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.

“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 simple and straightforward to implement. Yet, the opposite is true: RFID is a technology that requires the development of distinct systems, software and networks of considerable complexity. It entails antenna design, radio propagation analysis, low-cost integrated circuit production techniques (increasingly focused on printed electronics), receiver design and data encoding mechanisms, lightweight encryption and security protocols, materials technology, network discovery and information services, and novel interaction design approaches—to mention only some of the numerous engineering and computing disciplines involved in the development of complete RFID systems.

It is not surprising then that following Stockman’s statement in 1948 [1], RFID required almost 60 years of development to find its way to large scale applications [2]. And also, the fact that so many disciplines are involved implies that a full treatment of every aspect of RFID cannot be contained in a single survey. Instead, this paper addresses those aspects of RFID that are useful to the developer and the researcher of pervasive computing systems. We hope that the survey will be useful both as an introduction to the technology and as a guide to emerging research trends and novel applications.
1.1. Rationale and overview

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 internet infrastructures that can provide networked services to complement RFID and thus provide complete system functionality. These developments have allowed large-scale commercial applications in the supply chain [3,4], ticketing [5], asset tracking [6], maintenance [7], retail [8], and personal identification [9]. Due to these applications, RFID has become one of the most numerous computing platforms in use today: IDTechEx, a market research firm specializing in RFID, estimates that more than 3.7 billion RFID tags have been deployed in the field by mid-2007, with more than 1.6 billion new tags employed in 2006—and this trend is accelerating. The increasing popularity of RFID permits further cost reductions, and has revived interest in the use of RFID in pervasive computing research due to the unique opportunities it offers for low-cost large-scale experiments of novel systems and applications.

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 entity identifier and potentially a very limited amount of contextual information. This code is often employed as the handle that links a specific physical entity to its stored information. As a consequence, a considerable proportion of system functionality must be located on the network. For example network directory services could be used to relate the code retrieved from a tag to entity-specific metadata including the description and associated attributes of the tagged entity.

Pervasive computing applications can employ RFID to embed computation and communication capability into a variety of objects - whether natural, manufactured or handmade - thus making them fully-functional elements of a system. Applications can also use RFID to automatically identify humans and other living species, as well as locations, media content and data services. It is easy to anticipate pervasive computing environments that incorporate tremendous numbers of tagged entities, so that spatial tag density is very high. Such systems would be capable to automatically identify and collect information about entities and interactions between them in a manner completely transparent to users and in such a way that no manual intervention is required. Such infrastructures could convert simple observations into higher-level events and related abstractions that can be used for building end-user applications.

For RFID to enable this vision of truly pervasive computing in the real world, it must offer open, shared, scalable, and universally accessible infrastructures. Such infrastructures would support innovation at the application layer in a manner similar to the internet, adhering to the spirit of internet protocols and engineering drafts, which have allowed the deployment of a common global network that has supported the development of novel applications of global reach including email, the web and several peer-to-peer overlays. This vision for RFID has been dubbed – albeit inaccurately – the Internet of Things [10]. One of the main aims of this paper is to report 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 in the recent past. There are many valuable lessons discovered by this work that has moreover established the de-facto operational context for modern RFID. This appears to be unfamiliar territory for a large proportion of the pervasive computing research 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 wider research problems, 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 second part of this section we briefly outline three mature industrial applications of RFID which have played a central role in its rising popularity. In Section 2, we briefly review the basics of RFID technology and identify some of the tradeoffs in selecting a particular system. Section 3 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 capabilities.

1.2. Industrial applications

In this section we briefly review three applications of RFID that have been fully deployed in commercial systems. Emphasis is placed on those features demanded by large-scale and openness and are often not considered in laboratory-based explorations. The interested reader can find more details about these applications in [11].

