specifically created to provide solutions for network RFID. Contrary to all previous solutions discussed above, which offer commodity servers of network routing equipment modified for use with RFID, Reva has developed and markets a network appliance that can play the role of EM. In Reva parlance, this device is called a Tag Acquisition Processor (TAP) and offers all the facilities we have discussed as part of the EM specification.

To be sure, using TAP as the main network infrastructure component in the way proposed by Reva can offer several advantages. First, since this system is built for a dedicated purpose it is easier to carry out maintenance and provide simple management functions from a remote location. Moreover, its deployment leads to infrastructures that are fairly homogeneous and thus easier to manage and operate.

However, the greatest benefit offered by TAP is its close adherence to standards, which ensures interpretability and compatibility with a variety of third party reader equipment. In fact, Reva engineers have played a leading role in the development of reader management standards within EPCglobal and have made significant contributions in producing efficient and effective mechanisms for the management of RFID infrastructure in general.

6.5.5 Accada Open Source Platform

`Accada.org` is a reference implementation not only of the EM-related features but of the complete EPCglobal specification. The system is implemented as a collection of J2EE applications on top of the also open source Tomcat J2EE server and MySQL database. Although this is meant as an unoptimized reference implementation, its structure is fairly similar to other J2EE-based systems, and I include it here as it is the best way to learn about these systems. Access to the source code is a unique facility for anyone interested in these technologies, and I recommend it as the best starting point for the study of EPCglobal.

6.6 Summary

The implementation of RFID pushes a substantial proportion of system functionality from the network core to the edge. As a result, complete RFID systems have a distinct structure, which we have formalized in this chapter as the RFID stack. Organizing the operation of a network RFID system in these terms offers specific advantages in providing a single reference model for a variety of platforms and technologies but also because it can help identify the main elements required in the design of systems for specific application areas. A core component of such solutions is the development of an efficient event manager and we have briefly outlined how different platforms can offer specific functionalities to support this task.

7

RFID Middleware

In Chapter 6, we identified the central role of the RFID event manager in processing observations at the network edge. In this chapter, we consider the two main mechanisms that provide programmatic control of the event manager by specifying the particulars of the translation process of raw tag observations to application level events. First, we examine primitives that are useful in abstracting the main features of the process into logical constructs, and then we review filtering, aggregation and inference techniques used in the generation of events. Typically, such techniques are encapsulated in middleware, and we discuss specific systems and implementations.

7.1 The Role of RFID Middleware

Looking back at the RFID stack of Figure 6.2, note that the event manager interfaces with data capture devices at the lower layer, notably RFID readers, and with enterprise class applications and network services at the higher layer. A common scenario in terms of information flow would have observations of RFID tags captured at the device layer streamed into the event manager and transformed and subsequently streamed to consuming applications (see Figure 7.1). Observations coming from readers in particular can be described in terms of tuples of the form

$$(reader\_ID, tag\_code, time),$$

where $reader\_ID$ is the identifier of the reader that captured the observation of tag identifier tag_code at time $time$. During standard operations, readers will transmit streams of such tuples toward the event manager in response to observations. The event manager should be able to control the manner and frequency with which these observations are carried out and possibly also request that readers conduct some simple pre-processing; for example, to clean the input stream of incomplete readings or select only tag codes that correspond to cases but exclude items and pallets.

G. Roussos, *Networked RFID*, DOI: 10.1007/978-1-84800-153-4_7,

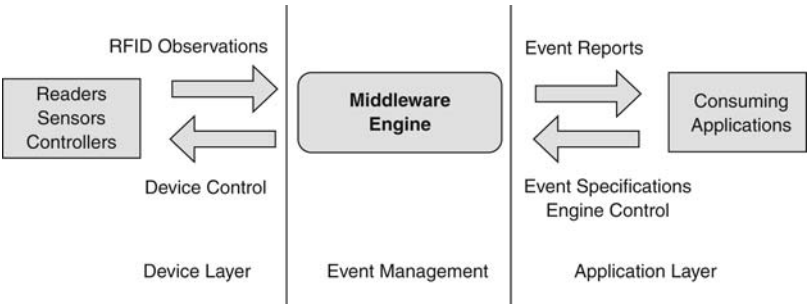

Fig. 7.1. RFID Middleware and its role within the RFID stack.

The event manager must also deliver notifications to consuming applications that are at the appropriate level of abstraction to be directly usable by them. For example, such an application-level event would correspond to the fact that a certain pallet containing a specific inventory of cases and items and related to the shipment of a particular order has left a warehouse and is on its way for delivery. This event would be important in the context of a warehouse management system, which would need to update inventory levels for instance. We discuss some common types of such application level events in Chapter 8. In the meantime, Figure 7.2 shows an example of an application-level event that combines observations with the physical and business location where they were recorded and the business process stage involved.

Although until recently there have been different views on how best to deal with such event management functionality, it is increasingly clear that it is best implemented as middleware. This choice provides for a clear system boundary with well specified interfaces and at the same time allows different system architectures that meet specific requirements for throughput and other performance constraints. In summary, middleware address the following needs of RFID processing:

1. Enhance application portability and interoperability. RFID middleware achieve this by decoupling applications from the physical layers of the infrastructure through an abstract API.
2. Reduce the volume of raw RFID data. RFID readers and other data sources generate high volumes of streaming data that must be cleaned and summarized. RFID middleware reduces the size of streams by accumulating data through counting and grouping operations over specified time intervals and cleans data streams by filtering them to eliminate duplicate identifiers and codes that are not of interest.
3. Infer higher-level events from raw observations. This task can have several components including combining observation tuples with other relations (for example via join operations) to provide context and employing situation-specific knowledge (for example, the details of the actual

```
<ObjectEvent>
    <EPCListType>
        <epc>
            urn:epc:id:sgtin-96:3.0037000.06542.773346595
        </epc>
        <epc>
            urn:epc:id:sgtin-96:3.0037000.06542.773346596
        </epc>
        <epc>
            urn:epc:id:sgtin-96:3.0037000.06542.773346597
        </epc>
    </EPCListType>
    <ActionType>
        ADD
    </ActionType>
    <BusinessStepIDType>
        urn:oid:shipping
    </BusinessStepIDType>
    <ReadPointIDType>
        urn:ean-ucc:gln:0073796001506
    </ReadPointIDType>
    <DispositionIDType>
        urn:oid:forsale
    </DispositionIDType>
    <BusinessLocationIDType>
        urn:epcglobal:fmcg:mda:gln:1234567890128
    </BusinessLocationIDType>
</ObjectEvent>
```

Fig. 7.2. Example of an application-level event that indicates that a number of items identified by their codes are cleared for sale and have been recorded as a specific location while being shipped to their final destination.

configuration layout of a particular reader installation) to decide on the likelihood of a particular observation.

4. Report application-level events to consuming applications. Reporting can be synchronous or asynchronous, periodic or single-shot, and can be conducted over a variety of transport mechanisms.

Using middleware to infer events from raw observations hides considerable complexity, and naturally the quality of its implementation directly affects the accuracy of event management. There are two aspects in this task that affect the results in different ways:

- *Statistical inference.* Due to the relatively low read accuracy of RFID, a reader will scan any tag or collection of tags many times and average out the observations. Effectively, a statistical filter is applied to tuple streams to decide the reliability of a group of observations and whether an event

is generated or not. The details of the filter selected have considerable bearing on the production of accurate results, and some examples will be discussed later in this chapter.

- *Reasoning.* Observation tuples must also be joined with related situation-specific data to produce meaningful higher-level events. For example, join operations may associate a specific tuple with the actual location of a reader and the class of the tagged item to produce higher-level context. In this setting, further observations may or may not strengthen conclusions regarding the actual state of particular items.

Later in this chapter, we will examine in more detail the various issues related to inferring events from observations, as the brief outline above raises more issues than it answers. Indeed, implementing robust and highly accurate RFID middleware can be a challenging process.

7.2 Docking Portal: A Motivating Example

To gain a better understanding of how RFID middleware would be used in practice, we examine a specific use case in detail. Recall that we have already identified docking doors as a highly suitable location to observe the movement of tagged items, as they provide a good control point through which all items have to pass. Revisiting this scenario from an identifier perspective, at the manufacturing facility products are fixed with individual tags encoding their EPC code, including their GTIN, which contains their item-specific serial number. Individual product items are then packaged in cases that are also tagged individually using EPC and assigned their particular SSCC (or in some cases a GTIN representing a case of product items). At this stage, each case SSCC is associated with the GTINs and all the items it contains, and this information is published on the local EPC IS. Cases are then loaded on pallets and often enclosed within some protective material, usually either cardboard wrap or transparent stretch film, and again tagged with their corresponding SSCC (see Figure 7.3). A particular pallet may contain cases from different product lines which are mixed due to the specific quantities included in the order placed by the retailer. The SSCC of each pallet is also associated with the SSCCs of the cases it contains, and this information is also published on the local EPC IS.

When all the necessary pallets are prepared for shipping, they are placed in a container and loaded on a truck for delivery to one or several DCs. This point offers the first opportunity to establish a control point for the movement of products downstream in the supply chain: readers located at the exit gates of the loading bay of the manufacturing facility scan the shipment as it is being loaded on the trucks and record every product item, case, and pallet identifier, grouping them together and associating them with the corresponding retailer's order details and DC destination. This information can be transmitted to the retailer so they can anticipate the arrival of the shipment.

Fig. 7.3. Typical RFID enabled warehouse loading bay portal.

On their arrival at the DC, pallets are individually unloaded and moved into the warehouse via a portal that records the arrival of the scanned EPC codes and cross references the recorded numbers against those expected. If the products are confirmed to be the ones expected for delivery, the warehouse management systems are automatically updated and the pallets are forwarded for storage in the facility. The same process is followed in loading the product cases for delivery to retail shops. At the shop, cases are again received, automatically checked against the expected deliveries, and, if confirmed, the store warehouse management system (WNS) is automatically updated.

In this process, a central role is reserved for the loading bay doors into the warehouse, as in most cases they represent the best location to place a control point for checking and updating the flow of products. As a result, dock doors are often turned into RFID-enabled portals where pallets are scanned and where the actual items delivered can be cross-referenced and inventories updated. This location works equally well as a control point for manufacturing plants as for distribution centers and retail stores and for both incoming and outgoing shipments.

Looking closer at the sequence of events involved in the operation of one such portal, the retail store receiving dock, the process starts with the receipt of an Advanced Shipment Notice (ASN). This is a common EDI message that is prepared and transmitted by the DC at the time when the pallets for a particular shipment have been loaded on a truck and leave the DC warehouse. The ASN is a notification of pending delivery and is sent to all parties responsible for the movement of freight from DC to store and the contents and configuration of a shipment. In this case, the ASN would contain at least the SSCCs of every pallet and possibly also those of the cases and the GTIN of the products included (the latter as a means of providing redundancy to EPC IS). The ASN would also record the total number of pallets and

the address and related details of the DC and retailer and can also contain numerous other related details.

At the store's receiving dock, the external motion sensor is tripped by the movement of the first pallet passing through the portal (see item 3 in Figure 7.4). Tripping the sensor results in the publication of a sensor-event message to the ALE engine operating on the warehouse event manager. This event marks the beginning of an event cycle that instructs the readers attached to the portal (item 2 in Figure 7.4) to begin collecting observations. The readers keep scanning and discover all tags marking individual products, cases and pallets. Each tag is typically discovered and read several hundred times and the observations are passed to the ALE engine either residing on the reader itself or at the event manager (depending on the model and the capability of the reader). Observations are processed according to the event cycle specification and reported to the WMS. An event cycle may be time constrained or terminated in response of a motion-sensing event tripped by the second, internal sensor (item 1 in Figure 7.4).

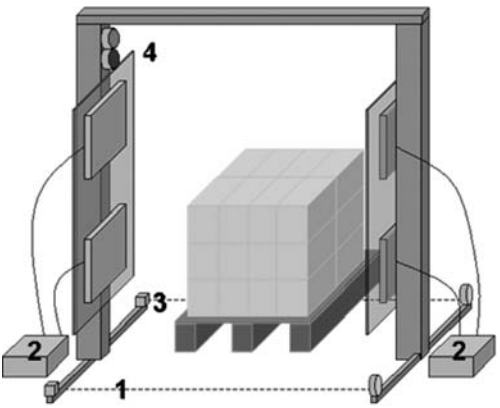

Fig. 7.4. Schematic of the components of an RFID-enabled warehouse portal: items 1 and 3 are motion sensors that activate and deactivate the portal; item 2 notes the locations of RFID readers, each of which has two external antennas attached; and item 4 represents red and green indicator lights that signal shipment approval or rejection.

Upon receipt of the event cycle report by the WMS, the list of products recorded is compared against the expected deliveries as specified in active ASN messages within the system. If the details match, then the pallet is expected and the portal switches on the green light on its frame (position 4 in Figure 7.4) indicating that the delivery has been accepted. At the same time, the inventory is updated with the new item received and cross-checked against the relevant purchase order. In case the codes retrieved by the pallet

are unexpected, the red light is switched on instead and the pallet returned to the truck.

During the aggregation cycle, the event manager filters duplicates, removes transients and codes that are not requested by the event cycle specification and returns the gathered EPC codes in a report. For example, if the event cycle specification requires that only pallet codes be collected, then all other types of tags (for example, item GTINs and case SSCCs) are observed but ignored.

In the following section we will describe the facilities supported by RFID middleware for the implementation of this scenario, with particular reference to the Application Level Events (ALE) specification developed within the EPCglobal systems.

7.3 ALE Middleware Abstractions

RFID middleware communicates with lower layers of the RFID stack to retrieve observations and higher levels of the RFID stack to deliver events (see Figure 7.1). Communication on both sides is organized in terms of cycles; that is, periods of accumulated activity that represent the smallest unit of interaction between middleware and readers or applications. Raw observations are collected in regular *read cycles*, and events are delivered in *event cycles*.

A read cycle represents the shortest unit of interaction between the middleware and a physical or *logical reader*. Physical readers correspond to single devices that have a unique network address and can capture information by scanning RFID tags but can also be any other source of RFID data, for example an EPC-enabled bar code reader or a terminal where EPC data are input manually. Note that there are no restrictions on the timing of read cycles by physical readers, which are free to decide on appropriate strategies. The reason for this is that for read cycles to be enforced, this would imply specific assumptions regarding the performance characteristics of reader devices which is undesirable in the context of middleware.

Logical readers, on the other hand, are abstract sources of EPC data that are often synonymous with specific locations. That is, several colocated RFID readers can be grouped together in a single logical reader so that they can be referenced as a single unit. For example, several docking portals located at the entrance of a particular warehouse, each supported by its dedicated reader device, can be grouped together into a single logical reader (for instance "WHentry") and handled as a single entity in that all observations captured through any of the associated readers will be processed in a similar manner.

An event cycle is the smallest unit of interaction between the middleware and a client application and can consist of one or several read cycles. Event cycles are defined implicitly through the determination of their boundaries, which can be declared in the following ways:

- as a specified time interval (for example, the next five seconds),

```
<logicalReaders>
   <logicalReader>WHentry</logicalReader>
</logicalReaders> <boundarySpec>
   <duration unit="MS">5000</duration>
</boundarySpec>

<reportSpecs>
   <reportSpec reportName="ThroughDockDoor"
               reportIfEmpty="true" reportOnlyOnChange="false">
       <reportSet set="CURRENT"/>
       <filterSpec>
           <includePatterns>
               <includePattern>urn:epc:pat:sgtin-96:20.*.*
                    </includePattern>
            </includePatterns>
       </filterSpec>
       <output includeEPC="true" includeTag="true"
               includeRawDecimal="false" includeRawHex="true"
               includeCount="true"/>
   </reportSpec>
</reportSpecs>
```

Fig. 7.5. Example of an event cycle specification. The specification demands that RFID data be captured via the WHentry logical reader, be repeated every five seconds, and report details of EPC codes that encode SGTINs for manager number 20.

- as recurring time intervals for example, every five seconds,
- in response to an external events for example, when a motion sensor at a dock portal is activated,
- as a period of time when no new observations have been recorded for example, if five seconds pass without any codes captured.

Each of these boundary types can be either the starting point of an event cycle or its termination point. In either case, the boundary condition is related to a trigger that fires when the conditions are met, thus signaling the beginning or the termination of the cycle. An event cycle specification with a specified time interval boundary is displayed in Figure 7.5.

For the dock portal scenario we considered in the previous section, start and termination boundaries are defined in relation to the activation of the external and internal motion sensors. The activation of these devices can be described in terms of triggers encapsulated in URI descriptions; for example, assuming that dock portal number 4 is a member of the WHentry logical reader, the definition of the boundary specification for the associated event cycle would be defined as:

```
<boundarySpec>
    <startTrigger>
```

```
            http://ale.com/startTrigger?triggerName=DockPortal4
        </startTrigger>
        <stopTrigger>
            http://ale.com/stopTrigger?triggerName=DockPortal4
        </stopTrigger>
    </boundarySpec>
```

Event cycle specifications can be assigned a unique handle or name and registered with a middleware implementation. For example, different client applications can register multiple event cycle definitions with an event manager according to their data needs. However, simple registration of an event cycle does not imply that the specification will be processed; in fact, the contrary is true. For the activation of a particular event cycle specification, it is necessary for at least one client application to have subscribed to it.

Applications declare their interest in specific event cycles and request one of three possible execution modes:

- Immediate mode presents the middleware implementation with an event cycle specification and requests that it be executed with immediate effect and the results returned synchronously to the caller. In this case, it is not necessary to refer to a named event cycle definition, but one could be provided directly to the implementation at the time of invocation.
- In subscribe mode, an application declares its interest in an event cycle specification and then receives data asynchronously as they become available. In this case, it is necessary to identify the event cycle specification of interest by name.
- Polling mode supports synchronous execution of an event cycle, but in this case reference must be made to a named specification. In this sense, the polling mode is equivalent to the subscribe mode, with immediate unsubscription after a single report has been received.

Looking at the dock portal example, we can use the subscription mode to receive regular reports every time a new shipment enters the warehouse and activates the motion sensors. Tag observations would be collected over this period of time and processed into an appropriate report. At the end of the process, the collected data will be delivered back to the client application via a registered URI. The sequence diagram representing this scenario is depicted in Figure 7.6.

7.4 ALE Filtering and Aggregation

Each cycle is also associated with one or more reports; that is, information collected during the specified event cycle that must be communicated to a client application. Report specs provide instructions on how to process the captured tag codes in terms of data that should be included or excluded and aggregation actions whereby specified code ranges are grouped together.

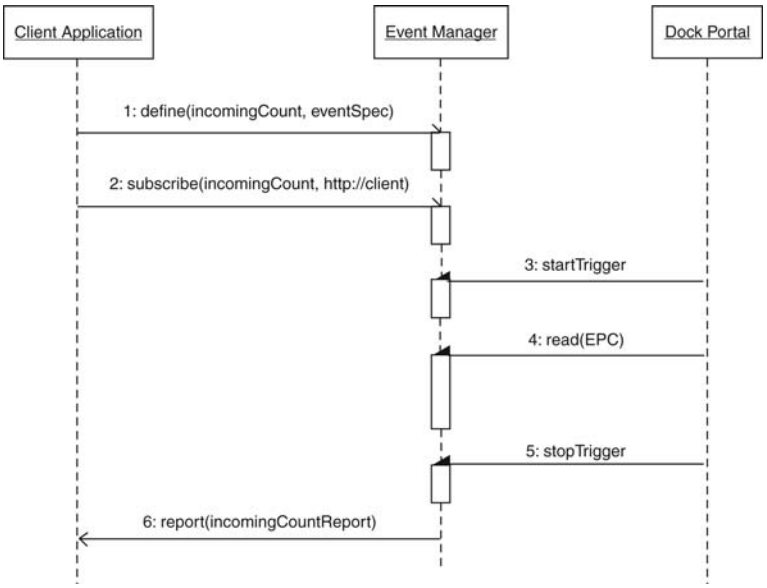

Fig. 7.6. A sequence diagram describing the middleware operation for the dock portal scenario.

