RFID in Pervasive Computing:
State-of-the-art and Outlook

George Roussos

School of Computer Science and Information Systems
Birkbeck College, University of London
Malet Street, London WC1E 7HX, UK

Vassilis Kostakos

Department of Computer Science
University of Bath
Bath BA2 7AY, UK

Abstract

RFID has already found its way into a variety of large scale applications and arguably it is already one of the most successful technologies in the history of computing. Beyond doubt, RFID is an effective automatic identification technology for a variety of objects including natural, manufactured and handmade artifacts; humans and other species; locations; and increasingly media content and mobile services. In this survey we consider developments towards establishing RFID as the cost-effective technical solution for the development of open, shared, universal pervasive computing infrastructures and look ahead to its future. In particular, we discuss the ingredients of current large scale applications; the role of network services to provide complete systems; privacy and security implications; and how RFID is helping prototype emerging pervasive computing applications. We conclude by identifying common trends in the new applications of RFID and ask questions related to sustainable universal deployment of this technology.

Key words: radio frequency identification, pervasive computing

PACS:

Email addresses: g.roussos@bbk.ac.uk (George Roussos), vk@cs.bath.ac.uk (Vassilis Kostakos).
URLS: www.dcs.bbk.ac.uk/ gr/ (George Roussos), www.cs.bath.ac.uk/ vk/ (Vassilis Kostakos).

Preprint submitted to Elsevier 16 April 2007
“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.”

Harry Stockman in “Communication by Means of Reflected Power” (1948)

1 Introduction

At first glance RFID and its application to pervasive computing appear to be straightforward and simple to implement. And yet the opposite is true: RFID is a technology that involves 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. It is not surprising then that following Stockman’s statement in 1948 [97], RFID required almost 60 years of development to find its way to large scale applications [66]. Even so, the fact that so many engineering disciplines are involved also means that a full treatment of every aspect of RFID cannot be contained in a single report.

That being the case, this survey is focused on developments related to engineering open, shared, scalable, universal infrastructures for the provision of wide-reaching pervasive services and applications. Such applications are able to cope with a variety of RFID-tagged objects whether natural, manufactured or handmade, irrespective of their origin. Applications can also use RFID to identify humans and other living species, and also locations, media content and data services. Pervasive computing environments constructed upon such infrastructures have high tag density with dozens or hundreds of tags co-located and guaranteed to inter-operate seamlessly. Such infrastructures also provide well defined points of access which aggregate lower level operational details into abstractions at an appropriate level suitable for building applications. Still, if RFID is to enable pervasive computing in the real world then it has to provide infrastructures that support innovation at the application layer in a manner similar to the internet. Indeed, internet protocols and engineering drafts have allowed a common planetary-scale network infrastructure to be deployed that has supported the development of novel applications of global reach including email, the web and several peer-to-peer overlays. As a result, this vision for RFID has been dubbed – albeit inaccurately – the Internet of Things [23]. This survey reports on progress towards this goal.

In conducting this survey we also have two additional objectives. First, we wish to introduce the pervasive computing community to the work carried out by the RFID commercial sector, which has been driving the development of the technology for the past decade and has established the status quo. There
are many valuable lessons discovered during this process that has established a de-facto operational context for RFID, which appears not be adequately disseminated within the academic community. Our second objective is to provide links between research in RFID systems and related investigations in other areas of pervasive computing with the view to highlight its relevance to the wider research community and the many possible common threads of investigation. This survey hopes to offer the baseline for setting and exploring research questions in this domain.

The remainder of the paper is structured as follows: in the following section we discuss selected well-established large-scale industrial applications of RFID that have driven its rise in popularity in recent years, and identify their main characteristics and requirements. In Section 3, we briefly review the basics of RFID technology and identify some of the tradeoffs in selecting a particular system. Section 4 deals with software abstractions for RFID middleware and related networked services and architectures. Following this, we review the critical issue of privacy protection and security. We conclude with a discussion of novel ways of employing RFID to enable pervasive computing and identify some of the challenges involved in developing new applications and functionalities.

2 Large-scale Commercial Applications of RFID

To be sure, recent years have observed an explosion of interest in RFID. The reason for this is twofold: first, the availability of very low cost passive RFID tags that require no battery to operate, and second, the wider availability of robust fixed and mobile communication networks that allow connectivity between RFID installations in diverse locations and centralized supporting services. These conditions have led to large scale commercial applications in the supply chain [89,92], ticketing [62], asset tracking [100], maintenance [68] retail [85] and identification [39]. Due to these applications, RFID has become one of the most popular computing platforms in use today: IDTechEx, a market research firm specializing in RFID, estimates that 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. Exceedingly, such applications are not the privilege of countries commonly thought as technologically advanced but it is equally possible they be encountered in the Middle-East, South America and the less developed regions of Asia. This popularity of RFID permits considerable cost reductions that have sparked further interest in RFID for research as it offers unique opportunities over alternative wireless sensor networking and communication technologies in instrumenting the physical environment.

Yet, the same features that make RFID such a popular technology are also
complicating its use. To offer 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 object 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.

In this remainder of this section we discuss three applications of particular significance namely, ticketing, e-passports and the supply chain. In all cases, RFID has been deployed in large-scale systems and is supported by networked services of varying complexity. These applications have quite different characteristics and requirements which are summarized in Table 1.

### 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 (MRTD) that use standard RFID technology to store personal and biometric information on passports, visas and travel cards. This development is seen as an improvement over manual processes in terms of efficiency and data entry precision at border control points but more critically, in increasing the security of air travel. At the time of writing, more than a hundred countries have implemented or are in the process of implementing the system.

The ICAO provisions call for ISO 14443-compliant RFID tags (discussed in Section 3) embedded in travel documents to hold personal and biometric information of the traveler. RFID readers operated by immigration services interrogate and retrieve traveler information without the need for manual inter-

<table>
<thead>
<tr>
<th></th>
<th>Ticketing</th>
<th>e-Passport</th>
<th>Supply Chain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tag Density</td>
<td>Low</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>Tag Range</td>
<td>Short</td>
<td>Short</td>
<td>Long</td>
</tr>
<tr>
<td>Tag Lifetime</td>
<td>Medium</td>
<td>High</td>
<td>Medium</td>
</tr>
<tr>
<td>Tag Complexity</td>
<td>High</td>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td>Tag Security</td>
<td>Strong crypto</td>
<td>Crypto (known key)</td>
<td>Password (32-bit)</td>
</tr>
<tr>
<td>Network support</td>
<td>Delay Tolerant</td>
<td>Local</td>
<td>High - real time</td>
</tr>
</tbody>
</table>

Table 1
Characteristics of three large scale industrial RFID applications.
vention. Due to the current capacity of tags, biometric data are restricted to photographs but the standard also provides specifications for iris 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.

While the MRTD system certainly facilitates faster and more accurate data entry, it has justifiably received strong criticism regarding its privacy and security performance and its unsubstantiated claims for improved safety of international travel. Counterfeiting the RFID component of an MRTD is a fairly straightforward process which can be achieved at relatively low cost. Moreover, access to personal data stored in the chip is facilitated rather than hindered compared against paper only travel documents since in some cases this is possible without the need to gain physical access to the document itself. Even in cases where additional external shielding is used to protect 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.