Indeed, there are two principal mechanisms for the selection and processing of observations of interest:

- *Filtering* is used to select for or exclude from further processing tag codes that follow prescribed patterns.
- *Aggregation* is used when only total counts are needed for a particular group of identifiers. In this case, the main task is to define one or more patterns that define group membership.

Both mechanisms depend on the application of identifier patterns that is, standardized expressions of EPC code ranges. Such EPC patterns are part of the specification of EPC codes and are represented in URI form. Pattern specifications are modeled on standard regular expressions as introduced by Unix utilities (and used by numerous systems since) and should be familiar to most.

The structure of an EPC pattern follows the general format

```
urn:epc:pat:TagFormat:Filter.Company.Item.Serial
```

with `TagFormat` referring to one of the identifier schemes defined under EPCglobal; for example, SGTIN, SSCC or SGLN. The remainder of the pattern spec corresponds to the four fields of the EPC code with the same name. Regular expressions can be defined for each of these four fields to select a specific value or a range of values.

Looking at specific examples, we can relate patterns to actions that can be taken by observing incoming pallets in the canonical dock portal example. First, we consider the case where an application requires notification of the fact that the Innocent Drinks 1-liter carton of pomegranate, blueberry and acai smoothie with EPC `urn:epc:gid-96:20.300.4000` passes through the portal. The application would specify its interest in this item by including the following pattern in its event cycle spec:

```
urn:epc:pat:gid-96:20.300.4000
```

If the application was interested in all smoothie cartons regardless of their serial number, the pattern to be used would be

```
urn:epc:pat:gid-96:20.300.*
```

instead. Or if only a specific batch of serial numbers should be identified, for example because they must be discarded due to contamination, their range would be specified in a suitable pattern:

```
urn:epc:pat:gid-96:20.300.[4000-4100]
```

The same regular expressions can be used with the Item and Company fields of the pattern to produce groupings that include (or exclude, depending on the type of filter) a range of products irrespective of their serial number, several manufacturers, or specific serial numbers irrespective of manufacturer and product type as appropriate.

Note that an EPC code is included in the report associated with an event cycle spec only when it does not match any of the patterns in the exclude filter and at the same time matches one or more patterns in the include filter. In case the include filter is not defined in the spec, the default action is to include all observations regardless of their code.

Aggregation patterns have an extra feature in addition to filtering as described above. When a field is marked with the expression `X` as in

```
urn:epc:pat:sgtin-64:X.X.X.*
```

this implies that one group should be constructed for each distinct value of this field so that in the case above there will be a single group for each unique combination of company prefix, item reference, and filter value. Following from our previous example, using this expression an application can request that all Innocent Drinks products observed at the dock portal be recorded and grouped by product type:

```
urn:epc:pat:sgtin-64:3.20.X.*
```

7.5 Other RFID Middleware

The facilities provided by ALE are matched in other implementations of RFID middleware that provide equivalent functionality. Nevertheless, ALE is by

far the more robust and well-developed system but is rather limited in that it provides a very restrictive model of events, and indeed its facilities for the specification of complex events are grossly inadequate. Moreover, ALE offers only a specification but does not touch upon how it may be possible to support its specifications with particular techniques that can guarantee good performance and correct results. Both of these issues are the subject of intense research activity that has recently produced important results that may in the future find their way into the specification.

Adaptive smoothing

ALE implementations often employ a sliding window approach in applying temporal smoothing filters to the incoming observation streams. However, this approach introduces tension between the desire to ensure completeness and at the same time accurately capture tag dynamics. As a result, the choice of smaller or larger aggregation windows is not an adequate solution to the smoothing problem, as its choice skews the results toward one or the other direction.

In [60], the authors propose that an alternative approach is desirable, whereby incoming streams are treated as a statistical sampling of tags observed in reality and employ advanced sampling theory to drive the smoothing. Specifically, they propose an approach that employs binomial sampling and π-estimators to drive the adaptive automatic selection of an appropriate window size that ensures superior results. This technique is encapsulated in their so-called SMURF mechanism, which also provides a fairly complete specification for the definition of complex event queries.

Complex event processing

In translating observations to application-level events ALE adopts a rather simplistic approach that has been considerably extended in [125] with event specifications that support the flexible use of negation, parasitized predicates, and sliding windows, while maintaining a relatively compact language specification. The authors show that it is possible to implement this approach with good performance compared with high-performance dataflow processors. This language is more complex and more expressive than that proposed by SMURF, but it still supports only one level of mapping, from observations to application-level events.

Yet application-level events as employed by RFID middleware are still not at the appropriate level of abstraction for use with business logic, for example. What is needed instead is the facility to define hierarchies of complex or composite events of ever-increasing complexity in the same way as real-world semantics compose complex hierarchies of meaning.

Higher-level programming models

When it comes to developing specific applications, the primitives provided by RFID middleware can be too low-level and thus do not facilitate relatively rapid development [67]. As a result, alternative higher-level approaches are necessary to capture application requirements in more effective ways.

There is considerable interest in approaches that attempt to address this problem, which is not unique to RFID. One approach advocated in [21] is to introduce domain-specific models and associated frameworks that provide system-level abstractions and use those to develop applications. An alternative is proposed in [36, 38, 96] that favors the development of general purpose frameworks that provide general purpose primitives that can be used to program RFID systems irrespective of the application. Finally, an extension to the EPC protocols specifically to support location tracking at a high level in a manner similar to the approach adopted in cellular networks is proposed in [23].

However, such high-level approaches do not always cater well to the peculiarities of specific RFID systems but are forced to abstract to the lowest common denominator, thus missing opportunities for improved performance. One example of this is provided by the different capacities of near- and far-field tags: assuming that only the lower memory capacity of UHF is available leads to a failure to use local storage on HF tags that can very considerably improve system robustness [69].

7.6 Summary

Systems operating at UHF frequencies can deal concurrently with a large number of tags that may additionally be moving at relatively high speeds. The resulting observations create a stream of RFID readings that must be cleaned, smoothed, and further processed into higher-level events that can be consumed by applications. This task is carried out by RFID middleware, and in this chapter we have reviewed some common techniques and systems as well as abstractions used to programmatically control their operation.

8

Network Services

In the pursuit to minimize costs, RFID tags have very limited resources both in terms of storage and computational capability compared with other types of computing systems. In many cases, and most notably in the case of UHF tags, the tag will hold little more than a unique identifier code. As a result, it is necessary to employ supporting services to provide for full system functionality. For these services, the identifier can be seen as a handle to retrieve information about the entity tagged. In some ways, the RFID identifier is similar to a primary key in a database system, and in others it is similar to a fully qualified URL, in that it describes the location where related information can be obtained. In fact, there are two types of services that are required for a complete system, namely directory and repository services, which we discuss in this chapter.

8.1 RFID Services Overview

Looking back at the RFID stack in Figure 6.2, note that its upper three layers require access to the full information held about the entity identified by a particular tag. Assuming that the tag is embedded in an artifact, this information can include details about its manufacture (for example, the date it was produced, the name of the manufacturing facility, its expiration date if relevant, and so forth). This information can also include details about the different locations where the artifact has been observed (for example, its arrival at a wholesale facility) or events related to changes in ownership or repairs (for example, replacement of a component).

The initial location of an entity has special significance, specifically in the case of an artifact, as its identifier would most likely be assigned by its manufacturer and assigned a code within the number space it controls. In fact, this is the main premise in the construction of virtually all identifier schemes that we discussed previously and has one significant implication inthat it will always point back to the manufacturer of the artifact. This will remain so

G. Roussos, *Networked RFID*, DOI: 10.1007/978-1-84800-153-4_8,

throughout the lifetime of the artifact and despite several potential changes of ownership or custody.

As a result, in attempting to map a unique identifier into information held about the tagged object, the process will inevitably start at the manufacturer's destination, a fact that has considerable repercussions for the design of RFID services. A second consideration relates to the usual case where an artifact is passed from one owner to the next, and thus data captured within the authority of each of these owners have to be linked together to create its complete history. This serialized chaining of information presents particular problems, as it depends on each link to provide a reference to the next one. If for any reason the chain breaks, then it is not possible to trace artifact-related information into the future. Clearly, if persistence is to be guaranteed, a different approach will have to be adopted, one that ensures that information is available long after the demise of a particular manufacturer.

Finally, this resolution process only identifies the location where information is available and the way to access this location, but it does not provide the data themselves. Consequently, a related service should be available at the end-point presented by the mapping, which can be further queried using standard interfaces. Since each such repository of artifact information would serve a very large number of items, it would represent a significant data management challenge. It would also have to provide high throughput for incoming and outgoing data, as it would be required to serve numerous data capture events and service requests for information.

Although the details of different solutions differ in their specific choices on how to address these issues, they are nevertheless remarkably similar in principle. Nevertheless, each has a unique flavor, often defined on the basis of its particular context and heritage.

8.2 Identifier Resolution Services

Mapping identifier codes to network service locations is a relatively straightforward task that can be easily accommodated within current internet infrastructures. In this section, we will examine the facilities provided by two of the main RFID standards to support this task. Further, we will outline current thinking about the chaining problem discussed in the previous section and consider some possible solutions.

8.2.1 Object Naming Service

The first and simplest code resolution service is the Object Naming Service (ONS), which is a specification overlaid on top of the Domain Name System (DNS). The ONS has a simple and straightforward role: to take an EPC RFID identifier and return a Universal Resource Identifier (URI) that specifies the location of information related to this code.

DNS facilities for the ONS

A simple and quick way to develop a complete service infrastructure for code resolution is by re-using the directory capabilities of the DNS. The DNS is established as a mature and robust system that already supports a variety of directories over the internet including mappings between IP addresses and host names, mail exchanges for internet domains, and voice over IP (VoIP) numbers to user locations, among others. DNS is virtually ubiquitous in its operation across the globe and has developed into a dependable operational systems despite several design limitations.

As part of its evolution, the most recent DNS specifications provide for a generic type of Resource Record held by the system. A useful facility for RFID is provided by the so-called Naming Authority Pointers (NAPTR) that were introduced by RFC 2915 [82] for use by the Dynamic Delegation Discovery System (DDDS). The purpose of this feature is to allow for lazy binding of strings to data, and dynamic delegation. Delegation is supported through the use of regular expressions to specify a delegation point within some other namespace. For example, NAPTR records are employed extensively by the Session Initiation Protocol (SIP) to manage session locations dynamically. SIP is a signalling protocol operating at the application layer of the TCP/IP protocol stack and is used to create, modify, and terminate sessions between two or multiple participants and is typically used during internet telephone calls and multimedia distribution. The use of NAPTR allows users to move from one network location to another and still be followed by their sessions; for example, telephone calls can always be delivered to the current network location of the particular user, as their number is dynamically mapped to their temporary position.

In addition to providing a ready-made infrastructure in support of this service, the DNS also has other advantages. For example, it provides a well-established API and tools for management and access to the directory. Even the simplest computing platform would provide some form of access to the system and as such can gain access to the identifier mappings. The DNS is also developed on top of a hierarchical administrative and operational model and a delegation paradigm that fits well with the structure of virtually all identifier systems for which manager's are responsible for maintaining their own section of the code space. Finally, compared with other internet-enabled directory services, for example X.500 and LDAP, the DNS offers a wider deployment basis and far superior performance, especially for short but numerous queries which best fit the requirements of RFID systems. However, the DNS also has some limitations, especially related to its update protocols and security, which we further discuss in Chapter 9. Nevertheless, on balance, the DNS is a good candidate for the provision of such resolution services.

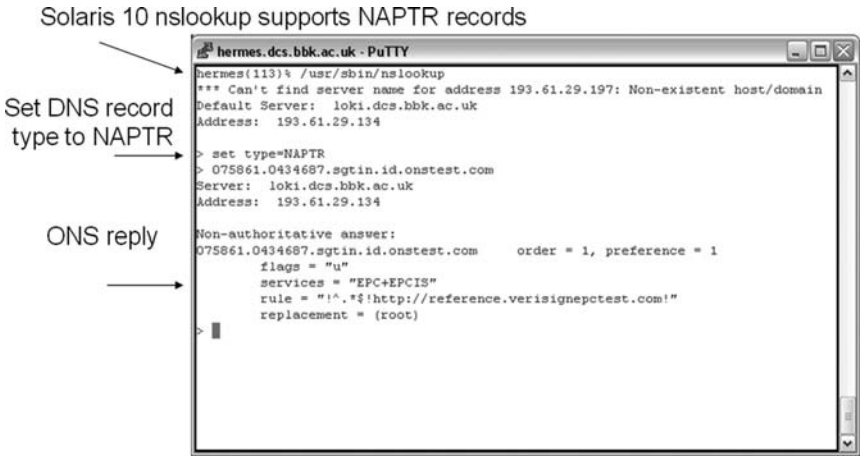

Fig. 8.1. Using the Unix nslookup utility to query the ONS.

ONS operation

The ONS specifies a thin protocol layer, which employs the NAPTR facility of the DNS to provide associations of EPC codes to URIs. DNS queries are issued using the EPC code reversed and appended to a pre-determined well-known domain name, which is currently defined to be `onsepc.com`. This is a relatively simple process and as noted earlier, benefits from numerous implementations of standard programming interfaces and utilities. To highlight the fact that this is so, in Figure 8.1 we show an example that uses the standard Unix nslookup utility to query the ONS.

Let's take a closer look at the steps of an ONS query and assume that the reader captures a 64-bit EPC code that is represented in hexadecimal from as

`0x10000000000006000000620`.

The conversion and query process takes the following steps:

Step 1: The raw EPC code is converted into URN format as

`urn:epc:id:sgtin:0614141.000024.400`

Step 2: The URN is passed onto the local ONS resolver:

`urn:epc:id:sgtin:0614141.000024.400`

Step 3: The ONS resolver converts the URN into the equivalent DNS NAPTR query (see Figure 8.1) by removing the URN prefix and the serial number from the EPC, reversing them and appending them to the onsepc.com domain name to obtain

`000024.0614141.sgtin.id.onsepc.com`

Step 4: The DNS returns the result set

O	P	F	Service	Regexp	R
0	0	U	EPC+EPCIS	`!^.*$!http://roussos.mobi/epcis!`	.
0	0	U	EPC+XMLRPC	`!^.*$!http://roussos.mobi/exist/epcis!`	.
0	1	U	EPC+WS	`!^.*$!http://roussos.mobi/ws/epcis!`	.

This process is depicted in Figure 8.2, which also shows how the DNS retrieves data by querying its servers. One last point is the interpretation of the NAPTR fields in the result set returned by the DNS. Note that the result of the query may include more than one records that describes different options for accessing the service. It is the responsibility of the client to parse the result set and select the appropriate service.

Each result consists of six fields, although the last one, called Replacement and denoted by R in the table above, is not used by the ONS. The first two fields, namely Order and Priority (denoted O and P), show the preference of the provider for the options provided.The third field is used to pass parameters of the NAPTR, but only one value is used in the ONS, where it is always set to "U." That means that the next field is not a DNS lookup but the output of the Regexp field is a URI. The Service field specifies the service available from this URI and may also specify the particular protocol that is used to communicate with a service. In the examples above, the service specified is EPC and the possible protocols for accessing the service can be one of many, in this case either a web service, XML RPC or the EPC Information Service. Using this information, a client now has adequate information to proceed with the next step and query for data.

A final point about the ONS has to do with management of the onsepc.com domain. This domain is owned and managed by EPCglobal, as are several of its top-level sub-domains that correspond to different types of identifiers. For example, SGTIN codes are grouped together under sgtin.id.onsepc.com as used above. Domains are delegated at the EPC Manager layer so that in the example below sub-domain 0614141.sgtin.id.onsepc.com would be managed and operated by the EPCglobal member with EPC Manager code 0614141. A mapping between EPC managers and their EAN.UCC codes as used in bar codes is also maintained on the ONS as a text record.

8.2.2 uID Resolution Service

An alternative solution to the re-use of existing DNS infrastructure is to develop a completely new network service specification that provides this simple mapping between identifier and URI via a secure overlay architecture. This approach is adopted by the ucode Resolution Service (RS) within the uID system (see Figure 8.3), which employs strong authentication and encryption to protect the system and the transmission of data. The complete redesign of the

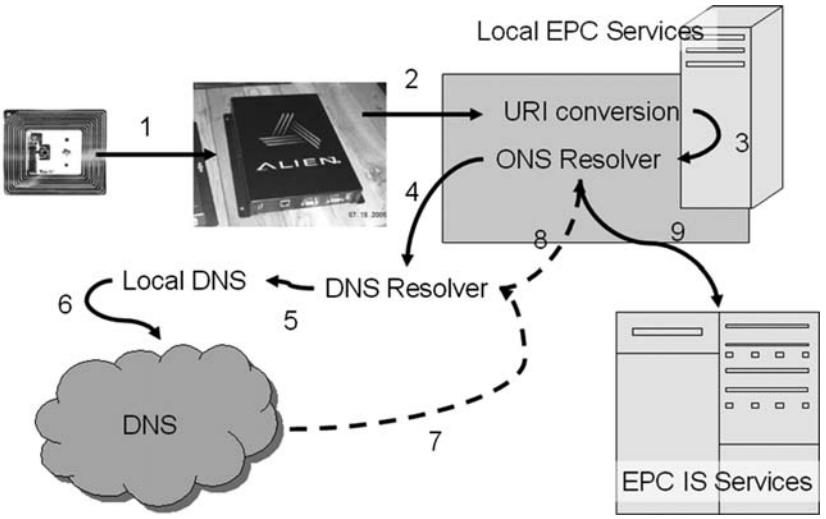

Fig. 8.2. Using the ONS to resolve an EPC code into its corresponding URI, which points to the associated EPC Information Service.

system also allows hierarchical resolution, meaning that mapping is carried out in two steps, first to identify the code classification scheme, and then to query the appropriate classification scheme provider for service details. However, as noted already in Chapter 2, the uID system is developed around a specific platform called eTRON and as a result is a far less attractive option in environments that are using multiple platforms. eTRON is an extension of TRON specifically for use with electronic commerce systems.

For those unfamiliar with TRON, it is a hardware and software platform that includes reference designs for common computing devices with a strong emphasis on embedded systems. Indeed, TRON is the platform of choice for a variety of industrial and consumer electronics but it is most often bundled under a manufacturer's brand. The main reason for its popularity, is that by co-developing hardware and software and insisting on a closed system approach, it offers very high levels of dependable operation and is far more robust than many general-purpose operating systems. TRON is particularly popular with Japanese and other Asian manufacturers and has a considerable following in the United States, although it is less well known in Europe.

Returning to ucode RS, the service has two main ingredients: authentication and secure messaging and a datastore management protocol. The former is provided by the eTP component of eTRON and is tightly integrated with this system. The latter provides a simple set of actions that can be taken to add and delete records and query the datastore. The authentication process establishes a session between client and server via mutual validation of certificates (issued via eTRON). eTP also provides facilities for the encryption of

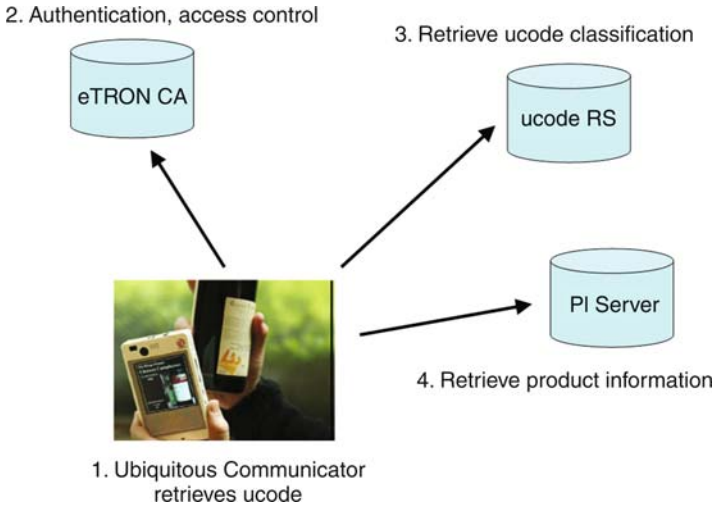

Fig. 8.3. Using the ucode Resolution Service to map a uID into its corresponding URI. Depending on the specific type of code represented by the uID, a secondary resolution service may be involved before product information can be retrieved from the ucode Product Information Service.

the transmitted data, as well as for the fragmentation of messages to support devices with lower resources.

When secure access to the datastore has been established, ucode RS offers three main code operations: registration, deletion, and resolution. Ucode registration inserts a ucode into the system and associates with it the address of a destination service that holds further information about the specific code. This address may take different forms and can be as simple as an IP address and a port number or as complex as a fully qualified URI. Note that due to the hierarchical structure of ucode, this address can either point to a so-called Product Information Service where authoritative information related to the identifier is held, or to a ucode resolution server that is responsible for the management of a particular sub-domain. For example, a ucode can encapsulate EPC identifiers and, as a result, when a code of this type is recognized, the RS would redirect to the appropriate resolver, which in this case would be the ONS. If the identifier is recognized as native ucode, then the resolution will point to the relevant Product Information service. Ucodes can also be removed from the system through deletion operations when the codes are no longer needed.

8.2.3 EPC Discovery Service

Moreover, both the ONS and uID RS are limited by the fact that they only retain the most recent service location related to a particular object ID; for

example, the URI published by the current owner of an artifact. This is hardly enough in many cases: in addition to the description of the current situation of the object, many pervasive computing applications need to gain access to historical use data collected during its lifetime or at least over a considerable length of time. This is not only due to the importance of context history for system adaptation but also because of a practical consideration: object IDs are assigned at production time from the address space controlled by their manufacturer, while the artifact itself changes ownership several times during its lifetime. As a result, such naive resolution of the object ID would point to the initial owner of the identifier rather than the current custodian of the artifact, and hence authoritative up-to-date information would no longer be available at the returned service location.

Moreover, the full object history is fragmented over different service locations corresponding to the different custodians that possessed the artifact at different times and a single service location is unable to provide the complete data set. As a consequence, using the ONS alone it is not possible to reliably support track-and-trace services, whereby a specific product item can be followed through its movement through the supply chain and located to its more recent custodian. In fact, one of the central arguments for the implementation of the EPCglobal network in terms of advantages to the consumer is its ability to considerably improve safety by automating product recalls through such product availability.

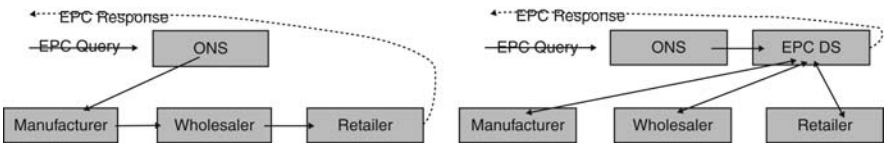

Fig. 8.4. Track-and-trace with the ONS (left) and the EPC Discovery Service (right).

Alternatively, rather than mapping an object ID to the initial URI provided by its manufacturer through the ONS, the resolution process could instead point to a secondary discovery service, which maintains the full record of the sequence of successive custodians from production to the most current recording available. This approach is adopted by the so-called EPC Discovery Service (yet to be finalized), which can be registered with the ONS and provide the list of URIs representing service points provided by all past custodians for a particular EPC.

This solution to maintaining a complete trace is preferable over the simpler alternative supported by the ONS, whereby the current custodian would be identified via a sequence of links through past holders (see Figure 8.4, left). Such chaining is vulnerable to broken links that can easily occur; for example, if any one of the custodians ceases to exist. In this case, one broken link

would be enough to result in the complete loss of the ability to trace the object history (Figure 8.4), while in the approach adopted by EPC DS it would only lead to the loss of the fragment of product item history under the control of the failed custodian.

8.3 Repository Services

Whatever the system employed, at the end of the resolution process, the inquirer has retrieved the details of a service location and its manner of access, which can now be used to manage information associated with a particular code. For example, in the previous section, we used the ONS to retrieve a number of alternative URIs that point to such services. The simplest of these alternatives would be to provide data management interfaces following the XML RPC protocol, which supports an easy standard way to query the repository.

As is suggested by its name, XML RPC (Extensible Markup Language Remote Procedure Call) uses XML to encode its calls and HTTP as a transport mechanism. Although none of the current systems specify a standard interface to identifier-related information using XML RPC, it is nevertheless a useful approach, especially in cases where the metadata stored in the repository are relatively simple and a lightweight approach is preferable. In the following example, we show a query-response pair that retrieves one of the attributes related to a particular tag. The call specifies an EPC identifier encoded in URN format and the names of the attributes that are requested

```
<?xml version="1.0"?> <methodCall>
  <methodName>epc_xmlrpc.getAttribute</methodName>
  <params>
    <param>
        <value><urn>epc:id:sgtin:0614141.000024.400</urn></value>
        <value><string>color</string></value>
    </param>
  </params>
</methodCall>
```

and the response returns the result of the query:

```
<?xml version="1.0"?> <methodResponse>
  <params>
    <param>
        <value><string>Blue</string></value>
    </param>
  </params>
</methodResponse>
```

Nevertheless, for more complex cases and especially when multiple custodians are involved such a metadata repository becomes a federated distributed database that should provide web service specifications to access object-specific data repositories. Such systems should also provide methods to record, retrieve, and modify event information for specific identifier codes. What does stand out, however, is the massive size and complexity of such a data repository, which—if successfully implemented—would be unique. This task is complicated by the complex network of trust domains, roles, and identities, which requires the careful management of relationships between authorization domains and conformance to diverse access policies and regulations. Yet these challenges are inadequately understood at the moment, as no system has attracted significant support.

8.3.1 EPC Information Service

One proposal put forward for the standardization of repository services comes from EPCglobal under their Information Service specification.[1] EPC IS provides an event-driven data model and two collections of interfaces with XML messaging bindings—one for recording information and one for querying the repository. EPC IS also specifies requirements for authentication and the cryptographic protection of XML messages, though it stops short of demanding a specific implementation or architecture.

In fact, an EPC IS server is not necessarily just a repository but can offer additional functionalities; for example, it can be a complete WMS. In practice, EPC IS interfaces can be used as a standard way for applications to access or update information about tagged entities, a facility that is useful in many circumstances, not only as a way to implement general-purpose repository services. It appears that EPCglobal are rather conservative regarding the actual deployment of such publicly available repository services and have opted for a staggered approach whereby complete infrastructures will emerge through a process of integration of individual systems. This evolution can be facilitated by such common interfaces, which are the sole target of version 1.0 of the standard, and moreover does not place any requirements on operational issues as is common with internet-based services.

EPC IS had been in development for several years before its first public release, in 2007. During this time, it had been extended considerably and, rather unsurprisingly, it has been far more tightly coupled to the requirements and specifics of supply chains. In any case, the current specification provides the following elements depicted in Figure 8.5:

- An event-based data model that describes the types of events specific identifiers can be involved in and their properties. Events and associated

[1] Note that EPC IS provides support only for RFID tags that encode EPC codes and cannot be used with any other type of identifier scheme.

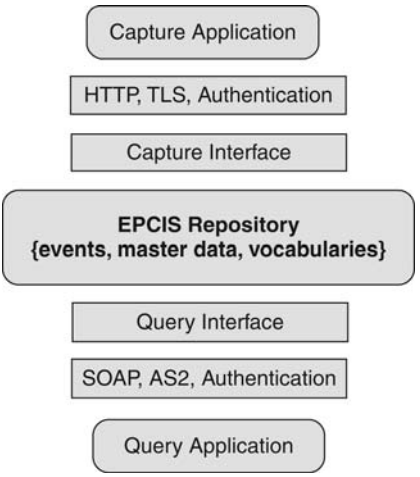

Fig. 8.5. The structure of the EPC Information Service.

meta-data are maintained by the EPC IS repository, which provides persistence and management.
- Interfaces for capturing events into the EPC IS repository and query interfaces for retrieving captured data. Capture is a one-way process from an application to the repository, while querying can be far more complex and can be conducted in synchronous or asynchronous mode.
- Bindings to XML messaging and transport over cryptographically protected channels.

Early versions of EPC IS defined fairly generic event templates that could be used in different domains without need for modification but more mature recent versions reflect typical situations often encountered in supply chain management. For instance, the majority of event attributes have been extended to directly relate to business transactions and business locations, and specific provisions have been made for particular processes reflecting the different stages of logistics. Nevertheless, EPC IS could still be used to support tagged entities outside this domain either by omitting those attributes that are specific to the supply chain or by extending the abstract event types, which is a feature provided by the specification.

Another noteworthy recent addition to EPC IS is its closer coupling with other GS1 provisions, in particular the GDSN (Global Data Synchronization Network; see Chapter 2), to provide so-called *master* data. Unlike event data that are produced as a result of object movement due to the conduct of specific business processes, master data provide slow-changing authoritative information about products. Events typically represent information recorded by observing tags (or other product identifiers, including bar codes), for instance

the fact that a particular object has been shipped through a specific dock portal of a named manufacturing facility. Event data would enter EPC IS in a constant stream and increase over time as a result of objects moving within the supply chain. They also provide the main input and output of EPC IS, and the details of the way that they are inserted or extracted by the service are specified through the capture and query interfaces. On the other hand, master data could include mappings between SGLNs (Serialized Global Location Numbers; see Chapter 5) and actual business locations or product classification categorizations.

Master data can also include additional user vocabularies; that is, user-defined attributes and keywords that describe the particulars of their specific application domain. For example, vocabularies can be defined to represent the different steps of a business process at the granularity of and using the terms appropriate for the specific domain. Event records would refer to these keywords and their interpretation would be provided by the master data held in the service. In any case, the primary objective of the EPC IS is distinct in that it aims to record and manage fast-changing data related to movement of products across the supply chain. In this context, master data can be viewed as a reference to authoritative registration information that provides a global vocabulary in support of EPC IS federation. Finally, note that the current version of the EPC IS specification does not define how master data are captured by or inserted into the system, so the actual automation between GDSN and EPC IS is not yet standardized.

Event model

The EPC IS event model aims to capture observations about where, when and why a specific RFID identifier has been recorded. The location and the time of such events are straightforward to record. The reason for the recording, however, can be rather involved and may include information relating to the details of the specific stage of the particular business process that caused the tag to be located at a certain place and time. For example, an object event could include pointers to the details of the shipment in which the associated product item was contained whether the tagged object is in transit or is ready for sale, and the environmental conditions during the capture (for instance, the temperature at that time). Since it is not possible to include all possible situations that can be relevant to a particular application domain, EPC IS supports the definition of vocabularies that extend the range of possible descriptions to fit the requirements of particular uses.

There are four main types of events defined by EPC IS: object, quantity, transaction, and aggregation. These definitions have been selected so as to represent typical data exchanges between businesses but are not intended to have a specific semantic interpretation within the standard. Nevertheless, the events would commonly have the following roles in capturing RFID data:

- *Object event.* This is the basic event type defined by EPC IS and aims to capture information related to a single observation that involves one or more physical object identifiers recognized by their unique EPC codes. For instance, in the standard dock door entry example, a single object event would be captured and recorded to associate a list of EPC codes with the receipt of a shipment of items (possibly packaged in cases and pallets) by a specific warehouse through a specific door.
- *Quantity event.* This event would typically be recorded to note a change in quantity of a particular type of product. For example, a quantity event would determine the increase in inventory levels of a particular product line stored by a particular distribution center.
- *Transaction event.* This event associates (or disassociates) one or more physical objects with a particular business transaction. For example, a transaction event would notify the shipment of an order from the manufacturer to the distribution center.
- *Aggregation event.* This event marks the assembly of more than one individual items into larger constellations. For example, an aggregation event would determine that several cases have been assembled into a single pallet. Note that such events are especially time sensitive and that they support both assembly and disassembly of composite entities but can also notify that a single element has been replaced with a certain constellation.

Users can add to these core types by specifying their own event types through a sub-classing mechanism. Aggregation events are of particular interest, and we will examine them in more detail later.

Interfaces

As noted earlier, EPC IS supports two types of interactions with client applications to

- capture event data generated by RFID readers and
- provide responses to queries related to specific tags or events.

Capturing event data is relatively straightforward in this scheme and involves a single capture call with event data passed as the parameter. EPC IS servers are not required to respond to confirm the successful receipt of events, and the standard makes the assumption that effective and efficient delivery will be facilitated through the implementation of enterprise message queuing infrastructures. This process is as simple as possible from the point of view of EPC IS and requires no further discussion.

On the other hand, querying is much more complex and there are several elements that must be developed to construct a complete solution. Similar to the modus operandi of RFID middleware, there are two alternative ways that information can be requested: either as a one-step question that requires immediate response or as a periodic process where new results have to be

retrieved at regular intervals. The former mode is relatively simple to implement, but the latter requires two additional facilities: a notification mechanism for the delivery of new result sets and an interface that can be used to control the execution of queries (see Figure 8.6).

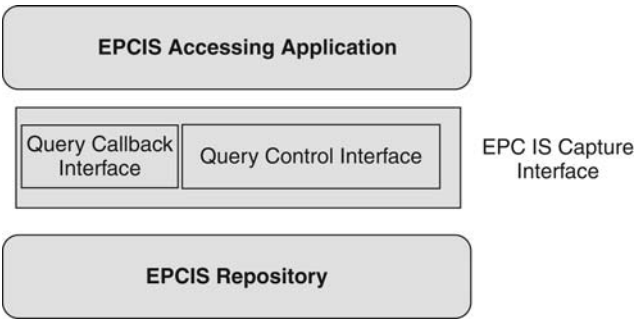

Fig. 8.6. EPC IS query interfaces.

Certainly, a remarkable aspect of the query interface is that it does not include a language in which to specify queries. Rather, the EPC IS specification offers just two parameterized queries (one for event and one for master data) that must be implemented by all systems and the method calls that are used to issue and manage them. The query control and callback interfaces provide equally basic facilities for issuing or subscribing to specific instantiations of these queries and for the receipt of results.

8.3.2 Containment Profiles

As noted earlier, aggregational events are a feature of EPC IS that merits further discussion. Aggregation events are used to compose single entities from multiple individual components so that a single handle can be used to reference the complete collection of individual elements. A typical case where this would be used is shipping, where several items are packaged in cases, which are subsequently assembled on pallets and then loaded on trucks for delivery. Figure 8.7 shows how the sequence of actions above can be captured by a series of aggregation events to establish a containment hierarchy of four levels.

There are several aspects of such containment profiles that require special mention:

- Containment relationships are time-sensitive; that is, they are initiated from and terminate at a specific time. As a result, such relationships do not exist outside their defined time frames.

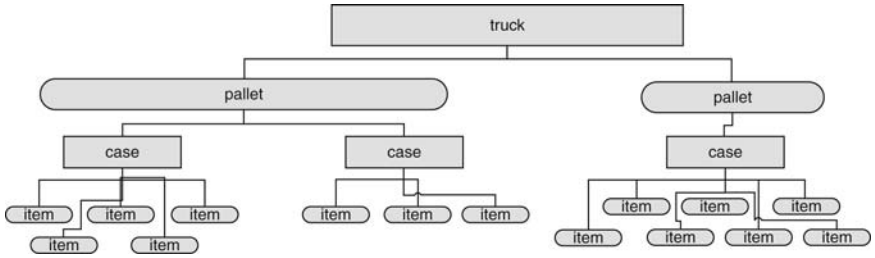

Fig. 8.7. A containment hierarchy defined through a sequence of EPC IS aggregation events.

- Each item defined through such aggregation events is uniquely identified through its EPC code, which is also used as the handle to define association and disassociation. A list of EPC codes is then related to a container identified by its so-called parent ID, which can also be an EPC code or any other URI.
- The aggregation hierarchy is defined implicitly by specifying the items included within a container.

A direct consequence of this is that queries that relate to such containment relationships are also time specific. For instance, it does not make sense to request the EPC codes of the items included in a specific container as this could change over time. Instead, it is meaningful to query for the items associated with a particular container at a specific point in time; for example, using the `MATCH_parentID` and the `GE_eventtime` attributes of the `SimpleEventQuery` template provided by EPC IS.

Why this feature is necessary is best highlighted in an example. Consider the case of an automobile that is made up of thousands of individual components, mostly sourced from third party manufacturers, which at a certain point in time come together to be assembled into a single entity. Over the lifetime of a particular car, these components will change as a result of maintenance, upgrades, or changing use. In most cases, the only requirement would be that the car as a whole be identified, but in others it would be necessary to identify individual components as well. The containment profile has been introduced to address exactly such time-dependent processes and is used within the EPC Information Service to group together components that are assembled into a new entity with its own unique EPC code. The composite object has an associated creation and expiration date, and its elements can be modified via updates to its aggregation event definition.

Clearly, queries related to containment hierarchies can return large amounts of data or require considerable processing resources to return a response. This is especially so when nested queries are to be iterated to the leaf nodes of the hierarchy tree over several layers of aggregation. This is an issue that has not been adequately explored within the current generation of EPC IS and

without a doubt requires further investigation and practical experience. As a sort-term measure, EPC IS allows refusals to serve excessively complex or large queries by returning error massages to this effect.

```
<rdf:RDF
    xmlns:rdf="http://www.w3.org/2/22/99-rdf-syntax-ns#"
    xmlns:rdf="http://www.w3.org/2000/01/rdf-schema#"
    xmlns:utad="urn:utad:schema:utad:base:0.0.0#"
    xmlns:utad="urn:utad:schema:pc:example:0.0.0#"
    <rdf:Description rdf:about="ucode:0123...cdef">
        <rdf:type rdf:resource="urn:utad:schema:pc:example:0.0.0#pc"/>
        <utad:version>0.0.0</utad:version>
    </rdf:Description>
/>
```

Fig. 8.8. A sample uTAD document.

8.3.3 ucode Product Information Service

Keeping with the approach adopted for the ucode RS, the ucode Product Information Service is also built on top of the core features of the eTRON platform. In this case, item descriptions are encapsulated within ubiquitous TRON Application Database (uTAD) documents (see Figure 8.8). These provide a standard way to record information about entities tagged with uID codes and can be distributed across several locations and represented in multiple formats, notably as RDF documents that can be exchanged between the ucode Product Information Service and a client (most likely a uID Communicator) so that the latter can acquire information related to a scanned RFID tag. Communication between the client and the server follows the approach described for interactions between clients and the uID RS, where sessioning, authentication, and security are handled by the relevant native eTP services.

8.4 Summary

Due to the restricted holding capacity and processing power of RFID, it is possible to maintain only a small fragment of information related to the tagged entity on the chip itself. As a result, many systems use network services to maintain and manage entity-related information that is both comprehensive and available. Typically, such services provide two stages: mapping RFID codes to service points where additional information can be obtained and repository services that can capture and retrieve information related to particular codes.

9

Privacy and Security

Issues related to the protection of user privacy and the security of systems that in some way employ RFID components have attracted considerable public interest. Many of the discussions refer to "RFID security" but what they often address are vulnerabilities that affect the system as a whole rather than specifically the operation of the reader and the tag. In many of these cases, it is the security of data repositories or middleware that has been compromised to reveal information collected through RFID. Nevertheless, RFID does indeed place extra stress on such systems due to the fact that it operates at the edge of the network where access cannot be as easily controlled, and that it can potentially create enormous amounts of detailed data. A considerable proportion of these data are harvested from end-user activities and can reveal the likes and dislikes or other private activities of specific individuals.

Moreover, reader and tag technologies have been related to a significant number of security compromises and present features that are easily exploited to invade the privacy of citizens and consumers. A surprising proportion of systems have employed the "security through obscurity" approach, which in modern computing is without doubt, a recipe for disaster. Rather than addressing the core limitations of their technologies, several providers have opted instead to attempt to prevent communication of these shortcomings to the general public through legal means, which in the long term severely damages the credibility of RFID technology.

In fact, RFID seems to lend itself naturally to hyperbole with highly exaggerated claims either for or against the use of the technology. In this chapter, we will attempt to separate fact from fiction and provide some useful insight on the causes of security and privacy problems and present potentially useful solutions. Still, several issues related to RFID remain open, especially in terms of privacy protection and the law, and we point these out wherever encountered.