E-passport applications have simple requirements both in terms of RFID tag access and supporting network services. Indeed, at any one time only a single tag is within the range of the reader and there is ample opportunity 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, during 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 restricted interoperability requirements since they operate within national scope and are commonly supported by proprietary database applications over secured private networks.
2.2 Ticketing

One of the earliest and most successful large scale applications of RFID has been in metropolitan public transport ticketing. Today such systems are deployed in numerous cities across the globe, in some cases being operational for nearly a decade. In ticketing applications, RFID tags are often embedded in credit-card sized reusable tickets which store either a seasonal pass or credit that can be used against travel. Such tickets hold a number which identifies uniquely either the ticket or the passenger and in some cases also holds personal information about the holder and a record of the most recent trips. Tickets are validated and updated in real-time at gates fitted with readers during entry and exit to the transport system. One of the largest RFID-based ticketing systems is the Oyster card [62] in London, UK, which supports more than 10 million active passengers and has deployed over 27 thousand readers.

A critical operational parameter of RFID ticketing related to control of passenger flow, is the time required to validate and update the information stored in the tag and operate the gates. Of course, this task must also be carried out in such a way as to guarantee correct and provable billing with ticket prices often calculated on the fly based on the type of the ticket and the trip details. In almost all cases, RFID is far superior to alternative magnetic stripe technologies and has allowed considerable improvements in the performance of transport systems. An equally critical feature of such systems is security in that they must ensure fraud prevention. In the case of the Oyster system for example, communication between tag and reader is protected by a proprietary symmetric cryptography protocol, a situation which although not ideal, it nevertheless still provides an acceptable tradeoff between cost and risk.

The operation of Oyster is supported by an extensive multi-tiered back-end information system that processes payment, 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 7 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, it 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.

Similar to the majority of RFID ticketing, Oyster is based on the ISO 14443 standard that provides specific facilities for transport applications. The relatively short range of this RFID technology, often referred to as near-field
communication (cf. Section 3), is used to the advantage of the 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 the passenger.
Moreover, data throughput requirements are very low compared with the sup-
ported read performance, and the timing overhead as perceived by commuters
is primarily associated with cross-checking and recording the transaction with
the on-line system at the station.

2.3 Supply Chain Management

Supply chain management deals with the movement of goods between orga-
nizations 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 pro-
cessed from supplying fresh food from the farm to the supermarket shelf, to
delivering uniforms from the manufacturer to the soldier in the desert. Nev-
evertheless, all supply chains share a common goal that is 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 in-
formation at frequent intervals and that supply chain costs be unequivocally
identifiable at all times. An essential element to 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-worked information services that can be used to track and trace goods
and services [12].

Similar to bar codes, RFID can be used to store globally unique product iden-
tifiers which moreover can provide item-level rather than class-level identifica-
tion granularity. RFID tags can also provide the means for automatic capture
and processing of this information by recording the movement of tagged prod-
ucts from site to site through significant control points, for example warehouse
portals [89, Chapter 2]. The utility of RFID in this role was highlighted in the
early-90s in work carried out by the US Department of Defense (DoD). This
feasibility study was initiated 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 became 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: Wal-Mart in the US, Tesco in the UK and Metro in Germany are all actively implementing RFID to track products across their supply chains. These deployments focus primarily on tracking product containers – commonly pallets and boxes – rather than individual product items across open supply chains. A typical example of such applications would be a dock door at a manufacturing or warehouse facility which is equipped with RFID readers that record the codes embedded in pallets loaded with incoming and outgoing products, and automatically update the company’s resource planning systems. Subsequently, RFIDs could be used for taking quick inventories using handheld devices [28]. A central role in highlighting the potential benefits of RFID in this context has been played by the Auto-ID Center established with the support of several manufacturers, suppliers and retailers of consumer goods [23].

Out of the work conducted at the Auto-ID Center also came one of the major candidates for the development of such automated open supply-chain systems. The so-called Electronic Product Code has been formalized by EPCglobal and provides an unambiguous numbering scheme to identify goods containers, services, companies, locations, and assets worldwide [23,89]. Moreover, EPC codes can be stored in RFID tags and transmitted wirelessly to enable their electronic reading wherever required by business processes. In this case, interest in RFID is as a replacement for barcodes over which it has considerable advantages primarily due to the fact that it does not suffer from line of sight limitations. EPCglobal operates under the remit of GS1 (previously know as the EAN/UCC system) that has also developed standard data structures characteristic for different supply chain applications. GS1 also provides a set of service specifications that envision a global scale network, overlaid on top of the Internet, which offers directory, information and global repository services that link any EPC code to all information available about the product from its manufacturer and subsequent custodians or owners. As a result, EPC-based supply chains represent good examples of open systems operating over standardized and federated infrastructures. An interesting challenge to EPC (further considered in Section 4) is the fact that during their lifetimes collections of tags that are initially assigned to individual entities, must be grouped into larger composite objects that must be treated as one at later stages of the supply chain for example, individual components of a particular automobile that become a single car. Such relationships are time dependent and cannot be recorded at the tag level.
Unlike ticketing and e-passport applications, supply chain management is a far more complex and challenging environment for computing. Clearly, one requirement is to read large numbers of tags in very short time periods as products move through warehouse portals and other supply chain control points. This task is made even more complex 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 that due to the characteristics of the environment of operation readers may need to be deployed in dense constellations so that their reading ranges overlap and complicate communication with individual tags. Unreliable reads and writes cause cascading effects and thus additional smoothing of the data has to be performed to identify significant events at the application (rather than the observation) layer.

Finally, despite the fact that early applications focused on container level tagging, it is becoming increasingly feasible to extend RFID to item-level tagging with several retailers currently implementing the technology in commercial applications [62]. This fact can potentially have the greatest impact on pervasive computing 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 (more on this in Section 6). Seen in this light, supply chain initiated activities may offer the greatest potential to yield practical facilities that will bring about the Internet of Things and hence the universal open infrastructures required for large scale pervasive computing outside the research laboratory.

3 RFID Operating Principle and Core Elements

RFID has several peculiarities compared to other common wireless communication technologies. In this section, we discuss the main elements of the identification process, describe their roles, and outline factors that affect their performance. Rather than detailing the lower level and physical layer protocols, which are covered elsewhere at different levels of detail (cf. [39, Chapter 2] for a layman’s description, [108] for a high-level description aimed at an engineering audience, and [27] for full details), we focus on the aspects of RFID that affect application performance and would be the primary concern of large-scale pervasive computing system development.

3.1 Operating Principle

Unlike other common types of wireless systems, RFID communication is asymmetric in that one peer takes on the role of the transmitter and the other the
role of the responder. To a certain extend this is the main reason for the success of RFID: rather than creating its own transmission the responder (called the tag) is instead modulating or reflecting the electromagnetic waves emitted by the transmitter (called the reader or interrogator) to communicate. This technique allows for a somewhat complex reader to be used with a very simple tag of small size, which can be built at low cost. A relatively small number of fixed or mobile readers can be used in deployments with very high numbers of tags. Moreover, in many cases the electromagnetic waves emitted by the reader carry enough energy to be harvested by the tag (using the coupling effect induced on the tag antenna by the electromagnetic carrier wave) and to be used as its source of power.