G. Roussos, *Networked RFID*, DOI: 10.1007/978-1-84800-153-4_9,

9.1 RFID in the Public Eye

Over the past few years, the security of RFID has been an issue that has moved out of the specialist domain of computer security and into the mainstream. To a certain extent, this interest has been fueled by sensationalist coverage, but there is enough substance to warrant the wider involvement of the public in this discussion. In fact, this interest should be rather welcome as it is positive to be able to discuss all related issues before the technology is implemented on a very large scale.

Even within the industry, there are significantly divergent views as to how RFID should be approached. For example, in 2004 the CEO of RSA, a provider of security solutions to business, was quoted as saying that he "would be very worried of his privacy"[1] when it comes to using RFID. On the other hand, the European VP of EPCglobal at the same time stated that "there are more myths in RFID than there are in Greek mythology"[2] in his attempt to squash privacy concerns regarding Gen1 of the EPC Class 1 tag that had recently surfaced.

Boycotting brands

At roughly the same time, a number of high-profile disputes between retailers and suppliers and consumer organizations have also attracted considerable interest. Among them, the challenge presented by CASPIAN, a consumer group campaigning against privacy invasion by supermarkets, to Sisley, a manufacturer of clothing and a Benetton brand, stands as the first to bring the matter into focus. Sisley announced in a joint press release with Philips on the 11th of March 2003 its intention to tag with RFID every item it produces, causing the immediate reaction of CASPIAN, which issued a call just two days after that asking support for a global boycott of the company. Initially, Sisley refused to respond to this call or provide additional details of the proposed scheme. This was followed by massive interest by the media and consumers leading Sisley to withdraw its plans on April 4, 2003, just a few weeks after they were originally announced. Even so, the damage to the Sisley and Benetton brands was considerable and in the context of RFID they remain synonymous with the complete failure to anticipate public reaction.[3]

Several cases have followed this one including similar calls for a boycott on the UK supermarket chain Tesco, which provoked ongoing protests outside some of its stores, Levi Strauss for embedding RFID in its jeans, Gillette for using tags on razor blade replacements that are used to trigger cameras recording shopper faces, and Wal-Mart for its involvement in the EPCglobal trials. Most, if not all of these boycotts certainly have solid grounding, and

[1] Quoted by InfoWorld in April 2004.
[2] Quoted by the BBC, also in April 2004.
[3] More details are available through http://boycottbenetton.org.

the companies involved have taken few measures to provide even minimal protection to the privacy of consumers.

Covert consumer tracking

A prominent case where the operator of an RFID-enabled system has violated its own privacy policy in a remarkably obvious way, involves Metro Supermarkets in Germany [4]. This is despite the fact that this relates to perhaps the most high-profile pilot investigation of item-level tagging. In any case, the privacy policy clearly displayed at the entrance of the so-called Supermarket of the Future states that

- wherever RFID is used, this is made visible, and
- the chips exclusively store product data but no customer data.

Nevertheless, upon further investigation, visitors have discovered that this is clearly untrue and that individuals are issued loyalty cards that contain RFID tags (see Figure 9.1, taken from the website that accompanies [4]). To be sure, this blatant and rather naive violation of the supermarket's own policy seems unnecessary and unjustified. Yet it is generally in line with the recommendations of an early briefing of the MIT Auto-ID Center to sponsors that urged them to capitalize on consumer apathy and push for item-level tagging, thus creating a de facto situation before consumer organizations could react [26]. This advice was enacted during early trials of EPC, famously at the Broken Arrow, Oklahoma, branch of Wal-Mart, where tags were embedded within lipstick sticks without any attempt to inform customers.

Fig. 9.1. Loyalty card issued to the visitors of the Metro Supermarket of the Future and an x-ray image of the card showing the embedded tag. Photos by Peter Ehrentraut, FoeBuD.

Tagging humans

At the same time, RFID popularity increased steadily with the introduction of e-passports (the security provisions of which we have already criticized in

Chapter 2), the rapid growth of access control technologies, and the commercialization of one specific type of RFID technology that has sparked further debate. The so-called Verichip is an LF RFID tag that is specifically designed to be surgically inserted into the human body [52]. In this case, specifically designed implies that it has been cleared for this use by the FDA in the United States, and that it has a special covering that binds closely to human tissue so that its removal is very difficult and requires further surgery.

The Verichip was developed with the stated intention of being used in healthcare and security applications but has found a variety of new potential uses including proposals to tag migrant workers in the United States and schoolchildren. Besides the obvious ethical questions this raises, there are also health issues to be considered[4] as well as questions regarding the security provisions for such a sensitive system. Indeed, the Verichip provides no form of control over access to the personal codes it holds and as such it represents an open invitation to the world to read and trace the individual who carries it.

More interesting is the case of the now defunct CityWatcher.com, a supplier of CCTV services, which demanded that in order to gain access to the company's central video repository, employees would have to accept being tagged [39]. Only an employee with an authorized tag would be given access to these facilities. Yet when the company filed for bankruptcy and released its employees, it also refused to be held responsible for the removal of the tags. Employees have been left with the tags still embedded in their bodies and would have to cover the expenses of surgical removal with their own funds.

9.2 Attacks on RFID Security

Following the discussions of the previous chapters, it should be clear by now that while RFID systems are unique in that they include readers and tags and have to support their asymmetric mode of communication, they nevertheless are also open to the usual attacks that are common to all information systems. Indeed, network connectivity, openness, and the highly distributed nature of these systems imply that in practice they present numerous opportunities for attack and as a result standard security assessments and penetration prevention measures should be implemented.

Yet there are features of RFID systems that are unique, and in this section we will look at several of these closely. We begin with a discussion of issues related to the operation of the reader and the tag, and then we turn our attention to certain features of RFID services that raise security concerns.

[4] Current research is inconclusive on this matter, and this is unlikely to be resolved until long-term studies become available. There is mounting evidence, though, that the use of RFID tags on livestock is causing increased incidents of cancerous tumors.

Specifically, there are four aspects of RFID security that have special interest or present new twists on existing techniques:

- Counterfeiting or cloning tags to gain access to a controlled system or environment.
- Eavesdropping and replay attacks.
- Malware.
- Forward and backward data security.

Although these areas of concern are fairly straightforward to identify, early RFID systems provided no protection to tags [123]. Anyone with a reader capable of interrogating the tags was allowed access [45, Chapter 2], and this practice is still in use in a surprising number of commercial systems.

Cryptography and access control

The majority of recent commercial deployments of RFID for access control incorporate private key encryption and are relatively robust to cloning and counterfeiting. Notable exceptions are e-passports and the Verichip, for both of which cloning has been demonstrated since 2006. Another case where access control has been inadequately implemented is the aforementioned Supermarket of the Future (see Figure 9.2), where access to item-level product tags appears to have no password protection despite the fact that they support this feature. Although there is an argument to be made in support of this practice in that this is not a fully commercial venture but rather a proof of concept, it is nevertheless a situation characteristic of the point of view that security is often an afterthought and is not factored in during the design phases of a system.[5]

Fig. 9.2. Reprogramming tags at the Supermarket of the Future.

[5] The possibility of modifying the data stored on the tags of product items at the SoF was demonstrated by Lukas Grunwald (see Figure 9.2) using his open source utility available via http://www.rf-dump.org/. The same person first demonstrated cloning of a German e-passport.

Moreover, it appears that several systems employ weak cryptographic implementations or use inadequate random-number generators [8]. Such systems are amenable to exploitation using one of a number of techniques, including brute-force attacks, timing and power consumption analysis, and relay or data injection.

An example of a successful brute-force attack on the Texas Instruments Digital Signature Transponder (DST) used in the Mobile Speedpass key fob used for gas payments has been demonstrated by Johns Hopkins University and RSA Labs [12]. The attack shows the feasibility of an exhaustive key search of the 40-bit encryption mechanism employed. An interesting feature of this attack is that it was achieved using modest resources,[6] including an array of 16 low-cost commercially available FPGA boards, and simulating DST output using a programmable radio device (see Figure 9.3). This attack successfully recovered 5 keys in under 2 hours, thus showing that using a single pair of challenge/response values an attacker could clone the DST.

Fig. 9.3. Cracking the encryption of the TI DST tag used in the Mobile Speedpass system using low-cost FPGA hardware.

In addition to averting unauthorized access, RFID tags need to provide *forward and backward data security*: even when the code of a tag can be retrieved, it still should not be possible to trace the tag through past and future event records in which the tag was or will be involved. This is particularly relevant for Gen2 tags, which have very low computational resources and are thus unlikely to provide any complex cryptographic mechanisms in the near future. Instead, Gen2 tags support a single access password that controls communication with the tag. This is an inadequate mechanism, especially taking into account that many different organizations should have access to the tag data during its movement through the supply chain.

[6] The hardware costs were estimated to be about 3,500 USD. Full details of the implementation are available via: http://rfidanalysis.org/.

This shortcoming of the Gen2 tag is certainly a major barrier for its wider use especially in consumer applications, and the successful development of efficient and effective backward and forward security techniques would go a long way toward addressing these concerns. As a result, the past few years have witnessed an unprecedented explosion of interest in this area, with techniques ranging from hardware mechanisms to lightweight cryptography. This work should produce results both in the short and the long term, but there will certainly be a considerable delay before some of these techniques become part of RFID standards.

To give a brief example of the flavor of such work, note that one of the first mechanisms for forward security proposed in the literature employs hash chains to renew the information contained in the tag [85]. Backward security is a much harder problem and was identified as a concern much more recently, with [63] describing one possible approach to this problem using one-time pads.

Relay and replay attacks

Relay and replay attacks are the application of classic so-called man-in-the-middle techniques to RFID. This type of attack was described in the abstract by many authors during 2004 and 2005 and was demonstrated in practice in 2006. Although it is often stated that one of the best security features of RFID is its short range, especially for HF and LF systems, this does not take into account that transmissions can be overheard at far longer distances than the few centimeters that the specifications support. This fact makes relay attacks usable in many practical situations.

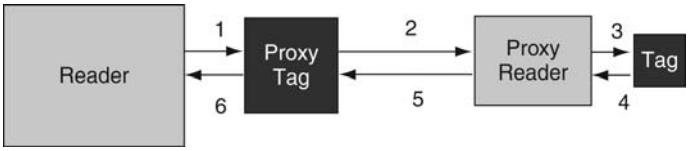

Fig. 9.4. RFID relay attack: using a leech and a ghost device to bypass challenge-response mechanisms.

In relay attacks, a "leech" device (a proxy reader) is positioned close to a valid RFID tag that will provide responses, and an associated "ghost" device (a proxy tag) is positioned near the reader that is targeted. The ghost device gains access to the system by relaying challenge-response queries between the target reader and tag via the leech. The complete process is outlined in Figure 9.4.

The leech and the ghost can be separated by a long distance; for example, they may communicate over a local wireless network or both be connected

to mobile phones communicating over a public cellular network, or the leech could be located in a fixed position and connected over the internet. The only requirement is that messages exchanged at steps 2 and 5 of the process be communicated quickly enough to avoid a time-out at the target reader.

A practical relay attack can be mounted on a ticketing system, for example. In this case, the leech would be positioned close to an authorized tag; for instance, close to the purse of a targeted commuter that holds their valid RFID ticket. The ghost device would be carried by the person who wishes to enter the public transport system and would be presented to the reader at the gates. The reader will begin the authentication and ticket validation process (step 1) and the ghost will simply relay all information to the leech (step 2), which would further replay to the target tag (step 3). The tag will respond to the instructions of the reader and release its information, which will follow the same path backward to the ghost and the reader at the gates (steps 4, 5, and 6). The reader will receive a valid response and operate the gates while charging the appropriate fare which will again be communicated to the tag through the same pathway.

The reason why this attack is practical is that the actual range over which messages in steps 3 and 4 can be transmitted is far greater than the few centimeters supported by ISO 14443, for example. Simple modifications to COTS equipment can provide effective operation for distances of up to 4 meters in steps 3 and 4. This attack was recently demonstrated in [54] against an ISO14443-A system (which employs fixed session identifiers and is somewhat easier to deal with compared with ISO14443-B systems, which are more commonly used in ticketing). Possible countermeasures against such attacks using a bounded communication limit are proposed in [34, 53].

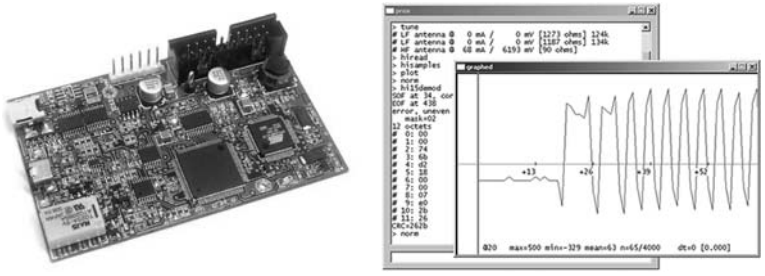

Fig. 9.5. The proxmark3 device and analysis of a captured signal. Photo by Jonathan Westhues, http://cq.cx.

It is also possible to capture and analyze the signal of a tag or a reader, and develop an attack depending on the specific details of the mode of communication. This is greatly facilitated by a special-purpose device (see Figure 9.5) that can be programmatically instructed to control every aspect of the received or emitted transmission. Such a device would be able to act as a tag

or a reader and analyze the signal received closely to measure, for instance, the power consumption to execute specific instructions (this information can be used for differential power analysis, discussed later in this chapter).

There are already several implementations of devices that can carry out this task, and principal among them is the so-called proxmark3, displayed in Figure 9.5.[7] It derives its flexibility from taking the received signal straight from the antenna to an analog-to-digital converter and then to an FPGA, which is more efficient in carrying out frequency filtering than a general purpose microcontroller. The result is finally passed on to the microprocessor, which is freed from the majority of the signal processing tasks. Since all signal-processing is performed either by software or by the FPGA, it is possible to reconfigure the device for a completely different communication scheme without any need for modifications to its hardware.

There are several other features of this device that make it particularly well-fit to explore vulnerabilities of RFID, including the fact that it can derive its timing from the carrier wave of a reader transmission, thus emulating the operation of a tag. Moreover, it is elementary to capture a transmission and replay it at a later stage or indeed to perform some detailed analysis off-line. For example, on the right of Figure 9.5, the trace shows the beginning of the `Inventory Response` command transmitted by an ISO 15693-compliant Texas Instruments Tag-It Plus tag in response to an `Inventory Request` command issued by the proxmark3.

Proxmarkii, a lower-capability predecessor of proxmark3, was used by its designer to clone a Verichip and many other similar systems that allowed full access without any authentication. The systems reported as vulnerable include military and other secure facilities, a fact that is remarkable in itself using the trivial technology provided. These devices are fairly sophisticated, but the same results can be achieved using much simpler electronics, and a design that can carry out the cloning functions only can be constructed at very low cost (see the proxmark website noted previously for more details).

Malware and other attacks

In addition to the attacks reported above, the increasing popularity of RFID and the higher complexity of the systems involved have acted as catalysts to explore new alternatives or more aggressive approaches to attack systems. In particular, two new types of threat have recently emerged that can have considerable bearing on the security of RFID although for the moment neither has been successfully applied to a real system.

One direction of investigation identified in [49] and implemented in proof of concept in [94] is the so-called RFID malware, also known with the misnomer RFID virus. At the core of this attack is the classic concept of buffer overflow

[7] Designed and developed by Jonathan Westhues. The complete specifications and sample software are available through http://cq.cx.

or data injection, a technique commonly used to gain privileged access to computer systems. In this case, the idea is that the identifier stored in a tag would be formed in such a way that it would cause the system to interpret the code as an instruction rather than data. This facility could then be used to steer the execution of a system toward a specific direction.

One potential scenario of application, proceeds by encoding SQL instructions in binary and storing them on the EPC memory of a Gen2 tag. However, following the discussion of Chapters 3 and 7, it should be clear that the technique as currently described has little if any chance of being practical due to the manner in which tag codes are read and processed. Although it is conceivable that there may be systems vulnerable to this attack, the vast majority of systems make the probability of this approach being successful extremely low. The excitement caused by the publication of this work can be attributed to its presentation,[8] which captured the public imagination despite its extremely limited scope.

A much more exciting possibility is presented by modern differential power analysis techniques that can be applied to RFID using the capabilities of devices similar to proxmark. Such cryptanalyses employ variations in the speed of computation (timing attacks) or power consumption (power analysis) of the tag with operations within the crypto-system to recover the key [16]. This type of attack falls under the general category of "side channel" cryptanalysis and has been described at a conceptual level, although it does not appear to have been implemented in practice yet.

This attack can be combined with physical manipulation of the tag chip to provide useful information to reverse engineer a tag that can be used to gain unauthorized access [5]. Finally, a proxmark-type device can be used for eavesdropping, whereby the attacker will remain silent but monitor communication between a reader and tag during a session. All these approaches provide fertile ground for further investigation and without a doubt will produce interesting results that may lead to practical attacks of considerable strength in due course.

Attacking RFID services

The attacks on readers and tags of course are only part of the story, as complete RFID systems can also be successfully exploited through their different components, for example by compromising their middleware. For instance, the EPCglobal network relies heavily on the DNS for its ONS service, which is convenient but possibly ill-advised, as the DNS has well-known vulnerabilities documented in RFC 3833 [113]. Among the limitations of the DNS, a few stand out that can play a role in circumventing ONS queries to rogue

[8] Reported as "does your cat carry a computer virus?" implying that the malware is carried into the home in the RFID collar of a pet. Such RFID collars are used to operate special doors for pets so that entry is allowed only to tagged animals.

servers, including packet interception, query prediction, cache poisoning, and server betrayal. Moreover, by using the DNS, the ONS inherits its well-known vulnerabilities, which can lead to highly successful denial of service attacks which directly affect the integrity and availability of the service.

Software errors in RFID middleware can also be used to compromise the system using standard techniques; for example, exploiting buffer overruns. Of course, such attacks are within the realm of traditional security and secure software development, but nevertheless RFID appears to make a difference in that systems are physically located at the network edge and as a consequence physical access is more likely compared with systems that operate within a trusted data center facility.

Other RFID network services, including those specified by EPCglobal and uID, would provide additional targets for potential attackers. However, the trusted operation of these and other RFID-related network services will remain mostly an open question due to their limited deployment in the field and their ongoing standardization. In any case, the scope and complexity of repository services in particular and the fact that they depend on multi-level trust hierarchies and federated databases raise significant concerns, which will have to be explored in the near future.

Mitigation

The only certain way to ensure that a tag will not divulge any information is of course to physically destroy it. Short of doing that, the implementation of strong cryptographic mechanisms should be a minimum requirement, and recent work has proved that this can be achieved in many cases; for example, through the lightweight implementation of AES mechanisms [30]. The usual advice regarding the longevity of cryptographic provisions also applies in this case; specifically it is preferable to avoid unpublished and proprietary algorithms, although in practice this can be difficult using the current generation of products.

Another possible way to prevent access to tags is by shielding them within a metal enclosure, which would absorb the reader signal altogether and completely prevent communication. Such an enclosure can be made out of a solid covering; for example, duct tape has been particularly popular for this task. A simple wire frame could also be used, as it would create a Faraday cage around the tag with similar effect. The latter approach has been adopted for the updated version of the US e-passport, although evidence suggests that it would still be possible to access the tag in this case if it is left even slightly open. There are several products currently marketed to cloak RFID embedded in a variety of documents, cards, and tickets, though it seems that good old duct tape and aluminum foil are still the most popular (see Figure 9.6).

Finally, two relatively recent proposals have sought to reduce the risks associated with tags still active in products after they are sold to consumers. The first concerns the introduction of the EPC Gen2 KILL command which

Fig. 9.6. A duct tape wallet used for cloaking RFID cards. Photo by Dustin Kirk.

will bring the tag into the permanent KILLED state, where it will not respond to any queries by readers, authorized or otherwise. This is seen primarily as a privacy and anti-counterfeiting mechanism, and its technical implementation is left to tag manufacturers. Typically, the kill state is implemented in software by setting a register in the tag memory, thus disabling the protocol state machine. However, since the tag would still need to execute a program to check the state of the register, it will still be vulnerable to differential power analysis. A more secure implementation would be to prevent the tag from being capable of operating by blowing up an embedded fuse, although this would require additional electronics and thus raise its cost.