These two core ideas behind RFID, namely communication by reflection and remote activation using radio frequency, were first discovered in the 40s and the 60s respectively. But it was not until the mid-70s that fully passive relatively long range systems became possible (for a more detailed discussion of the history of RFID see [66]) although early tags were still limited by the non-availability of high capacity, high-performance chips. At that time, RFID could only provide up to a dozen read-only bits on massive die sizes which occupied most of the tag volume. Shrinking electronics, especially in the 90s, have been critical in the development of the current generation of tags which are both significantly more power efficient and provide higher storage and computational capability - both as a result of miniaturization.

One particular type of RFID, the so-called active RFID tags, carries batteries so they are not wholly dependent on the reader to provide energy. Such tags have considerable advantages against passive tags that draw all their power from the reader signal, as they transmit at higher power levels and thus have longer range and support more reliable communication. Moreover, active tags can operate in particularly challenging environments for example around water, it is easy to extend them with additional sensing capability for example temperature sensors, and they can initiate transmissions, but they stop operating when their battery expires. Despite their advantages, 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 a unique advantage of passive RFID in the context of pervasive computing especially from a system maintenance perspective. Active RFID on the other hand is just one of an increasing number of wireless local area communication technologies and as such it is of limited interest to this survey. In the remainder of this paper we will not consider active RFID further, although many of the discussions herein apply, and we will refer to passive RFID simply as RFID, without further qualification.

RFID tags naturally split into two main categories: those that use the magnetic component generating the near field of the radio wave, against those that use the electric component, which generates the far field. Near-field 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 slightly shifted frequency from that of the reader (what is often referred to as the sub-carrier frequencies).

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. 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 distance proportionate 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, in practice ISO 14443 systems would consistently work at a maximum range of approximately 30 cm using medium size antennas on the reader (radius approximately 20 cm) and credit-card size tags.

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 it 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 loosing rapidly 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 of the card and the reader (but not of the electronics directly).

RFID systems using the far field of the carrier wave operate using a technique called backscatter rather than load modulation. This process is very similar to the operation of the 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, that is the signature of the component of the wave that has been sent back to the reader, and comparing it to 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 very considerable loss of power during the reflection process and readers have to be sensitive to
<table>
<thead>
<tr>
<th></th>
<th>HF (near field)</th>
<th>UHF (far field)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency</td>
<td>≈ 13 Mhz</td>
<td>≈ 900 Mhz</td>
</tr>
<tr>
<td>Spectrum Allocation</td>
<td>Uniform</td>
<td>Fragmented</td>
</tr>
<tr>
<td>Cost (per tag)</td>
<td>&lt; 15 cents</td>
<td>&lt; 15 cents</td>
</tr>
<tr>
<td>Range</td>
<td>&lt; 30cm (1m max)</td>
<td>&lt; 4m (10m max)</td>
</tr>
<tr>
<td>External Interference</td>
<td>No</td>
<td>Cellular phones</td>
</tr>
<tr>
<td>Memory capacity</td>
<td>4Kbits</td>
<td>256bits</td>
</tr>
</tbody>
</table>

Table 2
Comparison of HF versus UHF RFID technologies.

less than a microwatt in most cases.

Because of the involvement of the far field, tag and reader antennas are dipoles. 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 865-956 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 the US). 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 US and 950-956 MHz in Japan). Far-field systems allow for longer range communication and it is common to achieve between 3 and 4 meters using approximately 30cm antennas and 10 cm tags. Using larger antennas 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 [19].

3.1.1 Readers

An RFID reader or interrogator consists of three main 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.
RFID 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 readers would receive a scanning plan from a driving application or other middleware, which they would implement by issuing state transition instructions to the tags within their range. The latter step usually has three stages: broadcasting to all tags within range and receiving responses, selecting a particular tag as the peer for communication, and exchanging information with the selected tag. This process can be quite complex especially in the case where a large number of tags are within range or when two or more readers overlap. In such cases, additional collision avoidance techniques must be implemented to ensure that communication is organized in a structured way so as to allow all tags to participate in this process [27].

3.1.2 Tags

The tag is a far simpler device and consists of:

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

A typical example of a modern tag is the EPC Class 1 Gen 2 [19, Chapter 4] which operates at UHF frequencies. The chip has a relatively complex non-volatile memory structure divided in four distinct areas (cf. Figure 1). 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 that when presented permanently disables the tag. The EPC memory bank contains the Electronic Product Code assigned to the object, location or other 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 which identifies the tag itself. The user bank is optional and can be used freely by applications.

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 pseudorandom 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 Section 5. For this reason most recent tag protocols implement a randomization process whereby tags use a pseudo-random number each time they are interrogated by a reader so as to avoid easy tracing.

3.1.3 Identifiers

In the vast majority of applications and certainly in those with most interest to pervasive computing, RFID tags would be used to identify objects, persons, locations, services or other content rather than the tags themselves. If the scope of the auto-identification system used by this application is limited, for example it is an experiment running for a few weeks within a laboratory environment, a numbering scheme can be easily devised with IDs assigned temporarily, and later discarded without further penalty, in any way that is convenient within the specific context. On the other hand, if an object addressing scheme with wider scope is needed, for example one that operates over a prolonged period of time and involves potentially several organizations or authority domains that must co-ordinate, then it is necessary to ensure that code uniqueness guarantees are observed and that associated address assignment and management processes are in place and formally agreed. One such scheme is provided by the ISO/IEC 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).
Under this scheme, a guaranteed world-wide unique serial object identifier (i.e. the object ID) is associated with an artefact by its manufacturer at production time. ISO 15459 codes have four parts: data identifier (DI) header, issuing agency code, company ID, and serialized item code (cf. Figure 2 for an example). In conformance to previous related ISO standards each part of the code holds alphanumeric digits rather than numbers. The DI specifies the structure of the contents of the object ID and follows the specification of ISO/IEC 15418 encoded under ANSI MH 10.8.2 provisions. For example, DI set to 25S specifies that the object ID is a globally unique serial object number, and DI set to 2L specifies that the object ID is a location specified in a format defined in a subsequent field, for example a post code. Rules for the coordination of the address space are also defined in the standard with the Netherlands Normalization Institute being the only authorized registrar that can assign IACs. EDIFICE, an association of electronics suppliers, is such a registered issuing agency and can thus provide its members with their individual unique company identification numbers. Each member can then decide internally on how to structure the object serial numbers. A common approach is to separate the number in 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.

An important feature of ISO 15459 is that it accommodates a variety of existing product classification schemes that can be used as object identifiers. For example, the currently most popular way to tag objects is by way of a barcode, mostly using identifiers specified with in the EAN/UCC system. Setting the IAC to 00:EAN allows the automatic use of such codes. This is a considerable facility as it allows the immediate use of a very popular and widely used system without further administrative overhead. This approach also allows the incorporation into the system of a number of other domain specific numbering schemes under a unified hierarchical classification. For example, ISO 14223-2 defines a code structure specific for use for animal tracking including information on the the species and the premises where it is held. These codes are incorporated under ISO 15459 simply by 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 provide flexibility for future extensions.