An alternative to the kill command has been proposed in the form of the *clipped tag*, whereby the dipole antenna is made up of two parts, the second of which is detachable and would be removed at the point of sale. The extremely short length of the remaining antenna would provide very limited range and in most cases would require the reader to be almost in touch with the tag. This way, remote or casual scanning of the tag would be prevented while at the same time preserving its benefits for retailers.

9.3 Privacy Protection and RFID

There is no doubt that RFID presents many and multifarious challenges to privacy protection. The reason for this is twofold:

- RFID operation is transparent to the user. In this sense, RFID is the first technology that has become widely available from the new wave of computing, the so-called ubiquitous and pervasive computing. By deeply embedding numerous computing and wireless communication devices of very small size, this paradigm caters to "invisible, everywhere computing."
- Trust is a non-cognitive process and fundamentally depends on affective judgments [42]. As a result, it is hard if possible at all, to set a value to it or develop a formal representation, and thus it is hard to compute with. Note that trust is fundamentally different from trustworthiness, which is simply a value of the capability of a system to protect one's interests. Trust refers to its intentions as much as its capabilities.

In the following sections, we will explore some of the issues related to privacy protection and RFID, with particular emphasis on item-level tagging.

Item-level RFID can provide retailers with a unique source of information that can be employed for applications beyond supply chain management. Such applications may offer welcome new shopping facilities to consumers, but at the same time they also make possible new ways to violate personal privacy. Moreover, attacks on privacy enabled by item-level RFID are not limited to the physical confines of the store but to all purposes extend to any public space and even to the intimate space of the home. In these cases, the risk is not solely due to the use of RFID by the retailer but rather by third parties using the availability of the technology to mount independent attacks on consumers.

There are two main types of privacy attacks that can be developed by capitalizing on the widespread availability of item-level RFID tagging. In *tracking* attacks, the actions of individuals are recorded through the observation of RFID tags associated with their person and their future behaviors potentially inferred. For example, an RFID tag that remains embedded in an item of clothing long after its purchase can be used to identify its wearer wherever they go. *Information leaks* happen when personal or intimate information stored in RFID tags is revealed without the consent of its owner [45, Chapter 4]; for example, when personal details encoded in a tag are skimmed from an e-passport without owner consent. Both types of attacks become particularly likely when item-level tags affixed or embedded in consumer goods are not removed at the point of sale, so that stored identifiers can be retrieved by unauthorized readers, recorded, and processed without any visible indication to the user that this activity occurs.

A closer examination of tracking attacks identifies several distinct scenarios that become possible through item-level tagging [44]. For example, one of the earliest uses of RFID outside the supply chain that was explored during the development of the EPC system was in anti-theft applications. This is of particular relevance to items of small size but high value such as replacement razor blades, which are the most common targets of shoplifting. In this scenario, smart shelves would monitor high-value items placed on them, and in a case where a relatively large number are suddenly removed, a camera would take a photograph as evidence against a potential thief. But in practice it is hard to differentiate between lawful behavior and attempts to steal, and as a result photographs were taken in many more cases than was necessary. Although this may appear as a minor compromise of privacy it is nevertheless highly suggestive of the types of applications that are possible and how easy it is to develop applications using flawed heuristics.

Consumer privacy violations can be examined in finer granularity in terms of specific threats to pinpoint the many ways in which data analysis techniques, profile data, and the presence or absence of specific products can lead to violating one's rights [44]. As noted earlier, the widespread availability of RFID tagged products presents opportunities for covert data collection in locations and situations without the consent of the consumer. Individuals

associated with particular product item tags can in this way be linked with visits to specific locations at specific times. Even more, if readers observe several locations, sequences of visits can be reconstructed, and using simple inference techniques common behaviors, habits, or routines can be discovered.

Simpler but equally effective uses of the technology are also possible: a consumer carrying a particular type of product can be identified and approached with a discriminatory intention, for example because they carry a particular book title. A related use of the technology but with different intent would see the consumer being approached as a result of their possessing a particular item or brand that reveals their preferences. Identification of such preferences can be an effective marketing tool for competing retailers or simply used to identify the value of one's property and identify them as a worthwhile target for criminal intentions.

Such techniques are more effective when considering constellations rather than single products. Depending on the fact that a particular person is singularly associated with a specific product item may be haphazard. As products can be shared between several consumers, tracing collections of product identifiers moving together in a single constellation can provide much more accurate results. Even more so, when individual items are shifted from an established constellation into another, then it is possible to conclude that a transaction has taken place between the two persons involved.

Observing product items or product constellations over extended periods of time can provide adequate information to predict or infer preferences or behaviors. Although this is to some extent possible today through the use of loyalty schemes and cell phone records, tracking RFID tags does not require a contractual relationship with the consumer due to the technical characteristics of RFID. Moreover, RFID readers can be installed in such a way that there is no perceptible indication of their existence. Even when data collection is carried out in this way within the provisions of a mutually agreed upon contract, the wealth of information collected makes the indirectly enforced use of the technology through preferential pricing particularly attractive and can significantly reduce the capability of consumers to make free choices.

Last but not least, RFID tags can be used as a physical equivalent of cookies, with the vast majority of preferential pricing techniques developed for the web directly applicable [1, 2]. Indeed, historical information about acceptance or rejection of offers or other transaction opportunities can be stored on one or more tags carried by an individual and used to tailor future approaches to fit their profile. This is certainly feasible for the retailer that supplies the particular item used as the carrier, but due to the generally inadequate security provisions of RFID, this technique could well be accessible to third parties.

9.4 RFID and the Law

Although the consensus appears to be that RFID is a critical technology for future economic growth across several industrial sectors, it is also clear that its application must also be socially and politically acceptable, ethically admissible, and legally allowable. This aim becomes even more complex to achieve due to the universal scope of RFID technology, which must respect the policies, ethics, and law of every region and country where it is employed. To be sure, this is a challenging task, and in an attempt to make the main issues tractable from a computing perspective, in this section we will discuss the main considerations as they relate to the legal framework of the European Union.

9.4.1 Data Protection and Privacy

The EU founding treaty declares the fundamental freedoms that its citizens may expect, including liberty, democracy, and respect for human rights. Article 30 of the treaty in particular requires the enforcement of appropriate provisions for the protection of personal data, including the collection, storage, processing, analysis, and exchange of information. Moreover, Article 8 of its Charter of Fundamental Rights proclaims the protection of personal data as one of the freedoms that each citizen has a right to enjoy.

These principles are interpreted and implemented in practice through the legislative framework for data protection and privacy. The Data Protection Directive in particular has been developed with the aims of providing the general rules and the long-term vision, and to be robust despite technological innovations. Privacy protection is specifically addressed within the directive and is expressed in a way that is independent of the specific techniques and mechanisms employed in information processing and thus also applies in the case of RFID.

This directive is complemented by the more recent Privacy and Electronic Communications Directive (also known as the ePrivacy directive). This extension applies the general principles to the processing of personal data for the provision of public electronic communication services over public communication networks as well as to the recording and use of location data. It also specifies that direct marketing communications are only allowed when the recipient has agreed to be contacted in advance or in the context of an existing customer relationship, in which case companies can continue to market their own similar products on an opt-out basis. However, since RFID in most cases operates over private or corporate networks, it has been argued that the provisions of the ePrivacy directive do not apply, although this is only one interpretation, which does not take into account the case of the use of RFID readers in public spaces.

9.4.2 Commercial Transactions

The Electronic Commerce Directive regulates the process of contract offer and acceptance and applies to the fast-checkout process supported by RFID points of sale. The eCommerce directive has several provisions regarding appropriate ways of giving notifications of contractual terms and conditions and dictates that explicit consumer consent be given at all stages. Although exceptions apply to cases in which the interaction medium does not allow for information-rich interactions, RFID's predominantly silent operation stresses this requirement to its limit.

9.4.3 Governance

A central issue that affects the implementation of RFID is governance in the sense of access to RFID-related standards and infrastructures. The EU has been conceived as a vehicle for economic collaboration and has a tradition of creating a common open and non-discriminatory set of rules that strive to promote fairness and interoperable infrastructures. As such, its regulating bodies take a particularly negative view of any attempt to fragment public or shared infrastructures or the deployment of proprietary systems with the specific objective of preventing competitors from entering a market.

Arguably, the EPC system is tightly controlled by a group of companies and was developed with a view toward serving their interests and specific ends that relate to commercial, security, and political aspects of governance. Furthermore, the spirit of the community is one where protection is not limited to individuals but extends to companies, whose sensitive commercial information is also protected, as is the case with data within RFID-enabled business processes. As such, it is natural to expect the two opposite sides of the EPC proposition, namely rapid development of a new market sector and proprietary technology and infrastructures, will cause considerable friction and can potentially lead to closer regulation.

9.4.4 Spectrum Regulation

Recently, the EU has opted to liberate more spectrum for the growing demand for RFID usage, implemented through Decision 12 for RFID frequencies in the UHF band adopted by the Commission. This establishes a harmonized base for RFID applications across European states but nevertheless does not completely address the problem. In some cases, for example in distribution centers and shopping malls, it is necessary to operate hundreds or even thousands of readers in close proximity to each other in event driven mode. However, ETSI, the European Telecommunications Standards Institute, in standard EN 302 208 requires the use of Listen Before Talk to prevent a base station from transmitting if the channel is already occupied by another transmission. This limits the number of readers able to operate simultaneously in a particular

radio neighborhood to about 20 if all available channels are used and has some incompatibilities with Gen2 tag operation.

9.4.5 Environmental Issues

There are two directives that have already had very significant repercussions for electronics in general and RFID in particular, namely on waste electrical and electronic equipment (WEEE) and the restriction of the use of certain hazardous substances in electrical and electronic equipment (RoHS). RoHS in particular bans the use of certain hazardous substances that are rather common in electronics.

Relating to public health, the EU has some of the most strict regulation of the level of electromagnetic radiation that workers or the general public may be exposed to. Moreover, the Commission has in place a regular program of monitoring the possible effects of electromagnetic fields on human health through its Scientific Committees. Moreover, restrictions on EMF emissions from products available in any European state have been established to ensure the safety of both users and non-users. Although electromagnetic fields created by RFID equipment are generally low and thus exposure of the general public and workers is expected to be well below current limits, RFID nevertheless contributes to the total radiation in working and home environments and its widespread use may well have significant results especially when taking into account wireless networking technologies used in tandem.

9.5 Principles of Privacy Protection

Although there are already a large number of laws and regulations that define the rights and responsibilities of citizens as regards processing of personal data, RFID nevertheless seems to question the assumptions behind them and bring to the fore new questions and issues to be addressed. In the context of this discussion there are three core questions that have to be tackled:

1. *Initial entitlement.* This relates to the allocation of property rights regarding RFID-generated data, for example whether the record of trips recorded by a ticketing system is the property of the transport service or the commuter.
2. *Coercion and choice.* What is the tipping point where users do not have the freedom to exercise their free will but are de facto obliged to accept the use of RFID technology? For example, current ticketing policy in London may charge a non-holder of the Oyster card up to four times the fare that the holder pays for the same trip.
3. *Societal overrides.* When does society, regardless of the preference of particular individuals, have the right to access private data anyway.

These issues must be and are being debated in the context of particular nationalities and cultures and in all likelihood would have different answers in different cases. In any case, there are already a number of principles for the fair use of information agreed to in the context of the OECD that have been interpreted in [45] for the specific context of RFID. These seem to be an appropriate starting point for the discussions about the practical use of the technology:

- When a product contains an RFID tag, consumers should have the option to remove or destroy the tags when the product is purchased.
- Consumers should not be penalized for opting out of RFID use.
- RFID tags should not be used for price discrimination.
- Access to personal information recorded through the use of RFID should be given and mechanisms for modification of erroneous information should be available.
- Whenever an area is monitored for RFID, a clear and unambiguous notification should be provided.

9.6 Summary

The lack of adequate security and privacy protection provisions has sparked intense interest in this area of RFID technology. Several of the attacks recently demonstrated on RFID systems offer new twists on well-known techniques, though others have a uniquely RFID flavor. Moreover, little effort has been invested in providing comprehensive protection for the privacy of end users, especially within consumer systems. Despite this onerous situation, the main RFID standards have already been ratified and already hundreds of millions of RFID tags have entered circulation without adequate security and privacy protection provisions.

10

Epilogue

Since the initial development of communication through backscattering in the 1940s, RFID has grown into one of the most popular information and communication technologies with over four billion tags in active use. In previous chapters we looked at the state-of-the-art of RFID and associated network and software technologies required to build and operate complete open federated systems. In this chapter, we turn our attention to the future of RFID looking at work that aims to enhance the operation of tags and examining its relationship with pervasive computing. In particular, we examine the structure and implications of applications that become possible through the extensive item-level tagging of objects, places and individuals.

10.1 RFID Technology Development

One of the main obstacles to the wider adoption of RFID is the cost of the individual tag. Extensive item-level tagging of objects and metropolitan scale tagging of locations will not be possible without tags that cost only a few pennies. Taking consumer products, for example, it is relatively easy to confirm that this level of tag cost represents the tipping point after which tagging can become pervasive. Although item-level tagging for high-value products, for example formal business attire, has already been introduced, the highly competitive economics of commodity consumer goods prevents the implementation of RFID. Cost reductions can be achieved through novel material and packaging techniques, and we review some recent developments in this section.

Another way that RFID can offer more value as a system component is through the provision of additional sensing functionality. The role of RFID tags is currently to identify the presence or absence of particular entities through proximity detection but no other information about the context of capture. Simple low-power environmental sensors could be added to tags to offer this additional functionality, and considerable effort is being invested in this task.

G. Roussos, *Networked RFID*, DOI: 10.1007/978-1-84800-153-4_10,

Materials and packaging

Further reducing the cost of RFID tags to the level required for general-purpose item-level tagging above all requires innovation in production engineering. Driving the cost even lower than that would require considerable innovations in all aspects of the technology, primarily the introduction of fully printable tags. In the short term, the largest gains are expected from two techniques that are now entering maturity: tag antennas constructed using conductive ink and tag assembly processes.

Conductive ink is exactly what the term implies, its conductivity being due to the fact that it contains powdered silver and carbon. As a result it also has coupling capability and can be used to print tag antennas on a variety of materials including paper and self-adhesive thin film. Both UHF Gen2 tags displayed in Figures 4.8 and 4.9 have conductive ink antennas.

The savings from using conductive ink are due to reduced material costs and manufacturing costs. Using this approach, it is only necessary to apply the exact amount of ink to print the antenna, which is far more efficient than the construction of a metal antenna using copper or aluminum. It is also a simpler process since a printer similar to standard ink-jet technology can be used to create the antennas.

The second area where substantial savings can be realized is in the mass assembly of chips. The traditional way to attach the antennas to a chip has been to use the so-called flip-chip technique, whereby the chip attachment pads are treated and a small dot of solder is deposited on the attachments. The chip is then flipped over to bring the solder dots on top of the connectors of the antenna and melted using an ultrasonic process to complete the connection. Additional adhesive material may need to be added to form a stronger mechanical bond. This is a slow process and also quite costly when considering the target price of individual tags.

One alternative to flip-chip is the so-called fluidic self-assembly (FSA) technique, which has its origins in nanotechnology, where there is a need to assemble large numbers of very small devices in parallel. With FSA, individual parts are constructed in large numbers, separated, and placed at random in a fluid. The components are then transported using capillary forces created by heating the liquid to sites where they orient themselves and assemble in such a way that it is possible to achieve sub-micrometer alignment precision. The statistics of this process are such that the vast majority of components are aligned correctly with a small loss of components.

FSA and other similar techniques offer the promise of dramatically reduced assembly costs and also faster production of fully assembled tags that could help deliver large quantities of RFID to applications. However, such efforts have very high risks of failure; for example, parallel integrated chip assembly (PICA), a competitor to FSA that initially was claimed to be able to produce 70 billion tags per year, has failed to materialize due to the very high number of tags that failed to function properly.

Another area where materials play a central role in producing new generations of RFID tags is in the use of polymers to create a fully plastic tag. Early generations of this technology have already been demonstrated at lower frequencies and with elementary holding capacity, but work is under way on the development of HF tags. Such devices would cost only one or two cents when mass-produced due to the fact that they are manufactured through a relatively simple printing process. However, UHF tags are still well outside the reach of this technology and represent a major challenge, as they require all-printed high-performance transistors, which are hard to realize.

Augmented sensing for RFID

RFID is effective in identifying the presence or absence of tagged entities but can say little more about their situation. Even for the standard applications of RFID, there would be many cases where additional information about the setting of the auto-identification event would offer great benefits. Taking SCM as an example, it is often the case that products moving through the supply chain would be temperature-controlled, for example frozen products in the so-called cold supply chain. An additional temperature sensor that would indicate whether a certain threshold has been exceeded would provide considerable facility.

Unfortunately, even very simple temperature sensors require some form of battery to operate, and although active RFID tags integrating such sensors have been readily available for some time, this is not the case for passive tags. Nevertheless, recent proposals have identified ways in which temperature sensors can operate using passive tag technology [86], though actual prototypes are not yet in production. Other types of sensors based on simple micro-fluidic binary devices are also in development and in the near future it is expected that they could be used to identify pressure levels and simple forms of faults.

Surface acoustic wave technology

In addition to the LF, HF, and UHF frequencies, there are also a small number of tags that use microwaves for communication, usually in the 2.45 GHz band, which is also unregulated and available to industrial and medical applications. This frequency range has distinct advantages in that the antenna of the tag is very short (inversely proportional to the carrier wave frequency, see Chapter 4) and allows for a tag that occupies a very small area overall. It also offers the opportunity for far higher data transmission rates and thus higher-performance systems.

However, this band also has many limitations, especially the fact that metals and liquids absorb the majority of the transmitted energy and as a result severely affect the operation of the system. As a consequence, the majority of RFID systems that use this frequency are EAS tags commonly used with fabrics. Similar to UHF RFID, microwave systems use backscatter for

communication and due to various performance consideration they are often semi-passive rather than fully passive tags. This implies that the tag also carries a battery used to provide energy to the chip but not for communication.

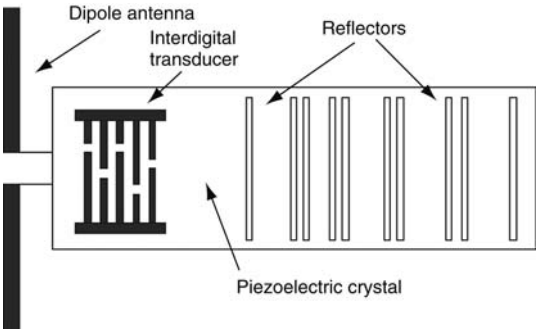

Fig. 10.1. The structure and operating principle of an SAW tag.

Recently there has been a lot of interest in the so-called surface acoustic wave (SAW) technology, which uses the piezoelectric effect to produce fully passive tags that offer good performance in the microwave range. In this case, energy is not harvested from the reader signal directly through a coupling effect but generated from ionic crystals embedded in the tag. One of the main components of an SAW tag is the interdigital transducer (IDT), a dual comb-like electrode structure that can convert microwaves into SAWs and vice versa. The remainder of the tag is made up of individual electrodes that are combined in a unique pattern (see Figure 10.1).

When the tag receives a microwave pulse generated by the reader, the dipole antenna captures the wave and supplies it to the IDT, which converts it into a SAW. This surface wave is then transmitted through the tag and a small proportion of it is reflected by the electrodes, the reflection pattern depending on the size and positioning of the electrodes. Note that the frequency of the surface wave generated by the IDT is the same as that of the incoming microwave signal.