Another proposal for a general RFID numbering system is the so-called Universal ID (uID) specification, proposed by the Ubiquitous Networking Laboratory.
at the University of Tokyo. uID has a similar structure to the ISO specification but it always has a set length of 128 bits which represent numeric rather than alphanumeric values. The code is subdivided in five sections namely version, top level domain (TLD) code, and class code (CC) which specifies the boundary between the remaining two fields: the domain code (DC) that specifies the type of the particular identification code (IC) which comprises the remainder of the uID. The DC is the mechanism by which other identification schemes can be incorporated to uID and there are already provisions for the most common types including JAN (the Japanese standard for bar codes), ISBN (for books), and the Electronic Product Code which we discuss next. 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.

Perhaps the most successful numbering scheme for RFID in terms of industrial adoption so far is the Electronic Product Code (EPC) which has already been introduced in Section 2.3. EPC codes are assigned to product manufacturers and end-users via GS1 national representatives (GS1 stands for One Global System which is to be taken as a statement of intent regarding the development of the Internet of Things). There are several different types of EPCs that can be used for products, locations and containers, and a common EPC type is the so-called Serialized Global Trade Identification Number (SGTIN) which comes in different lengths. The structure of the 64 bit SGTIN is very similar to both ISO and uID specifications but depends on the existing EAN-13 bar code standard for its product classification scheme. SGTIN-96 codes are made up of four parts namely, the Filter Value which allows the pre-selection of the object type (for example whether it is an item, a case, or a pallet), the Company Prefix Length which defines the actual length of the EAN/UCC Company Prefix which varies in length. The third field is the Global Trade Identification Number (GTIN) which is composed of a Company Prefix and an Item Reference which identifies the class of the object. The last section is the Serial Number which is a unique number assigned to an individual object by the managing entity (the holder of the Company Prefix).

Although not evident from the above description, EPC also supports interoperability with ISO standards albeit at a lower protocol layer. Gen 2 tags provide a parity bit as a toggle to indicate the type of identifier stored in their EPC memory bank (cf. Bank 01 in Figure 1, the Numbering System Identifier is part of the Protocol Control section) so that other numbering systems can be used instead of EPC. Although the EPC scheme clearly has few differences
with the previous ones and indeed several limitations when compared against ISO, it has nevertheless attracted considerable 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 (cf. Section 4). Moreover, several IT vendors have already integrated these specifications into their products and as a result the EPC standards have gained considerable advantage against competitors.

3.1.4 Sources of RFID Read Errors

Due to chip performance limitations, most RFID tags are not capable of supporting a full microcontroller instruction set and thus implement a simple state machine instead. For example, EPC Gen 2 tags can be only in one of seven different states, which they reach in response to current and past interactions with readers. Older tags, developed even a few years ago, would have far fewer states and would implement much simpler state machines. The considerably more complex operation of Gen 2 tags are due to the current need for RFID systems to operate in far higher tag densities and the associated requirement for multiple co-located readers with overlapping ranges. This situation is particularly challenging and together with the effects of harsher operating environments leads to very high error rates, which could be as high as 20 per cent even in relatively controlled circumstances.

The source of such errors varies. For example, Figure 4 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, which is not the case for two other types of events namely, partial and ghost reads. Writeable 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.

Efficient ways to share the air medium is a particular acute problem for modern RFID as larger systems imply a higher likelihood or errors. It is thus necessary to provide more effective coordination protocols that allow readers in particular to pinpoint and isolate their communication peer, and to maintain several concurrent channels while guaranteeing high performance. The majority of modern RFID systems address this requirement by employing some variant of 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 which can be addressed in isolation [27]. Such carefully designed techniques can solve in practice several of the problems identified above, there is good evidence for example that ghost reads in particular have been eliminated in Gen 2 tags.

Another significant source of errors is due to the presence of specific materials that cause disruption to the signal. For example, when metals are placed between reader and tag in most cases they would completely absorb 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 sensitivity to different materials: Near-field systems are mostly unaffected by dielectric or insulator materials for example paper, plastics, masonry, ceramics, but metals weaken the field (depending how ferrous they are) and they 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, in UHF frequencies tag-on-tag effects (like those depicted in Figure 4) are particularly strong due to higher tag densities and increased reader range.

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 extend by the use of multiple reader and tag antennas in complementary orientations and alignments so that the likelihood of a tag placed completely perpendicular to all antennas be minimized.
4 RFID Software and Network Services

To appreciate the complexities of large-scale RFID software development, it is first necessary to summarize the different stages of RFID processing. Moving sequentially from the lower level where observations are acquired by a reader towards application level processing, RFID captured data enters the following stages:

- **Collect observations**: Readers interrogate their vicinity for the presence of tags and subsequently request and retrieve object IDs and potentially additional data stored in the chip memory (some systems would require an intermediate authentication step to allow access to this information). Depending on the application, the duration of the interrogation cycle can vary considerably. For example for e-passport applications a read cycle could last up to a minute, while in supply chain applications several hundreds of tags would be read per second. The read phase could 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 for example temperature sensors could be used to record the environmental conditions in which a particular object has been observed and LED displays could be operated to indicate the state of the object.

- **Smooth observation data**: Raw observation data can be erroneous and incomplete as a result of read errors. Smoothing observations is the process of cleaning the collected data from incomplete reads that are discarded; from IDs recorded due to transient and thus irrelevant objects that must also be removed; from indeterminate reads must be resolved (for example using authoritative records from local persistent storage); and last but not least tags that have not been read must be rescanned.

- **Translate 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.

- **ID resolution and context retrieval**: Specific object IDs recorded in observations and events must be associated with object descriptions and related contextual-use data retrieved. This conversion requires access to network services that play a twofold role: (i) to map object IDs to network service locations that can be further queried about object details, and (ii) to respond to specific queries related to the current condition, the properties and the history of the object.
- **Dispatch and processing of event data:** Application level events must be returned to consuming applications for further processing. For example, a pallet entry event would trigger updates of inventory records to include the items contained in the identified shipment.

Of course 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. An orthogonal layer to the application execution profile is infrastructure management that is, maintaining configuration and status information related to the operating condition of RFID readers and other sensor elements [18].

The sequence of tasks outlined above is carried out by distinct network segments [16]: observations are collected at the reader level outside the IP network; observation processing and event translation at the network edge by the event manager; and application logic at the network core (or data center) level. A layer of mediation between the network core and edge is provided by the network services and other event consuming applications, which have the role of resolving identifiers into object descriptions and the subsequent querying for associated and context data. Put together, these distinct elements define the RFID stack depicted in Figure 5.

![Fig. 5. The RFID stack.](image)

What is different with RFID compared with traditional internet client-server style computing is that a considerable proportion of the functionality is shifted from the network core to the edge. Here, edge refers both to the edge of the corporate network (referring to satellite sites for example) or the service provision data center (a web farm for example), and of the internet (where IP functionality terminates). But most significantly, edge is a reference to the fact that a considerable proportion of system functionality operates at the ends of the IP network, a location usually reserved for clients. This shift towards edge...
computing is typical in pervasive computing applications and is observed in a variety of related situations, notably with wireless sensor networks [103]. In such cases, information processing, content routing and persistence are all located at the internet edge which gains far greater importance in its role within the system.