A small part of the reflection from each individual electrode is returned to the IDT, which converts the surface waves back into microwaves that are supplied to the antenna for transmission. The resulting microwave transmissions can be received by the reader. The number of returned pulses depend on the number of electrodes in the tag, and the delay between them corresponds to the spacing of the electrodes. As a result, the spatial characteristics of the electrode layout directly influence the structure of the transmission from the tag, and specific configurations can be used to represent a sequence of binary digits.

One characteristic of the whole process is that due to the involvement of the IDT and the fact that SAWs travel at much slower speeds through the tag

material than microwaves in the air, the response from the tag has considerable lag, on the order of milliseconds. This fact can be used to address the main problem of effective operation of RFID in this range; that is, noise generated by the reflection of microwaves on metal surfaces in the environment. Reflections from such surfaces, even ones placed hundreds of meters away from the reader, would return to the reader much faster than the response of the tag and thus can be clearly identified and isolated from the transmission of the tag.

Of course, the amount of data that can be encoded in an SAW tag depends on the number of electrodes that can be embedded in the tag and the spatial density with which they can be placed next to each other. This number is relatively low with the current generation of SAW technology, but it is expected to increase in the coming years. Another potential limitation of SAW in practice is that due to regulatory restrictions on transmission in the microwave band, the pulse generated by the reader and thus the range of the system, are restricted and it is expected to be only a few meters with current technology. More details on SAW and its principles of operation can be found in [31].

Writable tags

One of the most interesting facilities of RFID, which is also the least explored in the context of open systems (possibly due to the lack of workable security models), is the ability of tags to not only store fixed data but also write to their memory information provided by the reader in its transmission. Although in most cases the capacity of the tag is limited to less than 4 Kbits, it is nevertheless adequate to hold enough information to develop interesting applications.

One such application brings together RFID and robotics in a system that adopts a mode of operation inspired by ant colonies [78]. Robotic agents roam space and use RFID tags to exchange messages and notify each other about the discoveries they have made, thus enabling a persistent, location-specific, delay-tolerant mode of communication. Using standard bio-inspired mathematical techniques, the robots can achieve shared objectives by working collectively in a decentralized manner that would have been impossible without the use of significant infrastructure dedicated to coordination. A similar exchange of messages using tags to hold the information is of course also possible between humans, though this mode of communication has yet to be explored to any depth [92].

10.2 RFID in Pervasive Computing

Over the past 20 years, rapid advances in micro-electromechanical systems (MEMS) and wireless communication, coupled with significant improvements

in electronics manufacturing processes, have enabled the development of low-cost, low-power, multi-functional sensor and actuator nodes that are small in size and communicate untethered over relatively short distances. As a result, it has become practical for information processing devices to be embedded into an increasing range of physical artifacts linked together in dense, self-configurable, and adaptively coordinated networks. These networks can be integrated into more traditional information and communication systems or operate autonomically to establish what is known as the *pervasive computing paradigm*.

Pervasive, ubiquitous, and sentient computing or ambient intelligence are all visions for the future of information technology, which is no longer seen as a separate world of digital representations accessible only through a computing device. Instead, computing and communications are intimately coupled with the physical world so that, for example, artifacts automatically activate computation. This vision was first expressed in [122], where pervasive computing was described as the technology that "activates the world, is invisible, everywhere computing that does not live on a personal device of any sort, but is in the woodwork everywhere, makes a computer so imbedded, so fitting, so natural, that we use it without even thinking about it."

This vision for the future of computing is gradually becoming fact, but its implementation at the scale called upon by its full application involves enormous costs and requires massive resources. As a result, and despite its numerous limitations, RFID provides an excellent candidate for the exploration of the opportunities and the implications of pervasive computing applications. Indeed, RFID has been used since the earliest ubiquitous computing experiments to provide object auto-identification functionality, and in the following paragraphs we will look at several applications where RFID has been used in innovative ways.

Context awareness

One of the main ingredients for successful pervasive computing systems is support for context awareness; that is, the ability of systems to self-adapt to fit their particular context of use [24, 106, 119]. Context awareness can vary from relatively simple tasks (for example, changing the resolution of media streams to facilitate the changing wireless bandwidth available to a mobile user) to the highly complex (for example, applications that have an appreciation for the social context of their use and adapt to address the requirements of leisure activities or of professional duties). The first step in this type of adaptation is to capture the context of use in a suitably expressive model, and RFID can provide this automatic identification functionality.

For example, a straightforward application of RFID is in supporting context awareness by supplying facts relating to proximity between specific objects and users [83] and thus the information required by the system to adapt

and offer flexible functionalities to its users. Typical examples of such adaptation using RFID include:

- medical facilities (for example, hospital beds that fit the needs of particular patients with particular ailments [10]),
- smart home environments that adapt to provide useful information specifically to the member of the family that is using a particular system (for example, the Aware Mirror, which identifies the member of the family in proximity by their use of their RFID-tagged toothbrush and retrieves and displays traffic and weather information specific to their trip to work [41]), and
- assisted-living applications for those with mental disabilities, including autistic children and elderly persons with Alzheimer's disease (for example, the iGlove, which employs RFID to record interactions with objects so as to learn common patterns of behavior with a view to helping the disabled with everyday activities [33]).

One particularly interesting application in this area is in preserving memories. This use of RFID was first explored in the context of the Cooltown project, where museum exhibits were installed with readers and connected to a server containing additional associated content [35]; for example, further suggested activities based on the phenomena discussed in the exhibitions (the system was used at the Exploratorium, a hands-on science museum). On entry, visitors were issued a card carrying a tag, which was the primary means of augmenting their interaction with the exhibits during their visit (for example, to operate embedded cameras). Such interactions could be easily recorded and hence the visitor tags could also be used for bookmarking exhibits of particular interest. These bookmarks could then be reconstructed as single web pages on the museum website and accessed using the unique ID of the visitor. In this way, visitors were able to reconstruct the most memorable aspects of their visit.

RFID and wireless sensor networks

Wireless sensor networks (WSN), also often called smart dust, are one of the main ingredients for pervasive computing scenarios. WSN nodes have a number of different sensing capabilities, which frequently include environmental conditions, the existence of specific chemicals, motion characteristics, and location. For example, WSN sensors can record temperature, humidity, vehicular movement, lighting conditions, pressure, changes in the magnetic field of the Earth, soil makeup, different types of bio-sensing, noise levels, the presence or absence of certain kinds of objects, and mechanical stress levels on attached objects. Nodes may also be able to detect movement characteristics of the objects they are embedded in such as speed, direction, and the size of the object itself. Sensor readings may produce a continuous signal, detect a

particular event or an event ID, or identify the location of an object (relative or absolute, numeric or semantic).

WSN nodes may also have action capabilities so that they can control mechanical, electric, or biological actuators and long-range optical communications. For example, WSN nodes include biomanipulators, fluidic microvalves, mixers and micropumps, adaptive optical micromirror arrays, rotary engine and microbial power systems, polymer micromachining processors, and last but not least longer-range communication systems, including atmospheric lasers.

One view of RFID is that it is the simplest possible—albeit more mature—type of wireless sensor network [55]. Seen in this light, RFID "networks" have two fundamental limitations that prevent them from supporting relatively complex data-processing operations: tags cannot initiate communication, and to operate they need to be within the range of a reader. One of the implications of this is that, contrary to fully functional wireless sensor networks (for example, systems made up of motes), RFID can only support one-hop networking, which severely limits its deployment capability.

On the other hand, RFID tags are increasingly capable of more advanced sensing than simple identification. For example, in the first section of this chapter we discussed how temperature sensing capability can be added to passive RFID and work in the same vein is being carried out by many researchers.[6] Moreover, devices that combine features of both RFID and motes are becoming increasingly available. For example, the Memory Spot chip by HP operates in a manner similar to RFID but provides far superior storage and data transmission capabilities, and the standardization work carried out within the IEEE RuBee working groups is expected to produce similar results in the short term. In fact, in their quest to achieve ever-higher longevity, wireless sensor network nodes are acquiring power-harvesting capabilities so as to become independent from battery power. Sources of power in the environment abound and are not restricted to electromagnetism but can also be thermal, vibrational, and piezoelectric.

RFID and motes have also been used side by side to provide complete systems. In Chapter 7, we discussed the use of motion sensors within RFID systems to improve the accuracy of supply chain applications, and in the same context temperature sensors are used to provide audit facilities for the conditions under which food is stored (for example, so that it is possible to provide proof that food safety regulations have been observed). Another application where motes and RFID have been used together is media production [79], and it is expected that more will find their way into actual deployments. Another application of RFID in WSN has been to provide location information for mobile sensor nodes that can retrieve details about their position and plan accordingly (see Figure 10.2). Finally, RFID has been used as part of the wireless sensor network platform itself to provide asynchronous communication by using a second channel that provides wakeup on demand [84], although its power-saving capability appears to be limited.

Fig. 10.2. A mobile sensor node (the Roomba cleaner) with a surrogate device capable of sensing its location from RFID tags.

Public displays

In Chapter 5, we looked briefly at the service identification schemes developed under the NFC Forum. One specific use scenario outlined by NFC is the communication between NFC-enabled mobile phones and interactive posters that provide links to further information or services related to their content. This is simply an informational facility, but it is possible to extend this idea to use RFID to enable social interactions around such public displays [81].

More advanced situated displays can sense colocation of several persons around them and respond by displaying information of common interest retrieved from its associated database of user profiles, thus providing prompts to discussions. The infrastructure required for this type of application need not be developed specifically for this task but can be based on the registration and session access facilities that would be particularly useful in a conference environment as detailed in [121]. Such ideas have been further explored to bring together combined physical and digital coincidence; that is, to investigate concurrent proximity of authorship and location again within an academic conference environment [68]. In this case, public displays are also used to explore common links to the same social groups as the enabler of further interaction between conference attendees.

10.3 RFID and Pervasive Retail

Supply chain management has been one of the three principal applications discussed in Chapter 2 as a way of motivating the discussions in this book. The

use of RFID in SCM aims to increase the efficiency of automatic identification by replacing lower-resolution technologies already in use and thus improve the quality of the captured product movement data, which can be used to better plan production and logistics. RFID is expected to achieve this by increasing the accuracy of the data, the speed by which it is captured, and the location within the replenishment cycle where this is due to happen.

The majority of these objectives can be achieved to a large extent by the current generation of container-level tagging which monitors product movement through the supply chain at the case and pallet levels. This is more than adequate information for improved forecasting and logistics and removes some of the most significant inefficiencies from the current system. Yet the long-term planning for SCM optimization calls for tracking the flow of individual items from the manufacturer through to the retailer and the consumer. This facility would provide a far more detailed view of the supply chain and even higher-quality predictions.

However, while container-level tagging has a clear value proposition, the cost-benefit analysis of item-level RFID is not equally positive. Regardless of the cost of the tag, which would certainly need to fall to the level of a few pennies, item-level tagging seems to be a high-risk strategy that can only produce marginal returns if monitoring of tags ends at the point of sale. Instead, this approach would be far better suited as the enabler of a variety of consumer applications that capitalize on the availability of tags in every single product item.

Such universal tagging combined with other pervasive computing technologies offers unique opportunities for conducting commerce. It allows for the personalization of the shopping experience, but more critically from a vendor perspective, it allows for the development of novel marketing and customer profile analysis tools that make item-level tagging far more appealing than its simple use in SCM. In the following sections, we present a selection of pervasive retail applications and discuss some of the findings of user studies with these systems.

10.3.1 The New Consumer

While suppliers and retailers have invested considerable effort in optimizing their supply chains through ECR and other initiatives, they must also respond to considerable social and market changes that directly affect consumer behaviors. Furthermore, competition in the grocery sector is intense, and as a result retailers must expect lower profit margins to remain competitive. At the same time, socio-demographic changes such as the increased number of dual-income, single-parent, and technology-familiar households have significantly altered shoppers' expectations, demands, and spending patterns [67]. Among other factors identified, a recent survey highlighted the decline of the "traditional family" [18]: it is estimated that in the United Kingdom by 2021 the average household size will be 2.21 persons, compared with an average of

2.70 in 1981. Moreover, a 30% increase in one-person households is expected, followed by a decrease of 33% for married couples. Finally, it is estimated that the share of total retail expenditure accounted for by groceries and food will decrease to 40% by 2004 compared with 50% in 1984.

These findings indicate that forging stronger consumer relationships and establishing successful consumer retention strategies will become increasingly important. Thus, appropriate relationship-building strategies for specific consumer groups must be a fundamental building block for the successful economic future of tomorrow's food retailers. A core component of such strategies is the development of attractive consumer experiences.

It is worth observing that the overall consumer shopping experience is affected by a number of store-related factors, which include ambience (temperature, scent, music, and so on) [9], service quality in the store [7], perceived image of the store [76], situational elements such as crowding, time, and budget availability by the consumer and so on [25]. Failure to provide an effective consumer experience results in increased consumer stress levels [7], which translate into consumer rejection of shopping and have been seen to conspire to create apathetic shoppers—consumers who have no interest in, or actively dislike, shopping and appear to endure rather than enjoy the whole experience [93].

At the heart of the matter lies the fact that in the new consumer situation traditional factors of competition (for example, price level, selection, and location), although still important, are no longer sufficient to achieve competitive differentiation. As a result, retailers must concentrate on enhancing the end-to-end shopping experience, aiming to win customer loyalty by inventing innovative ways of satisfying the new consumer needs.

10.3.2 Revisiting the Shopping Experience

In the previous section, we looked at how the changing market situation and the emergence of a new type of consumer are exerting pressure on retail shopping. In this section, we discuss opportunities offered by RFID and other pervasive computing technologies to enhance the shopping experience during a visit at the store but also in extending the relationship beyond the actual visit. Appropriate use of these technologies can make significant contributions towards meeting the needs of this new market and the requirements of the new consumer specifically.

According to traditional retail management theory, a shopping experience can be driven toward the maximization of efficiency or toward entertainment [77]. Yet the current situation demands that both these objectives must be met, clearly a challenging task. As a result, a critical component of any solution is that the stakeholders of the retail value chain must jointly discover the actual consumer needs and implement new shopping experiences. Of course, the rapid deployment of new technologies presents both opportunities and risks for those retailers eager to innovate. Nevertheless, the retail sector is

particularly IT-oriented, constantly experimenting with new technologies that promise to streamline and optimize core operations within the store or the warehouse and communication within the entire value chain.

One of the facilities that can increase supply chain efficiency is to improve on the current practice of late collection of sales data at the POS. Earlier collection of consumption data will provide more accurate demand forecasting and more agile replenishment strategies. For this reason, one possible extension of the ECR philosophy would use technology, to support the collection of data directly from the shelves or even earlier from the consumer's home. Indeed, the replenishment process starts when the consumer runs out of a particular product.

Gaining such early information and using it in supply chain optimization can potentially increase considerably the accuracy of predictive replenishment strategies to a degree that is well beyond what is possible today. RFID and other pervasive computing technologies can fulfill exactly this requirement. One possibility for achieving this has been explored by the MyGrocer project, a second-generation pervasive retail system and one of the early prototypes in this area.[1]

Nevertheless, extending the supply chain in this way has significant repercussions for the consumer, who is now involved in the data-processing pipeline. Pervasive commerce services collect and employ personal data associated with individual consumers in ways that can be used to reconstruct their private activities at an unprecedented level of detail. Several of these issues have been discussed in Chapter 9, where it was noted that no comprehensive solution to this exists yet. Moreover, recent studies indicate that the implementation of these technologies influences the consumption experience at home and creates a novel retailtainment experience during the supermarket visit [70]. This change transforms a particular retail ecology to another [102], and thus it should come as no surprise that consumers show considerable skepticism toward pervasive commerce value propositions.

10.3.3 Pervasive Retail Scenarios

The MyGrocer project developed three scenarios where RFID and a combination of other pervasive computing technologies were combined to deliver a new shopping experience. The three scenarios related to an RFID-enabled shopping cart used during the supermarket visit, the use of a mobile phone for on the move shopping, and finally an RFID-enabled home that monitors consumption.

[1] For a review of other work in this area, refer to [70]. Other approaches to conducting commerce using item-level RFID are explored in [28, 29, 46, 47]

RFID shopping cart

The consumer enters the supermarket and selects a "smart" shopping cart fitted with RFID readers and a tablet computer. She identifies herself to the system with her RFID-enabled loyalty card and her password. The system logs her in, responds with a welcome message, and then proceeds to present a "suggested" shopping list based on monitored home inventory and actual consumption data. The consumer starts shopping, walking through the supermarket aisles, picking up products from the shelves, and placing them into her cart.

The consumer sees on her suggested shopping list that she needs to buy a new bottle of shampoo. She asks the cart to display the quickest way to the shelf that holds the particular product, she picks it up from the shelf, and places it in her shopping cart. The cart identifies that the shampoo bottle has been placed in it and triggers the following event sequence: the product serial number is used to query discovery and product identification services and to retrieve related information that is used to update the shopping list and the total cost of the shopping cart contents. The consumer also needs to buy a specific brand of hair conditioner, and as it happens, the retailer is currently promoting a particular brand for customers with her profile. When she places the shampoo in the cart, the system displays the relevant offer on the screen together with instructions on the shortest path to the aisle and shelf where the conditioner is held.

The consumer continues her shopping, picking up and putting products into her cart, and the display constantly updates her suggested list and the list of cart contents. Later, she decides to remove one can of orange juice from her cart and replace it on the supermarket shelf. The system updates the shopping list with the new total amount and the new contents of the cart.

When the items on the shopping list are exhausted, the consumer proceeds to the checkout. When she approaches the till, the system re-scans all the items in her shopping cart, calculates the total value of the products, displays that information on the till display, and prints out a receipt. The consumer pays at the till or charges everything to her account.

Smart home scenario

The consumer returns home and places her shopping items into her RFID-enabled storage (for example her fridge, cupboards, and so on). New product information is recorded by her home server and consolidated to the home inventory data. The home maintains data on inventory levels as well as consumption. Periodically, the consumer gives permission to her home server to upload her new shopping list to the system.

On-the-move

The following day, on her way to work, the consumer uses her mobile phone to check which products she needs to replenish before the weekend. After logging in, the system displays her current home inventory and her suggested shopping list. The consumer decides to add new items to her shopping list for the dinner party she is planning for Saturday night. When she is happy with the contents of her new shopping list, she looks at the totals. The system displays the cost of her shopping at her usual supermarket. However, she is unhappy with the price and she decides to look for cheaper alternatives by initiating a reverse auction between available retailers.

A short while later, the consumer receives offers by different retailers and selects the best. The consumer selects "home delivery" and confirms the order. Later in the day, the system notifies the consumer via SMS to her mobile phone that baby diapers are going to run out in the following hours and requests confirmation of her instant replenishment order. The consumer confirms and the order is placed.

10.3.4 A Case Study in Pervasive Retail

The MyGrocer project identified the following features of a pervasive retail system as critical for its success:

(a) a compelling interface, enabling seamless interaction between shopper and system,
(b) an efficient product scanning mechanism that would minimize shopper involvement, and
(c) an integrated information system that would enable the provision of value-added retail services.

The details relating to the technical development of this system have been presented elsewhere [99] although since that time a variety of new technologies and service components have become available and a similar system built today would would be quite different. Nevertheless, the main ingredients of the user interface developed by MyGrocer are still relevant and in this section we will review the functionality of the smart shopping cart in particular.

Experience design

The shopping cart was modeled around a touchscreen mounted on it as seen in Figure 10.3. Five distinct areas were identified to facilitate the shopping experience:

- *Shopping cart content*: lists products placed in the shopping cart;
- *Total cost*: shows the total value of products in the cart and the total amount of reductions due to promotions and offers;

- *Shopping list*: lists products that are marked as regular buys and those that have been indicated as for replenishment due to consumption;
- *Offers and promotions*: details offers and promotions for the particular shopper;
- *Additional information*: displays either detailed information on the last product scanned (for example, weight, cost, nutritional value and so forth) or on the terms and conditions of the last promotion triggered.

Unlike general-purpose product browsing appliances, where the design should address all possible cognitive processes of the user, adopting the so-called appliance argument offers several benefits. The consumer may still perform all the activities usually associated with web browsing and shopping; that is, finding products, information, and general browsing, transacting, and communicating. However, the particular focus of the system implies that all of the user's goals may be achieved much more efficiently. Whenever additional computing power or storage resources are required, such problems may be offloaded to a central server and results exchanged wirelessly.

Formative evaluation

In project MyGrocer, the collection of user requirements aimed at understanding both how to integrate pervasive retail with the systems of the supply chain actors and how to cater to the needs of the end users. To this end, research was carried out to assess the appeal of pervasive retail as a value proposition to the consumer as well as to identify barriers to acceptance. The approach adopted was qualitative in nature and used focus groups. Market Analysis, a market research firm, was commissioned to conduct the field research. The

Fig. 10.3. Prototype shopping cart implementation: shopping cart with tablet PC, wireless network, and RFID reader (left). User interface: login screen, shopping options, including shopping cart content monitor, shopping list, total shopping cart content cost, offers and personalized promotions, additional product information, and finally fast checkout (right).

target audience consisted of women age 25–34 who are responsible for grocery shopping within their household and demonstrated some familiarity with information and communication technologies, either as regular users of personal computers and mobile telephony at home or at work; women with the same background but from the 35–50 age range; married couples with both partners between the ages of 25 and 34, both responsible for shopping and with backgrounds similar to groups one and two; and couples as in the previous group but from the 35–50 age range.

During the discussion, the participants were first introduced to pervasive retail concepts through a presentation based on concept drawings with explanatory text, which the moderator used to discuss selected usage scenarios. Following the introduction, participants were encouraged to discuss their thoughts, feelings, and reactions to this novel approach to retail as well as express their responses regarding attitudes and purchase behavior in this environment. The discussions of all groups were recorded in audio and video with the permission of the participants. At the end of the discussions, participants were given a voucher for one of the retailers participating in the project.

The pervasive retail proposition attracted significant interest from most participants as a shopping option in addition to the ones available today. In particular, the in-store scenario received the most favorable response, with the main benefits perceived to be the improvement of the shopping experience, which was understood to be faster, easier, and offering better value for money. The features that proved most attractive were:

- constant awareness of the total cost of the shopping cart contents which offers the opportunity to accurately control spending during a shopping trip,
- access to complete and accurate descriptions of products, including price, size, ingredients, suitability for particular uses, and so forth,
- the ability to compare the value of similar products,
- the provision of personalized, targeted promotions that reflect the individual consumer profile in addition to the usual generic promotions as well as the fact that the participants could access all offers available in the specific supermarket at a single contact point,
- the proposed in-store navigation system, especially in the case of hypermarkets, where orientation is particularly complex, and
- the smart checkout and the ability to bypass queues and reduce waiting time.

However, the findings highlighted one of the main concerns of the participants: the use of personalized purchase statistics by the retailer and collaborating service providers. A large number of participants were particularly concerned about the collection and storage of personal data, even though they were aware of the provisions (albeit not the practicalities) of the data protection act. Their negative reaction to data collection was triggered primarily after the eponymous authentication during the initial use of the shopping

cart when, after entering personal identification credentials, they were presented with a personalized shopping list derived through the analysis of their purchase history.

The two main issues arising related to the immediate recognition of the fact that for the construction of the personalized shopping list their data are recorded, preserved, and processed. This reaction was more pronounced when trust of third parties was also involved—a common feature of pervasive retail systems. The main source of concern was that private data collected in the sheltered space of the home could be delivered to external sources without the explicit consent of the consumer. The vast majority of participants did not trust a service provider to protect their privacy, irrespective of whether it was a contractual obligation or not.

Another major concern related to the overall shopping experience, which was perceived to point toward a technology-controlled, fully standardized lifestyle. Two issues interrelate on this point. On the one hand, participants rejected the claim that a software system could accurately predict their wishes just by collecting historical data and monitoring habitual purchases. Indeed, due to its ability to preempt their wishes, this aspect of the system appeared patronizing and overtly rationalized but most importantly contrary to the experience of being human. In fact, the majority of participants discarded the possibility of a computer system that could successfully predict their wishes, while some of them were offended by this suggestion. On the other hand, the participants of the study perceived that the pervasive retail system reviewed promoted primarily the interests of the supplier, while the consumer only received marginal benefits.

Finally, several participants observed that adoption of pervasive retail could result in a transformation of traditional family roles. They emphasized that product selection and maintenance of appropriate home inventory levels are a means to establish roles within the family unit and the responsibility to carry out these activities an integral part of the identity of the person or persons in charge. Elimination of this responsibility was perceived as undermining the status quo, and pervasive retail was consequently treated with mistrust and hostility.

Consumer trust

While the value proposition of MyGrocer did indeed attract substantial interest from consumers, at the same time it was also evident that, if implemented as described in the user scenarios, several aspects of the system would create considerable friction and would pose barriers for the wider adoption of the system. A short-term solution to this problem was to invite loyalty club members to participate in the second phase of system testing, a fact that offered two distinct advantages: it capitalized on the established trust relationship between the consumer and the supermarket and allowed for the regulation of the relationship via a contractual agreement.

Indeed, participation in a loyalty program often implies a relationship built over a longer period of time, which fosters mutual trust and helps develop a set of reasonable expectations. Furthermore, having agreed on a contract, the two parties clearly understand their rights and responsibilities to each other and have an explicit set of rules for interacting. It is thus easier to explore the extension of the relationship to include the new ubiquitous commerce services. In practice, this approach proved very successful and allowed for the evaluation of the deployed system in conditions where security and privacy were not the dominant factors.

Arguably, some of the research findings of the previous section should be seen within the context of the study, especially with respect to the evolution of retail practice. To this end, we should briefly discuss the timeliness of the emergence of supermarkets as the dominant retailing option and of the adoption of credit cards in Greece. Until the early 1980s most grocery shopping was done in small, neighborhood shops with very few large supermarkets, primarily located in the two main metropolitan areas in the south and the north. Over the decade, this situation changed at an accelerated pace, with most of the local shops disappearing and by the end of the decade almost completely being replaced by super and hypermarkets. Today, even in rural areas most grocery shopping is done in supermarkets that belong to one of the five national chains.

Fig. 10.4. Shopping with a smart cart, and quick checkout during the MyGrocer trials.

The end of the 1980s also witnessed the rapid adoption of credit cards for electronic payment. Deregulation of consumer loans at the beginning of the decade played a key role in making credit cards commonplace and accessible to most within a few years. Since the mid-1990s, supermarket shopping and payment by credit card has been as common as in any other Western European country or the United States, although middle-aged Greeks still prefer to use cash and would opt to shop from a smaller grocer if possible. At the same time, the traditional family roles have also changed significantly. With the urbanization of the population in the 1950s, more women entered higher education and joined the professions.

Today, especially in urban areas and with younger couples, the norm is that both partners work outside the home and share the responsibility of running the household. In particular, it is likely that either the husband or the wife would be responsible for the replenishment of home supplies, although women take up this role more often than men, certainly in middle-aged couples.

MyGrocer highlighted several aspects of researching pervasive computing systems that may have wider implications. Unlike more traditional information systems, where interaction is mediated by a computing device, for example a desktop or mobile computer, in pervasive computing things seem to happen transparently in a space that cannot be approximated through a real or even a representational one. Thus, users are confused by their lack of appropriate language to describe it and will need to be offered other abstractions to replace the device. In our case, consumers attempted to express their opinions by anthropomorphizing system behavior so they could relate it to their existing experiences.

One aspect that appears to be highly relevant (but we were unable to investigate in-depth) is the question of how pre-existing attitudes toward privacy affect consumer views of ubiquitous commerce. Previous studies have indicated that there are considerable variations in how people deal with such issues, and there is a reasonable expectation that some of these attitudes would directly affect their perceptions of pervasive commerce.

The novelty of pervasive computing means that for more significant observations to be made, one has to allow for an extended period of interaction with the system. Unlike system functionality, habits and practices take much longer to develop and often what seems novel and threatening at first glance quickly becomes part of the routine. Longer-term implications of use cannot be discovered without ethnographic studies. Of course, the problem with this approach is the very high cost for deploying and maintaining the required infrastructure at the required scale and time frame.

This last observation points to another aspect of trust that is often overlooked. Indeed, trust in information systems is often seen in the tradition of cognitive psychology, which was also the basis for machine learning and artificial intelligence in the early 1960s. While this approach has made considerable contributions to computer science and systems engineering, we expect that it may not facilitate further development of our understanding of trust. Indeed,

in the technical literature, trust is considered as a purely cognitive process. It is often treated as a utility function that system users try to maximize for their own benefit. We believe that this approach is better suited as a measure of trustworthiness, which is quite different from trust, and moreover that trust is a non-cognitive function that cannot always be approximated well by mathematical constructs. Hence, in the intimate computing context of ubiquitous commerce, the development of concepts of trust on this basis is of restricted use.

Approaching trust within its social context may provide a more productive alternative. It appears that this is particularly relevant in cases where there is little information on which to make a judgment of the trustworthiness of the other party and thus the decision to trust or not depends mainly on non-cognitive elements. Clearly this aspect of trust played some role in the case of our studies, where the information to make an unambiguous and provable trust judgment was not available. In fact, the system frequently created significant levels of stress to the participants, which they could not justify in concrete and objective terms.

10.4 Summary and Conclusions

Several of us would have you believe that RFID is the greatest technological invention: it will deliver cheaper, better quality, and safer food for the global market; it will simplify the manufacturing of cars and airplanes; it will save the environment by allowing every single product to be recycled; it will save human lives by preventing medical mistakes; it will make the world a safer place by averting acts of terrorism; it will do away with counterfeiting, especially of drugs; and of course it will spark the next computing revolution by creating the Internet of Things. Few computer technologies have sparked such excitement.

Yet not everyone agrees with this view. Some believe that RFID is the source of many evils: it will restrict our ability to exercise our civil liberties; it will generate unique opportunities for attacks on privacy, well beyond what is currently possible; it will make us more predictable and diminish choice and free will; it will create a situation where every object spams passersby in its attempt to attract attention; and it will further pollute the environment. Even the more adverse of these claims can be, and have been, argued for with conviction. Few computing technologies have been anticipated with such apprehension.

But where does the truth lie? Are any of these views accurate, or are they all confused notions caused by our over-inflated expectations for yet another new technology—an attitude certainly not alien to computing? The fact of the matter could be a lot simpler: despite its long history, RFID is merely a first-generation pervasive computing technology that is very well fit for specific

applications. It can deliver only some of the revolutionary or indeed the deeply dark applications and systems foreseen by either its proponents or its critics.

For all its limitations, RFID is a significant milestone in the history of computing, as it marks the first practical technology to intimately link the material and the digital and thus bring about pervasive computing. Indeed, in some ways the controversy around RFID is only by reflection, or the consequence of the fact that "some are born great, some achieve greatness, and some have greatness thrust upon them." For RFID, greatness has been achieved by being the precursor to pervasive computing systems of the future.

As a result, the controversy around RFID is in fact a discussion of the opportunities and threats presented by this new paradigm where everything and everyone is automatically identifiable by computers and every movement captured and processed. For the research community, RFID offers the unique ability to prototype and experience firsthand pervasive computing, which is often the true source of excitement about this technology, as it can be used to transform the way we experience computing and communications.

Outside the laboratory the situation is different. RFID tags embedded in travel documents and ID cards do not improve public safety but do open up threats to privacy; they also improve data capture accuracy but at a disproportionate cost. RFID is not economically viable for universal item-level tagging of consumer goods, and it will remain so for many years to come, especially when data management costs are taken into account. It can, however, create considerable improvements when used for container-level tagging in the supply chain, although the benefits are not shared equally between all partners. RFID for ticketing and asset tracking is simple and effective and the more mature application of this technology.

In any case, our current experiences with RFID point to two areas of particular concern: fair use of the technology and the potential environmental impact of its disposal. Regarding the former issue in particular, there are many open questions in terms of device and data ownership, the use of the technology to coerce and control, and when this technology should be used in ways that promote the common good but suppress individual rights. These questions are not rhetorical but already have had an impact, as we discussed in Chapter 9.

At a more basic level, security and privacy protection mechanisms implemented in the vast majority of commercially available tags are too simplistic and completely ineffective, if they exist at all. But that is not to say that RFID tags are impossible to secure, and in fact the research community has taken considerable strides toward this end. Nevertheless, there is solid evidence that the technology can be used to violate personal privacy in new and powerful ways, but addressing these issues will require more practical experience, research, and deliberations.

Another almost totally unexplored implication of the use of RFID, which never arises in the research lab, is disposal. Bits can easily disappear—in fact, one may say too easily—but atoms are much harder to ethically and

responsibly dispose of. RFID tags in particular cannot be recycled by the processes in common use today, and as a result they make all tagged packaging material non-recyclable. For example, wooden pallets used in the supply chain and glass containers are both 100% recyclable, but embedding or attaching a tag to either type of object completely prevents them from being recycled. With waste management becoming a necessity in many parts of the developed world, the potential of RFID to retrograde this capability at large scale is clearly of considerable concern.

Acronyms

AES	Advanced Encryption Standard
ALE	Application Level Events
ANSI	American National Standards Institute
AON	Application Oriented Networking
APDU	Application Data Unit
ASK	Amplitude shift keying
BAC	Basic Access Control
CASPIAN	Consumers Against Supermarket Privacy Invasion and Numbering
COTS	Commercial-off-the-shelf
CSN	Card Serial Number (ISO 14443A)
CW	Carrier wave
DDDS	Dynamic Delegation Discovery System
DES	Data Encryption Standard
DI	Data Identifier
DI	Device Infrastructure (IBM)
DNS	Domain Name System
DoD	Department of Defense
DoS	Denial of service
DRM	Dense reader mode
DST	Digital signature transponder
EAN	European Article Number
EAS	Electronic Article Surveillance
ECR	Efficient Consumer Response
EDI	Electronic Data Interchange
EDIFICE	European User Group for companies with interests in computing, electronics and telecommunications
EIRP	Equivalent isotropic radiated power
EM	Event Manager
EPC	Electronic Product Code

EPC DS	EPC Discovery Service
EPC IS	EPC Information Service
EPC RP	EPC Reader Protocol
ERP	Effective radiated power
FDA	Food and Drug Administration
FMCG	Fast moving consumer goods
FPGA	Field-programmable gate array
FSA	Fluidic Self Assembly
GDSN	Global Data Synchronization Network
Gen2	EPC Class 1 Generation 2 tag
GIAI	Global Individual Asset Identifier
GID	General Identifier
GLN	Global Location Number
GRAI	Global Returnable Asset Identifier
GTIN	Global Trade Identification Number
HF	High frequency
HTTP	Hypertext Transfer Protocol
IAC	Issuing Agency Code
ICAO	International Civil Aviation Organization
IDT	interdigital transducer
IP	Internet Protocol
ISBN	International Standard Book Number
ISO	International Organization for Standardization
J2EE	Java 2 Platform, Enterprise Edition
LBT	Listen before talk
LF	Low frequency
LLRP	Low Level Reader Protocol
MRTD	Machine Readable Travel Document
MRZ	Machine-readable zone
NAO	National Audit Office
NAPTR	Naming Authority Pointers
NDC	National drug code
NFC	Near Filed Communication
NNI	Netherlands Normalization Institute
NRZ	Non-return-to-zero
OCR	Optical character recognition
ODS	Object Description Service
ONS	Object Naming Service
OSGi	Open Services Gateway initiative
PC	Programme Control (Gen2)
PC	Protocol Control (Gen2)
PET	Privacy-enhancing technologies
PICA	Parallel Integrated Chip Assembly
PIE	Pulse interval encoding
PIN	Personal Identification Number

PS	Premises Server (IBM)
PSK	Phase shift keying
QSP	Query Slot protocol
RF	Radio frequency
RFC	Request for Comments (internet specification)
RFID	Radio Frequency Identification
RN16	16-bit random number (Gen2)
RSA	Rivest, Shamir and Adleman cryptography
SAW	Surface acoustic wave
SCM	Supply chain management
SEM	Sensor Edge Mobile (oracle)
SES	Sensor Edge Server (Oracle)
SGLN	Serialized Global Location Number
SGTIN	Serialized Global Trade Identification Number
SIP	Session Initiation Protocol
SKU	Stock-keeping unit
SNMP	Simple Network Management Protocol
SOAP	Simple Object Access Protocol
SoF	Supermarket of the Future
SSCC	Serial Shipping Container Code
TAP	Tag Acquisition Processor (Reva)
TCP	Transmission Control Protocol
UHF	Ultra-high frequency
UID	Universal Identifier
uID	Ubiquitous identifier
uPIS	ucode Product Information Service
URI	Uniform Resource Identifier
URL	Uniform Resource Locator
URN	Uniform Resource Name
USB	Universal Serial Bus
VMI	Vendor Managed Inventory
XML	Extensible Markup Language
XMLRPC	XML Remote Procedure Call

References

1. A. Acquisti and J. Grossklags. Privacy attitudes and privacy behavior: losses, gains, and hyperbolic discounting. In J. Camp and S. Lewis (eds), *The Economics of Information Security* (Kluwer Academic Publishers, 2004) 165–178.
2. A. Acquisti, Ubiquitous Computing, Customer Tracking, and Price Discrimination. In G. Roussos (ed), *Ubiquitous and Pervasive Commerce* (Springer, London, 2006) 115–132.
3. H. Ailisto, L. Pohjanheimo, P. Vlkkynen, E. Strmmer, T. Tuomisto and I. Korhonen, Bridging the physical and virtual worlds by local connectivity–based physical selection, *Pers. aUbiq. Computing* 10(6) (2006) 334–344.
4. K. Albrecht and L. McIntyre, *Spychips: How Major Corporations and Government Plan to Track Your Every Move with RFID*. Thomas Nelson Publishers, Nashville (2006).
5. R. J. Anderson and M. G. Kuhn, Low cost attacks on tamper resistant devices. In Proc. SPW97, Lecture Notes in Computer Science, Vol. 1361 (Springer, Berlin, 1997) 125–136.
6. G. Avoine, Bibliography on Security and Privacy in RFID Systems, online at `http://lasecwww.epfl.ch/ gavoine/rfid/`
7. R. Aylott and V.W. Mitchell, An exploratory study of grocery shopping stressors. *Int. J. Retail Distribution Management* 26(9) (1998) 362–373.
8. G. Avoine and P. Oechslin, RFID traceability: A multilayer problem. In A.S. Patrick and M. Yung (eds), Proc. Financial Cryptography, *Lecture Notes in Computer Science*, Vol. 3570, (Springer, Berlin, 2005) 125–140.
9. J. Baker, The role of the environment in marketing services: The consumer perspective. In J.A. Czepeil, C. Congram and J. Shanahan J. (eds), *The Services Challenge: Integrating for Competitive Advantage* (American Marketing Association, Chicago, 1986).
10. J. E. Bardram, Applications of context–aware computing in hospital work examples and design principles, in *Proc. ACM SAC 2004* (ACM Press, 2004) 1574–1579.
11. J. Bohn, V. Coroama, M. Langheinrich, F. Mattern and M. Rohs, Living in a world of smart everyday objects – Social, economic, and ethical implications. *J. Hum. Ecological Risk Ass* 10:5 (2004) 763–785.