The principal enabler of this edge processing capability within the RFID stack is the event manager [13], which has the role 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.
- Offer a single interface to applications.

In many ways the event manager is very similar in scope and function to peer-age architectures that are common in overlay networks including peer-to-peer [78] and pervasive computing overlays [93], as well as dynamic proxy servers in application level active networks [35]. 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 the complete system functionality. Finally, the event manager can be seen as the RFID-specific implementation of a system role that is common to pervasive computing namely, translating lower level to higher level context [59].

4.1 Programming RFID

In processing observations into events, the event manager defines a scan and query plan using a device abstraction layer which is relayed to and executed by the reader infrastructure. A scanning plan specifies the frequency of data acquisition, how many attempts are made, triggering conditions, and so on. It may also include information about the specific components of each participating reader that is employed for example which of the attached antennas will be activated. Naturally, this device abstraction layer also provides facilities for the discovery of reader capabilities (for example supported functionality, attached components, software versions and so forth) and can also request the pre-processing of the observation data if this functionality is supported by the reader. Finally, the device abstraction layer can also potentially support actions predicated on a triggering observation for example when a motion sensor detects movement. Examples of such device abstraction layers are offered by the Reader Protocol [RR] part of the EPCglobal standards, the Ubiquitous
Communicator API [95] and the generic interface of WinRFID [80]. Particular reader manufacturers have also developed such abstract device interfaces but these are less useful as they can only be used with readers from specific suppliers.

The event manager provides application programming interfaces for event discovery, subscription and reporting [34, 84]. This allows client applications to find what events are available and define new ones, subscribe to those of interest and receive reports with results. Events are defined over event cycles that is, delimited time intervals over which observations are processed. Note that although observations and events are related to read and event cycles correspondingly, the event manager decouples their respective domains and provides a clear separation of scope (cf. Figure 6). While the adoption of cycles as the main modus operandi for the event manager may appear limiting this is not so, as in addition to defining cycles either periodically or within fixed time slots, it is also possible to have arbitrary bounds defined on triggers fired by specific observations or by software interrupts or by external notifications.

Filtering and aggregation processing by the event manager aims to identify specific patterns in the event data and to summarize data collected from different readers over several event cycles correspondingly [102]. Filters work by applying include or exclude regular patterns that is, by setting rules that define ID lists or ranges to be included (or excluded) in the processing of observations. For example, following the EPC filtering specification, the exclusion filter \texttt{epc:gid-96:18.[321-326].*} encountered while processing EPC tags specifies that the product range that corresponds to product codes between 321 and 326 will not be processed irrespective of the serial number of the objects recorded. Similarly the aggregation pattern \texttt{epc:gid-96:*.*.X.*} results into grouping observations by product code and reporting only the total number of observations for each class of product. Due to relatively frequent read errors such filtering and aggregation techniques are rather complex to implement in practice and recent work highlights the significance of statistical techniques to improve data fidelity [53, 106].

The programming interface provided by the event manager can be implemented using different methods: the Application Level Events (ALE) specifi-
cation [11] is a middleware specification and the Java RFID System provides
the same abstraction as a language specific implementation of a component
model built on top of the Jini event management framework.

While there seems to be some consensus about the desired functionality of the
application event interfaces, the actual implementation of the event manager
can be done in several alternative ways. These alternatives are not mutually
exclusive but adapt to their operational context and explore different tradeoffs
between levels of functionality and performance guarantees [51,60]. In practice,
the event manager may consist of one or more distinct physical devices and
logical service end-points with the responsibility for specific tasks shared be-
tween them. A rather interesting consequence of this is that traditional barrier
lines that differentiate between computing and networking infrastructures are
 overrun, with data servers taking on data routing roles and network routers
acquiring application processing characteristics.

For example, Oracle’s Senor Based Computing platform [13] defines an event
manager to be a completely software service residing on a computing device
attached to the network edge. Unsurprisingly, the emphasis in this case is
in providing persistent storage of all captured observations, filtering and ag-
gregation as an extension of the relational model and event routing using
store and forward over web services (but also as raw database triggers). IBM
favors a similar solution but gives emphasis primarily on adaptation, acces-
sibility, system management and programmability of the edge services using
the Websphere platform [16]. On the other side, Cisco’s Application Oriented
Computing (AON) implements event managers as extensions to traditional
IP content-based routing. This is achieved by applying transformation filters
to particular data flows identified via inspection of packet content. Such fil-
ters are software components developed following the AON framework and
dynamically deployed to edge router devices. A common type of such a filter
would operate on XML data generated by RFID observations at a networked
reader (most current EPC-based readers would support this functionality) and
apply event generation rules. These differing approaches can be combined to
produce complex systems of very large scale.

4.2 RFID Network Services

To provide full functionality, the upper three layers of the RFID stack of Figure
5 require access to discovery and repository management services accessible on
the internet. Discovery services resolve captured object identifiers into network
service locations where repository services reside. Repositoy services in turn
can be further queried via standard service profiles to obtain trace and other
meta-data related to a particular ID.
Discovery Services. Mapping object IDs to network service locations is a relatively straightforward task, which can be easily accommodated within current internet infrastructures. One way to accomplish this is by simply using the directory capabilities of the Domain Name System, which can support an extended collection of record types. This approach is advocated by the Object Naming System (ONS) specification within the EPCglobal family of standards, which employs the Naming Authority Record [72] to provide associations of EPC codes to Universal Resource Descriptors. Under ONS, the serial item segment of an EPC code is removed, and the remainder segments reversed and appended to a pre-determined well known domain name (as of this writing onsepc.com). Of course, one problem with this approach is that ONS inherits and perpetuates the well known limitations and vulnerabilities of DNS, though some of these issues are addressed by the use of a single domain where delegation and updating can be handled with greater effectiveness.

An alternative solution would be to develop a completely new network service specification that provides this simple mapping via a secure overlay architecture. This approach is adopted by the uID Resolution Service within the uID system, which employs strong authentication and encryption to protect the system and the transmission of data. Unfortunately, uID RS currently supports only the TRON operating system and cannot be considered a general purpose service yet.

Moreover, both 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 artefact. 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 could not represent the complete data set.

Hence, rather than mapping an object ID to the service point provided by its manufacturer, the resolution process could alternatively point to a secondary discovery service instead, which maintains the record of the complete sequence of successive custodians, from production to the present day. This approach is
implemented in the so-called EPC Discovery Service which can be registered with the ONS and provide the list of URIs of all custodians for a particular object ID. This solution to maintaining a complete trace is preferable over the alternative whereby the current custodian would be identified via sequence of links through past holders. Such chaining is vulnerable to broken links that can easily occur for example, if any one of the custodians seizes to exist. One broken link would be enough to result in the complete loss of the ability to trace the object history.