12. S. Bono, M. Green, A. Stubblefield, A. Juels, A. Rubin, and M. Szydlo, Security analysis of a cryptographically-enabled RFID device, in *Proc. 14th USENIX Security Symposium* (USENIX Press, 2005) 1–16.
13. C. Bornhövd, T. Lin, S. Haller and J. Schaper, Integrating automatic data acquisition with business processes – experiences with SAP's Auto–ID infrastructure, in *Proc. VLDB04* (2004) 1182 - 1188.
14. S.A. Brown, *Revolution at the Checkout Counter: The Explosion of the Bar Code*. (Wertheim Publications in Industrial Relations, Harvard University Press, Cambridge, MA, 1997).
15. R. Caneel and P. Chen, *Enterprise Architecture for RFID and Sensor Based Services* .(Oracle Corporation, Redwood Shores, 2006).
16. D. Carluccio, K. Lemke, and C. Paar, Electromagnetic side channel analysis of a contactless smart card: First results. In Proc. Ecrypt Workshop on RFID and Lightweight Crypto (electronic proceedings, 2005).
17. D. Carluccio, K. Lemke-Rust, C. Paar, A.-R. Sadeghi, E-Passport: The global traceability or how to feel like an UPS package. In J.K. Lee, O. Yi, and M. Yung (eds) Proc. WISA 2007, *Lecture Notes in Computer Science*, Vol. 4298 (Springer, Berlin, 2007) 391–404.
18. D. Carter and I. Lomas. Store Of The Future. *Proc. ECR Europe* (Berlin, Germany, 2003).
19. J. Chamberlain, C. Blanchard, S. Burlingame, S. Chandramohan, E. Forestier, G. Griffith, M.L. Mazzara, S. Musti, S-I. Son, G. Stump and C. Weiss, *IBM WebSphere RFID Handbook: A Solution Guide* (IBM Redbooks, Raleigh, 2006).
20. J.J. Chen and C.Adams, Short-range wireless technologies with mobile payments systems. In M. Janssen, H.G. Sol, R.W. Wagenaar (eds) *Proc. ICEC '04* (ACM Press, 2004) 649–656.
21. H. Chen, P.B. Chou, S. Duri, J.G. Elliott, J.M. Reason and D.C. Wong, A model–driven approach to RFID application programming and infrastructure management. In F. Lau, H. Lei and X. Meng (eds) *Proc. ICEBE05* (IEEE Press, 2005) 256– 259.
22. J-P. Curty, M. Declercq, C. Dehollain and N. Joehl, *Design and Optimization of Passive UHF RFID Systems* (Springer, Berlin, 2006).
23. P. De, K. Basu and S. K. Das, An ubiquitous architectural framework and protocol for object tracking using RFID tags. In T. Finin, C. Ghidini, T. La Porta and C. Petrioli (eds) *Proc. MobiQuitous* (ACM Press, 2004) 174–182.
24. A.K. Dey, G.D. Abowd and D. Salber, A Context-based infrastructure for smart environments, in *Proc. MANSE99* (1999) 114–128.
25. R.J. Donovan and J.R. Rossiter. Store atmosphere: An Environmental Psychology Approach. *Journal of Retailing* 58:34 (1982) 52.
26. H. Dunne, Message Development, Auto–ID Sponsor briefing (June 2002).
27. D.W. Engels, J. Foley, J. Waldrop, S. Sarma and D. Brock, The networked physical world: an automated identificationarchitecture, in *Proc. WIAPP 2001* (IEEE Press, 2001) 76–77.
28. A. E. Fano, Mobile valet: enabling collaboration between remote services, mobile users, and task location, *in Proc. 2001 AAAI Fall Symp. Int. Inf.* (2001).
29. A. E. Fano and A. Gershman, The Future of business services, *Comm. ACM* 35:12 (2002) 83–87.

30. M. Feldhofer, S. Dominikus, J. Wolkerstorfer, Strong authentication for RFID systems using the AES algorithm. In *Proc. CHES 2004*, Lecture Notes in Computer Science, Vol. 3156 (Springer, Berlin, 2004) 357–370.
31. K. Finkenzeller, *RFID Handbook: Fundamentals and Applications in Contactless Smart Cards and Identification* (John Wiley & Sons, London, 2003).
32. L. A. Fish and W. C. Forrest, The 7 Success Factors of RFID, *Supply Chain Man. Rev.* 5 (2006) 26–32.
33. K.P. Fishkin, M. Philipose and A. Rea, Hands-on RFID: Wireless wearables for detecting use of objects. In *Proc. ISWC'05* (IEEE Press, 2005) 38–43.
34. K.P. Fishkin, S. Roy and B. Jiang, Some methods for privacy in RFID communication. In *Proc. WSASN04*, Lecture Notes in Computer Science, Vol. 3313 (2004) 42–53.
35. M. Fleck, M. Frid, T. Kindberg, E. O'Brian-Strain, R. Rajani and M. Spasojevic, From informing to remembering: deploying an ubiquitous system in an interactive science museum, *IEEE Perv. Comp.* 1:2 (2002) 13–21.
36. C. Floerkemeier, M. Lampe and T. Schoch, The smart box Concept for ubiquitous computing environments. In *Proc. SOC*, Grenoble, France (electronic proceedings, 2003).
37. C. Floerkemeier, R. Schneider and M. Langheinrich, Scanning with a Purpose – Supporting the fair information principles in RFID protocols. In *Proc. UCS 2004*, Lecture Notes in Computer Science 3598 (Springer, Berlin, 2005) 214–231.
38. C. Floerkemeier and M. Lampe, RFID middleware design – addressing application requirements and RFID constraints. In *Proc. SOC–EUSAI* (ACM Int. Conf. Proc. Ser. 121, 2005) 219–224.
39. K.R. Foster and J. Jaeger, RFID inside: The murky ethics of implanted chips. *IEEE Spectrum*, 44:3 (2007) 24–29.
40. M. Fry and A. Ghosh, Application level active networking, *Comp. Net.* 31 (1999) 655–667.
41. K. Fujinami, F. Kawsar and T. Nakajima, AwareMirror: a personalized display using a mirror. In *Proc. Pervasive2005*, Lecture Notes in Computer Science 3468 (Springer, Berlin, 2005) 315–332.
42. F. Fukuyama. *Trust: Human Nature and the Reconstitution of Social Order* (Free Press, New York) 1996.
43. S. Garfinkel, RFID in Ubiquitious Commerce. In G. Roussos (ed) *Ubiquitous and Pervasive Commerce* (Springer, London, 2005).
44. S.L. Garfinkel, A. Juels, and R. Pappu, RFID privacy: an overview of problems and proposed solutions. *IEEE Security and Privacy* 3:3 (2005) 34–43.
45. S. Garfinkel and B. Rosenberg, *RFID: Applications, Security, and Privacy* (Addison–Wesley, 2005).
46. A. Gershman and A. E. Fano, Customer service with eyes. In *Proc. Work. Ubiq. Communications* (electronic proceedings, 2003).
47. R. Ghani and A. E. Fano, Building recommender systems using a knowledge base of product semantics. In *Proc. Work. Recomm. Personalisation E-Commerce* (2002).

48. P. Golle, M. Jakobsson, A. Juels, and P. Syverson, Universal re–encryption for mixnets. In *Proc. CT–RSA*, Lecture Notes in Computer Science 2964 (Springer, berlin, 2004) 163–178.
49. L. Grunwald, RF-ID and smart labels: myth, technology and attacks, Black Hat Breifiengs, Las Vegas, USA (22-24 July 2004).
50. O. Günther and S. Spiekermann, RFID and the perception of control: the consumer's view. *Comm. ACM* 48:9 (2005) 73–76.
51. D. Haehnel, W. Burgard, D. Fox, K.P. Fishkin and M. Philipose, Mapping and localization with RFID technology. In *Proc. ICRA 2004* (IEEE Press, 2004) 1015–1020.
52. J. Halamka, A. Juels, A. Stubblefield and J. Westhues, The security implications of VeriChip cloning, *J. Am. Med. Inform. Assoc.* 13 (2006) 601–607.
53. G.P. Hancke and M.G. Kuhn, An RFID distance bounding protocol. In *Proc. SECURECOMM 2005* (IEEE Press, 2005) 67–73.
54. G. P. Hancke, A practical relay attack on ISO 14443 proximity cards. In *Proc. ISSP 2005* (IEEE Press, 2005) 328–333.
55. J.M. Hellerstein, W. Hong and S.R. Madden, The sensor spectrum: technology, trends, and requirements. *ACM SIGMOD Rec.* 32:4 (2003) 22–27.
56. T.S. Heydt-Benjamin, H-J. Chae, B. Defend and Kevin Fu, Privacy for public transportation. In *Proc. PET06* (electronic proceedings, 2006).
57. S. Inoue and H. Yasuura, RFID privacy using user–controllable uniqueness. In *Proc. RFID Privacy Work.* (2003).
58. J.E. Hoag and C.W. Thompson, Architecting RFID middleware, *IEEE Int. Comp.* 10:5 (2006) 88–92.
59. R.J.K. Jacob, H. Ishii, G. Pangaro and J. Patten, A Tangible Interface for organizing information using a Grid. In *Proc. CHI 2002* (ACM Press, 2001) 339–346.
60. S.R. Jeffery, M. Garofalakis and M.J. Franklin, Adaptive cleaning for RFID data streams. In *Proc. VLDB05* (2005) 163–174.
61. A. Juels, R. L. Rivest and M. Szydlo, The blocker tag: Selective blocking of RFID tags for consumer privacy. In *Proc. ACM CCS03* (ACM Press, 2003) 103–111.
62. A. Juels, RIFD Security and privacy: a research survey. *IEEE J. Sel. Areas Comm.* 24:2 (2006) 381–394.
63. A. Juels, Minimalist cryptography for low–cost RFID tags. In *Proc. SCN 2004*, Lecture Notes in Computer Science 3352 (2004) 149–164.
64. A. Juels, P. Syverson, and D. Bailey, High–power proxies for enhancing RFID privacy and utility. In *Proc. Work. Priv. Enh. Technologies* (electronic proceedings, 2005).
65. G. Karjoth and P. Moskowitz, Disabling RFID tags with visible confirmation: Clipped tags are silenced. In *Proc. Work. Priv. Elec. Soc.* (ACM Press, 2005) 27–30.
66. E. Katsiri, J. Bacon, and A. Mycroft, Linking sensor data to context-aware applications using abstract events, *J. Perv. Comp. Syst.* (2007).
67. Y. Kim, M. Moon and K. Yeom, A framework for rapid development of RFID applications. In *Proc. ICCSA 2006*, Lecture Notes in Computer Science 3983 (2006) 226-235.

68. S. Konomi, S. Inoue, T. Kobayashi, M. Tsuchida and M. Kitsuregawa, Supporting colocated interactions using RFID and social network displays. *IEEE Perv. Comp.* 5:3 (2006) 48–56.
69. S. Konomi and G. Roussos, Ubiquitous computing in the real world: lessons learnt from large scale RFID deployments. *Pers. Ubiq. Computing* (2007), to appear.
70. P. Kourouthanassis and G. Roussos, Developing consumer-friendly pervasive retail systems. *IEEE Perv. Comp.* 2:2 (2003) 32–39.
71. M. Kritzler, L. Lewejohann, A. Krger, M. Raubal and N. Sachser, An RFID–based tracking system for laboratory mice in a semi–natural environment. In *Proc. PTA 2006* (electronic proceedings 2006).
72. M. Lampe, M. Strassner and E. Fleisch, A Ubiquitous Computing Environment for Aircraft Maintenance, in Proc. ACM SAC04 (ACM Press 2004) 1586–1592.
73. J. Landt, The history of RFID, IEEE Potentials 24:4 (2005) 8–11.
74. J.S. Lee and H.J. Kim, RFID code structure and tag data structure for mobile RFID services in Korea, in Proc. ICACT 2006, Vol. 2 (2006) 3–6.
75. C. Legner and F. Thiesse, RFID-based maintenance at Frankfurt airport. *IEEE Perv. Comp.* 5:1 (2006) 34– 39.
76. M. Levy and B.A. Weitz. *Retailing Management* (McGraw-Hill, New York, 2003).
77. D.M. Lewison. *Retailing* (Prentice Hall, 1997).
78. M. Mamei, F. Zambonelli, Pervasive pheromone-based interactions with RFID tags. *ACM Trans. Aut. Adap. Systems* (2007) to appear.
79. A. Marianantoni, H. Park, J. Friedman, V. Holtgrewe, J. Burke, M. Srivastava, F. Wagmister, W. McDonald and J. Brush, Sensor networks for media production. In *Proc. EmNets 2004* (ACM Press, 2004) 325–325.
80. A.J. Martin. *Infopartnering: The Ultimate Strategy for Achieving Efficient Consumer Response* (John Wiley & Sons, London, 1995).
81. J.F. McCarthy, D.W. McDonald, S. Soroczak, D.H. Nguyen and Al M. Rashid, Augmenting the social space of an academic conference. In *Proc. CSCW 04* (ACM Press, 2004) 39–48.
82. M. Mealling, *Dynamic Delegation Discovery System (DDDS) Part Three: The Domain Name System (DNS) Database* (IETF 2002).
83. C.S. Nam and S. Konomi, Usability evaluation of QueryLens: implications for context-aware information sharing using RFID. In *Proc. IASTED-HCI 2005* (Acta Press, 2005) 90–95.
84. W. Nosovic and T.D. Todd, Scheduled rendezvous and RFID wakeup in embedded wireless networks. In *Proc. ICC* (IEEE Press, 2002) 3325–3329.
85. M. Ohkubo, K. Suzuki and S. Kinoshita, Cryptographic approach to "privacy-friendly" tags. In *Proc. RFID Priv. Work.* (electronic proceedings, 2003).
86. K. Opasjumruskit, T. Thanthipwan, O. Sathusen, P. Sirinamarattana, P. Gadmanee, E. Pootarapan, N. Wongkomet, A. Thanachayanont and M. Thamsirianunt, Self-powered wireless temperature sensors exploit RFID technology. *IEEE Perv. Comp.* 5:1 (2006) 54–61.
87. S. Ortiz Jr., Is Near-Field Communication close to success? *IEEE Computer* 39:3 (2006) 18–20.

88. L. Pelusi, A. Passarella and Marco Conti, Opportunistic Networking: data forwarding in disconnected mobile ad hoc networks, IEEE Comm. Mag. 44:11 (2006) 134–141.
89. T. Pering, Y. Anokwa and R. Want, Body movements: gesture connect facilitating tangible interaction with a flick of the wrist. In *Proc. TEI07* (ACM Press, 2007) 259–262.
90. Pew Research Center. *Pew Internet and American Life Project Trust and Privacy Online: Why Americans Want to Rewrite the Rules.*
91. B. S. Prabhu, X. Su, H. Ramamurthy, C-C. Chu and R. Gadh, WinRFID A middleware for the enablement of radio frequency identification (RFID) based Applications. In R. Shorey, C.M. Choon, O.W. Tsang and A. Ananda (eds) *Mobile, Wireless and Sensor Networks: Technology, Applications and Future Directions* (Wiley-IEEE Press, 2006).
92. P.O. Rashid, P. Coulton and R. Edwards, Mobile: The mobile phone as a digital SprayCan. In *Proc. ACM SIGCHI ACE06* (ACM Press, 2006).
93. R. Reid and S. Brown, I hate shopping! An introspective perspective. *Int. J. Retail and Distribution Management* 24(4) (1996) 4–16.
94. M. Rieback, B. Crispo and A. Tanenbaum, Is your cat infected with a computer virus? In *Proc. PERCOM 2006* (IEEE CS Press, 2006) 169–179.
95. J. Riekki, T. Salminen and I. Alakarppa, Requesting pervasive service by touching RFID tags. *IEEE Perv. Comp.* 5:1 (2006) 40–46.
96. K. Römer, T. Schoch, F. Mattern and T. Dübendorfer, Smart identification frameworks for ubiquitous computing applications. *Wireless Networks* 10:6 (2004) 689–700.
97. G. Roussos, P. Kourouthanassis and T. Moussouri, Appliance design for mobile commerce and retailteinment. *Per. Ubiq. Computing* 7:3-4 (2003) 203–209.
98. G. Roussos and T. Moussouri, Consumer perceptions of privacy, security and trust in ubiquitous commerce. *Pers. Ubiq. Computing* 8:6 (2004) 416–429.
99. G. Roussos, D. Spinellis, P. Kourouthanassis, E. Gryazin, P. Pryzbliski, G. Kalpogiannis and G. Giaglis, Systems architecture for pervasive retail. In *Proc. ACM SAC 2003* (ACM Press, 2003) 631–636.
100. G. Roussos, *Ubiquitous and Pervasive Commerce: New Frontiers for Electronic Business* (Springer, London, 2005).
101. G. Roussos, Enabling RFID in Retail. *IEEE Computer* 39:3 (2006) 25–30.
102. T. Salvador, G. Bell and K. Anderson, Design ethnography. *Design Management Journal* 10(4) (1999) 35–41.
103. S. E. Sarma, S. A. Weis, and D. W. Engels, Radio-Frequency identification: security risks and challenges. *Cryptobytes* 6:1 (2003) 2–9.
104. E.W. Schuster, S.J. Allen, D.L. Brock, Global RFID: The value of the EPCglobal network for supply chain management (Springer, London, 2007).
105. J. Scott, P. Hui, J. Crowcroft and C. Diot, Haggle: A networking architecture designed around mobile users. In *Proc. IFIP WONS 2006* (Springer, Berlin, 2006).
106. T. Selker and W. Burleson, Context-aware design and interaction in computer systems. *IBM Sys. J.* 39:3/4 (2000) 880–891

107. K. Shindo, N. Koshizuka, and K. Sakamura, Large-scale ubiquitous information system for digital museum. In *Proc. IASTED03* (2003).
108. J. Smaros and J. Holmstrom, Reaching the consumer through e-grocery VMI. *Int. J. Retail Distr. Management* 28:2 (2000) 55–61.
109. H. Stockman, Communication by Means of Reflected Power. *Proc. IRE* 35 (1948) 1196–1204.
110. M. Strassner and T. Schoch, Today's Impact of Ubiquitous Computing on Business Processes. In *Proc. PERVASIVE02 Short Paper Proceedings* (2002) 62–74.
111. C. Tan, B. Sheng, and Q. Li, Serverless search and authentication protocols for RFID. In *Proc. IEEE PERCOM07* (IEEE Press, 2007) 3–12.
112. F. Thiesse, E. Fleisch and M. Dierkes, LotTrack: RFID–based process control in the semiconductor industry. *IEEE Perv. Comp.* 5:1 (2006) 47–53.
113. D.R. Thompson, N. Chaudhry, and C.W. Thompson, RFID security threat model. In *Proc. ALAR05* (electronic proceedings, 2006).
114. H. Vogt, Efficient object identification with passive RFID tags. In *Proc. Pervasive 2002*, Lecture Notes Computer Science 2414 (2002) 98–113.
115. J. Waldo, Virtual organizations, pervasive computing, and an infrastructure for networking at the edge. *Inf. Sys. Frontiers* 4:1 (2002) 918.
116. D. Wan, Magic medicine cabinet: A situated portal for healthcare. In *Proc. UBICOM 99* (1999).
117. D. Wan, Magic wardrobe: Situated shopping from your own bedroom. In Proc. *UBICOM 2000* (2000).
118. F. Wang and P. Liu, Temporal management of RFID data. In *Proc. VLDB05* (2005) 1128–1139.
119. R. Want, K.O. Fishkin, A. Gujar and B.L. Harrison, Bridging physical and virtual worlds with electronic tags. In *Proc. CHI99* (ACM Press, 1999).
120. R. Want, An introduction to RFID technology. *IEEE Perv. Comp.* 5:1 (2006) 25–33.
121. T. Watanabe, S. Inoue, H. Yasuura, J. Sasaki, Y. Aoki, K. Akimoto, An RFID-based multi-service system for supporting conference events. In *Proc. AMT 05* (IEEE Press, 2005) 435–439.
122. M. Weiser, The computer for the 21st century. *ACM SIGMOBILE Mob. Comp. Comm. Review* 3:3 (1999) 3–11.
123. J. Westhues, Hacking the prox card. In S. Garfinkel and B. Rosenberg (eds) *RFID: Perspectives, Policy, and Practice* (Addison–Wesley, 2005).
124. S. Willis and S. Helal, RFID information Grid for blind navigation and wayfinding. In *Proc. ISWC05* (IEEE Press, 2005) 34–37.
125. E. Wu, Y. Diao and S. Rizvi, High-performance complex event processing over streams. In *Proc. 2006 ACM SIGMOD Int. Conf. Man. Data* (2006) 407–418.
126. Y.Z. Zhao and O.P. Gan, Distributed design of RFID network for large-scale RFID deployment. In *Proc. ICII06* (IEEE Press, 2006) 44–49.

Index