**Repository Services.** The second element of RFID network services aims to manage and maintain object usage information and is provided by custodians. Conceptually, it is little more than a federated distributed database, and provisions for this task are offered by both EPCglobal and uID, within the EPC Information Service and the Ubiquitous Product Information Service correspondingly. From a usage perspective both standards are little more than a set of web service specifications to access object specific data repositories. Both provide methods to record, retrieve and modify event information for specific object IDs. 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 neither system has attracted significant support.

One feature of such repository services that merits further discussion today is the so-called containment profiles. This technique is necessary to form single objects out of individual components and be able to reference them directly. Consider the case of an automobile for example: it 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 in 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 is 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 expiry date and its elements can be modified via related containment interfaces.
4.3 Mobile Service and Content Discovery

The approaches developed in Section 4.2 are adequate for auto-identification of objects, humans and other species, and locations. RFID can be used to also identify mobile services and media content in which case additional facilities are helpful in providing a comprehensive solution. For example in many cases it is necessary to provide a means for mobile devices to bootstrap communication and agree on a communication channel, media format and so on, which would benefit from a localized rather than network approach. To address these additional requirements the concept of the so-called Near-Field Communication (NFC) was introduced by Philips and Sony in 2002 and in 2004 the technology was taken over by the cross-industry NFC forum which is responsible for the further development and promotion of the technology [77]. Although initially NFC was seen as a technology that operates between two sophisticated mobile devices, it has been recently extended to include interactions with RFID tags and readers and specific provisions for consumer applications such as ticketing and payments [17].

NFC-enabled mobile devices can interact with RFID tags that provide service discovery and can exchange data to the extend that devices can deposit information in tags embedded in the environment. NFC operates at 13.56 MHz and interoperates with ISO 14443 RFID tags. Typically, an NFC-enabled mobile phone can act both as a reader and as a tag. Although NFC does not necessarily require a cellular network, the vast majority of NFC devices are smart mobile phones and the standard is specifically focused to this use context.

Compatibility across NFC devices is facilitated via standard data formats for example the NFC Transfer Interface Packet (NTIP) which defines how collections of NTIP records can be assembled and exchanged with other devices. Such records can be used to describe actions for example to instruct the phone to call a specific number, access particular content online or send an SMS. In this way, RFID tags effectively become physical hyperlinks or physical shortcuts for specific actions. For instance, users can affix tags on walls and annotate their content with associated actions that will be executed when the phone touches the tag. This physical hyperlinking feature is greatly enhanced by the availability on most NFC phones of a related feature provided by the embedded Java virtual machine, namely the ability to automatically launch an application in response to instructions from a tag. For example, in [1,79,83] this facility of NFC is used to use mobile telephones as the tangible interface to pervasive computing systems.

The selection of near field RFID technology as the basis for NFC has the added advantage that RFID tags in this frequency can support extended memory capacity compared to UHF tags and thus facilitate longer service
specifications. Nevertheless, the rapid growth of far field technology and especially its popularity in the marketplace offers an attractive argument for its use also for service identification in a manner similar to NFC. In this case, a challenge to be addressed is the low memory capacity of the tag which in most cases can only hold a single identifier. This problem is addressed by the Mobile RFID protocol [67] that employs ISO 18000-6C for communication between smart phone and tag which is compatible with EPC Gen 2 tags. 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 and compatible with both EPC and ISO 15693 systems. Since it is not possible to store complete service descriptions within a single tag, service IDs are mapped to service descriptions using the Object Description Service (ODS) which is very similar in operation to the ONS as it uses the same technique to utilize existing DNS infrastructures. In response to ODS queries, mobile devices receive URIs containing the service endpoint which in the case of Mobile RFID is often a simple WAP URL.

5 Security and Privacy

Whenever applications of RFID are discussed issues of privacy and security are almost always raised, and with good reason. The invisible operation of RFID in particular opens up unique opportunities for collection of personal information by authorized and un-authorized parties alike, without any evident indication to or the consent of the person observed. This perilous lack of user control over this technology is aggravated by the surprising laxness with which such issues have been addressed (or rather not been addressed) in the recent past, and in some cases by deliberate attempts to create de-facto situations where privacy protection is intentionally compromised.

The vast majority of far-field standards used within open systems and the supply chain in particular, fail to make adequate provisions for security and privacy protection. The situation is better in closed/proprietary RFID systems which often implement strong private key cryptographic protection and offer some degree of security. Note that the strong cryptographic techniques discussed in the following sections have not yet had significant impact outside the academic community, which has only recently focused its attention on these problems. Research in this area is expanding rapidly, and as a result it is not possible to provide a full review of current developments in the context of this survey. Readers interested specifically in this area should refer to [38] for an overview and to [39] for a detailed discussion of implications for privacy. The comprehensive survey of security techniques in [55] is a good starting point for details on particular mechanisms and the current state-of-the-art can be
always reviewed in the online bibliography maintained in [5]

5.1 Privacy

Privacy violations due to the use of RFID can be broadly split in two classes: tracking, whereby the actions of individuals are recorded and their future behaviors potentially inferred using RFID tags associated with them; and information leak, where personal or intimate information stored in RFID tags is revealed without the consent of its owner [39, Chapter 4]. Both problems are related to the invisible and unsupervised operation of RFID, which implies that if left uncontrolled, tags would effectively broadcast their unique IDs to any party interested. These identifiers can be recorded and correlated without any visible indication that this activity occurs.

Privacy protection concerns are caused primarily by the expected widespread use of RFID for item-level tagging, although other systems can also be amenable. Despite the fact that such item-level tags are primarily intended for supply chain applications, if they are not removed or otherwise disabled at the point of sale they remain operational and can be interrogated while the products are in the possession of consumers. For tracking in particular, it is not necessary to gain access to the object ID but the technique can be successful using the session or tag ID only. Of course, poor access control of RFID tags can lead to further information leak of object IDs which can be used identify a person through their ownership of specific artifacts. In such cases, scanning individuals and their garments reveals valuable information about their favorite brands and thus their habits and financial status. It can also have other consequences for example, to indicate that the individual suffers for a specific illness as a result of the fact that they carry some type of medication suitable for their condition.

Consumer privacy violations can be are 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 infringement of ones rights [38]. Such threats include the collection of data in unauthorized locations; the lack of control over the data collection process; the prediction or inference of behavior; and the subsequent reduced capability to make free choices. Such threats have been identified from early on in the use of RFID in the context of retail and have caused significant concerns among consumers [86]. The principal complaint is related to the invisibility of operation of RFID and the subsequent perception of loss of control: consumers feel that they have little or no say over both the RFID infrastructure and the use of the collected data. It is clear that before the technology and related services can be widely accepted at the marketplace, it’s users must be able
to exercise their will over it, at least to a significant degree [43]. Although the technology itself is the cause of the majority of such concerns, early commercial applications have not helped to develop public confidence as many events show. For example, Metro Supermarkets in Germany violate their own stated privacy policy by embedding covert RFID tags in their loyalty cards, and an early briefing of Auto-ID Center sponsors urged them to capitalize on consumer apathy and push for item-level tagging thus creating a de-facto situation before consumer organizations could react [22].

There is of course a straightforward solution to addressing all privacy concerns, namely to ensure that tags are removed or completely disabled at the point of sale. However, this approach to RFID privacy would prevent retailers to develop after sales programmes which potentially can provide very significant competitive advantage via differential pricing [3,4], direct marketing [24] and customer relationship management [40,41]. This would remove their incentives to implement item-level tagging and would make the deployment of the technology very unlikely. Removing the tags would also cancel any potential uses of RFID outside the scope of retail for the development of useful pervasive computing applications. For example, a number of proposals have been published recently in which item-level tags are used in robotic and other assistive technologies [44,104,105], to help the blind navigate home environments [111], or the elderly suffering from Alzheimer’s disease [29]. In any case, the balance in the value proposition of item-level RFID would remain on the side of the supplier rather than the consumer [96].

A definitive solution to tag removal is to place them so that they are clearly visible and easy to remove and discard, much like conventional price tags. It is also possible to identify and destroy RFID tags in a variety of ways, though many of these would also cause considerable damage to the object in which they are embedded. One way to safely permanently disable a tag is specified by the EPC Gen 2 standard that provides the so-called “kill” command which permanently prevents the tag from communicating further – however, access to this command is protected via a simple 32-bit password and this approach appear to be extremely limited at scale. Furthermore, the vast majority of tags implement this command in software and are still vulnerable under differential electromagnetic emissions analysis.

Short of completely discarding RFID tags, approaches to protecting privacy require access control to the object ID. A potential solution is to either suppress access to the object ID altogether at the point of sale for example using hashing [91], or allow consumers to change the identifier and control the granularity of data that becomes available to interrogators [50]. However, when tag constellations can be observed neither approach is completely effective in preventing tracking. This threat can be addressed by equipping a tag with a pre-defined set of pseudonyms that the tag can emit in alteration [56], thus
allowing only authenticated readers to correlate the various pseudonyms associated with each tag. The scalability of this solution has been improved by incorporating a time-memory trade-off in the tag computations, or by enabling the reader to keep track of the pseudonyms without a need for a central server. Alternatively, a periodic renaming strategy can be developed on a hash function [75], or a physically un-clonable function [9] which leverages the inherent manufacturing differences of each tag’s circuit layout to provide a light-weight cryptographic function. Finally, a rather promising approach has recently been proposed whereby the antenna of the tag is physically severed thus radically reducing the read range of the tag which remains operational but practically unusable without contact [58].

To address the issues inherent in relying on public RFID readers to enforce privacy, consumers might choose to carry their own privacy-enforcing devices. An example of such a device is the so-called Watchdog Tag [33] which can monitor read attempts and collects reader information. Similarly, the RFID Enhancer Proxy [57] mediates between readers and consumer tags to enforce privacy-preserving communication. Other approaches aim to completely prevent tags carried by particular individuals from being read for example tag cloaking using a Faraday cage which can be easily constructed using for example duct tape or tin foil, and active jamming [54]. Alternatively, tags can incorporate low-level circuitry that can calculate the distance to a reader, and choose to only respond to requests from very proximate readers which can be inspected visually [30]. Finally, a system designed specifically for public transport ticketing proposes the use of anonymous credentials and proxy re-encryption to increase passenger privacy [49].

Of course, any technological approach is only part of the solution [88]. It is also necessary to develop a secure context of use where consumer rights are protected and to this end there is already related legislation in several US states. Moreover, both the US Federal Trade Commission and the European Commission are carrying out extensive consultations on these matters. Until such legal endeavors become mature provisions for consumer notice and choice will be hotly debated as in the case of the RFID Bill of Rights suggested in [37].

5.2 Security

On the related issue of security, there are three aspects of RFID that present specific challenges for pervasive computing applications:

- use of unauthorized or cloned tags to gain access to a controlled system,
- forward and backward data security,
- eavesdropping and replay attacks.

Although these risks are fairly straightforward to identify, early RFID systems provided no protection to tags [110]. Anyone with a reader capable of interrogating the tags has been allowed access [39, Chapter 2] and this practice is still in use in a surprising number of commercial systems, notably the VeriChip tag which is specifically designed for use with humans [45].

Most recent commercial deployments incorporating private key encryption are relatively robust to tag cloning. However, weak encryption implementations do exist and are open to attack, as are implementations based on inadequate random-number generators [6]. Both types of systems are open 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 DST RFID device is described in [10], where the authors demonstrate the exhaustive key search of the employed 40-bit encryption mechanism. This is achieved using modest resources including an array of 16 FPGA boards, and simulating DST output using a programmable radio device. This attack demonstrated that using a single pair of challenge/response values an attacker could clone the DST.

In addition to averting unauthorized access, RFID tags need to provide forward and backward data security that is, even in cases where the object ID can be retrieved, it is still not possible to trace the tag through past and future events in which the tag was or will be involved. A mechanism for forward security is proposed in [75] that employs hash chains to renew the information contained in the tag. Backward security is a much harder problem and has only been recently identified as a concern with [56] describing one possible approach to this problem using one-time pads.

In relay attacks, a “leech” device is positioned close to the RFID tag and an associated “ghost” device is positioned near the target reader. The ghost device gains access to the systems by relaying challenge-response queries between the target reader and tag via the leech. This attack has recently been demonstrated in [47] against an ISO14443-A system (which employs fixed session IDs and is somewhat easier to deal with). Possible countermeasures against such attacks using a bounded communication limit are proposed in [30,46]. Eavesdropping attacks are carried out by rogue readers that remain silent but can monitor communication between a reader and tag during a session. Attacks on RFID can also use variations in the speed of computation (timing attacks) or power consumption (power analysis) of the tag when it carries out calculations to decide whether to accept or reject a presented credential [14]. Finally, physically tampering with a reader can provide useful information to reverse engineer a tag that can be used to gain unauthorized access [2].
The attacks on readers and tags of course are only part of the story as complete RFID systems can also be successfully exploited through different components for example by compromising their middleware or network services. For example, the EPC system relies heavily on DNS which has well known vulnerabilities that can be used to penetrate ONS [101]. Software errors in RFID middleware can also be used to compromise the system using standard techniques for example exploiting buffer overruns [82]. However, such attacks are within the realm of traditional security and RFID appears to have limited effect in this regard except perhaps opening another avenue of attack.

6 Advanced Applications of RFID

Despite its numerous limitations, RFID nonetheless provides a more than adequate mechanism to explore the requirements, the opportunities and the implications of novel pervasive computing applications. To this end, RFID has been used since the earliest ubiquitous computing explorations to provide object auto-identification functionality and in this section we review some of these applications and the issues they raise for the future use of this technology.

Context awareness. An immediate application of RFID is in supporting context awareness by supplying facts relating to proximity between specific objects and users [73] and thus the information required by the system to adapt and offer flexible functionalities to its users tailored to their specific needs within their particular situation [21,94,107]. Typical examples of such adaptation using RFID include medical facilities for example, hospital beds that fit the needs of particular patients with particular ailments [7]; 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 that identifies the member of the family being proximal by their use of their RFID-tagged toothbrush, and retrieves and displays traffic and weather information specific to their trip to work [36]; and assisted living applications for those with mental disabilities including autistic children and elderly persons with Alzheimer’s disease for example, the iGlove that employs RFID to record interactions with objects so as to learn common patterns of behavior with a view to helping with everyday activities [29].

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 where installed with readers and connected to a server containing additional associated content [31] 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 with 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 be then 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.

**Socialization and Social networks.** In the majority of situations considered in this survey, the emphasis has been on situations where a person is interacting with one or more objects and the role of RFID is to provide mechanisms for their automatic identification. Yet, perhaps the most interesting applications of RFID and indeed those closer to the ubiquitous computing spirit, are the ones that use RFID to enable individuals to interact with each other and in some cases to even support complete social networks. One way that RFID can enable such socialization is through the use of of the object itself as the interface to carry out collaborative work [52], that is RFID can and has been used as the foundation for basic but functional tangible interfaces.

In some cases such interaction can be asynchronous, for example a physical label holding an RFID can be used to hold information that can be disseminated to others. The application of this technique can facilitate the wireless provision of instructions for a game played in the real world via a mobile device [81]. In this particular case, it is not the object itself that is being manipulated but rather a mobile RFID reader (integrated into a smart phone) is used to facilitate this exchange of instructions.

Another way that RFID can enable social interactions is in coordination with public displays [71]. Such situated displays can sense co-location 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 [109]. Such ideas have been further explored to bring together combined physical and digital co-occurrence that is, to investigate concurrent proximity of authorship and location again within an academic conference environment [61]. In this case, public displays are also used to explore common links to the same social groups as the enabler of further interaction between conferees attendees.

**Writable RFID.** One of the most interesting facilities of RFID that is also the least unexplored – possibly due to the lack of workable security models – is the ability to not only read but to also write information into the tag. Although in most cases the capacity of the tag is limited to less than 4Kbits,
it is nevertheless adequate to hold enough information to develop interesting applications. One such application brings together pervasive computing and robotics in a system that adopts a mode of operation inspired by ant colonies [69]. 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 working collectively in a decentralized manner that would otherwise have been impossible without the use of significant infrastructure dedicated to coordination. Similar exchange of messages using tags to hold the information is of course also possible between humans though this mode of communication is yet to be explored at any depth [81].

**RFID and wireless sensor networks.** One view of RFID is that it is the simplest possible – albeit more mature – type of wireless sensor network [48]. Seen in this light, RFID “networks” have two fundamental limitations that prevent them for supporting relatively complex data processing operators: 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 rather than simple identification. For example, [76] discusses how temperature sensing capability can be added to passive RFID and work in the same vein is carried out elsewhere. 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 Section 2 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 that 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 [70], and it is expected that more will find their way to actual deployments. 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 wake-up on demand 
[74], although its power saving capability appears to be limited.

**Higher-level programming models.** In Section 4, we discussed the sys-
tem components that are required to create and program complete RFID in-
frastructures. However, 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 [60]. As a result, alternative higher-level 
approaches are necessary to capture application requirements in more effective 
ways.

There is considerable interest in explorations that attempt to address this 
problem which is not unique for RFID but affects the majority of pervasive 
computing software development. One approach advocated in [18] is to intro-
duce domain-specific models and associated frameworks that provide system-
level abstractions and use those to develop applications. An alternative is 
proposed in [84] which favors the development of general purpose frameworks 
that provide general purpose primitives which can be used to programme RFID 
systems irrespective of the application. Finally, an extension to the EPC pro-
tocols specifically to support location tracking at a high level in a manner 
similar to the approach adopted in cellular networks is proposed in [20]

However, such high-level approaches do not always cater well to the peculiar-
ities of specific RFID systems but are forced to abstract to the lowest common 
denominator thus missing opportunities for improved performance. One ex-
ample 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 
failure to use local storage on HF tags that can very considerably improve 
system robustness [62].

7 Summary and conclusion

This survey has reviewed developments towards establishing RFID as the cost-
effective technical solution for the development of first-generation, open, shared, 
universal pervasive computing infrastructures. RFID is an effective automatic 
identification technology for a variety of objects including physical, manu-
factured and handmade artifacts; humans and other species; locations; and 
increasingly media content and mobile services. Because of the limited capa-
bilities of the tags however, RFID systems have to rely heavily on network 
services to support full system functionality.

RFID has already found its way into a variety of large scale commercial ap-
lications and arguably it is already one of the most successful computing platforms. Yet, RFID falls short of delivering the full pervasive computing vision although without doubt it is very useful in prototyping and other exploratory investigations in this context. Furthermore, RFID is only the simpler of a number of rapidly advancing wireless sensor network technologies that are gaining new capabilities and offer improved performance, and may soon challenge RFID in certain applications. Especially in cases where more than automatic identification functionality is required, RFID may soon provide a less attractive tradeoff compared with newer technologies.

Moreover, further development of RFID technology depends on successfully addressing two fundamental concerns: economic and environmental sustainability, and adequate privacy protection. Indeed, despite the massive reductions in cost of the past decade, RFID is still not cost-efficient for a variety of tagging targets. To achieve such universal taggability it is necessary that production costs for individual tags fall below a certain level – the exact tipping point is as yet unclear [108] – which seems to be unlikely in the short term due to market, intellectual property and manufacturing reasons. Such very large scale deployment of RFID that would allow in practical terms the assumption that every object is automatically identifiable is perhaps possible in the long term and it will require further technological advances. Universal deployment of RFID also depends on robust business cases that satisfy all parties involved and are far from forthcoming at the current state of development. There are reasons to doubt that this situation will change even in the medium term.

Perhaps more critically, RFID manufacturers are yet to address to any extend the considerable implications of this technology for the environment. Current regulation in both the EU (Directive on Waste Electrical and Electronic Equipment) and the US (California’s EoCycle legislation for electronic waste) demand that electronic components are recycled and yet there are no such provisions for RFID and none are planned. Even more so, product packaging containing RFID tags poses a very clear and substantial threat to the wider recyclability of both residential and industrial waste as it contaminates the materials and prevents effective processing of paper, plastic and glass, but also of pallets, boxes and other supply chain containers. Without doubt, these concerns have to be address effectively before we embark on RFID deployment at global scale. Indeed, allowing for such sustainable practices should be a central consideration and the costs of cleaning up must be included in the estimates.

Such universal tagging also has very considerable privacy implications that are far from being adequately addressed by the proposals reviewed in this survey. Nevertheless, it is unlikely that these concerns can be addressed within our current framework for discussion and it appears to be necessary that we rethink a number of provisions within the technological framework but also within our
legal systems. Without doubt universal RFID deployment would involve some type of tradeoff whereby individuals will be required to give up a proportion of their privacy in exchange for added value. Although this tradeoff can be clear and welcome in some cases, in others, the relationship between benefit and cost is less evident and in some cases may inevitably lead to coercion and suppression of choice.

References


[107] R. Want, K.O. Fishkin, A. Gujar and B.L. Harrison, Bridging physical and

25-33.

Based Multi-Service System for Supporting Conference Events, in Proc. AMT

RFID: Perspectives, Policy, and Practice (Addison-Wesley, 2005).

[111] S. Willis and S. Helal, RFID Information Grid for Blind Navigation and

[112] Y.Z. Zhao and O.P. Gan, Distributed Design of RFID Network for Large-Scale