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 document (MRTD), that uses standard RFID technology to store personal and biometric information on passports, visas and travel cards. Due to the current capacity of tags, in practice, biometric data are restricted to a photograph of the bearer but the standard also provides specifications for iris and fingerprint records. At the time of writing, more than one hundred countries across the globe have implemented or are in the process of implementing the system. 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.

The ICAO provisions call for ISO 14443-compliant RFID tags (discussed in Section 2) embedded in travel documents. RFID readers operated by immigration services interrogate and retrieve traveler information without the need for manual intervention. 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 [105]. A long-term goal of ICAO is 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.

1.2.2. 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 having been in operation for 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 code that identifies uniquely the ticket or the passenger, and in some cases also maintains 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 [5] in London, which supports more than 10 million active passengers and has deployed over 27 thousand readers. Other popular systems include the Octopus card in Hong-Kong and the Suica card in Tokyo, both of which are notable for the fact that their use has been extended beyond ticketing to general-purpose micropayments accepted at a large number of retail outlets.

Similar to the majority of RFID ticketing, the Oyster card is based on the ISO 14443 standard that provides specific facilities for transport applications. The relatively short range of this RFID technology 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 supported read-write 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.

1.2.3. Supply chain management

Supply-chain management (SCM) has been one of the main drivers of RFID technology in recent years. SCM deals with the movement of goods between organizations across a supply chain—from raw materials obtained by the manufacturer to finished products delivered to the consumer. Each supply chain is distinct and reflects the unique needs of the range of products that are processed, ranging from the delivery of fresh food from the farm to the supermarket shelf to army uniforms transferred from the manufacturer to the soldier in the desert. Nevertheless, all supply chains share a common goal: to keep the process simple, standard, speedy, and certain. To achieve this objective, it is necessary to ensure that all participating organizations exchange accurate information at frequent intervals and that related costs be unequivocally identifiable at all times. This in turn requires the use of open, worldwide data standards for globally unique product identifiers and product classification schemes, combined with internet-worked information services that can be used to track and trace goods and services [12].

By the early 2000s, RFID tags had become cost-effective as a means to automatically identify products and as a consequence, several pilot projects pursued their use to obtain superior-quality data collection for SCM. The perceived benefits of the technology convinced some of the largest retailers 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 – cases and pallets – rather than individual product items across open supply chains. A typical example of this would be an application installed at the dock door of a warehouse facility to record and cross-reference the codes retrieved from incoming and outgoing product containers. Such a portal would be equipped with RFID readers to read the tags of interest, and would automatically communicate with the company's resource planning systems to update the captured information.

The Auto-ID Center, established with the support of several manufacturers, suppliers and retailers of consumer goods [10], played a central role in highlighting the potential benefits of RFID in SCM. In particular, the Auto-ID Center introduced the Electronic Product Code (EPC) which provides an unambiguous numbering scheme used to identify product items and containers, services, companies, locations, and assets worldwide [10,3]. Auto-ID also developed a family of related standards which define how EPC codes are stored in tags and transmitted wirelessly as well as how to use an EPC to obtain entity-related data through discovery and information services on the internet. In the case of EPC, interest in RFID is as a replacement for barcodes as the principle means to encode information on physical entities. In this respect, RFID offers considerable advantages due to its higher data holding capacity and the fact that it does not require line of sight between reader and tag.

The current custodian of the EPC specifications is EPCglobal, a subsidiary of GS1 (previously known as EAN/UCC), an industry association that develops standards for SCM. EPCglobal aims to develop a full set of standards and federated service support infrastructures that support complete traceability of EPC-related information as a means to facilitate more efficient SCM. Such EPC-based supply chains are far more complex and challenging environments for computing than ticketing and e-passport applications and come closer to the vision of an Internet of Things. Their task is further complicated by the fact
that due to the characteristics of their operational environment readers may need to be deployed in dense constellations so that additional smoothing, filtering and aggregation of the captured data has to be performed to identify significant application-level events.

2. RFID basics

RFID has several peculiarities compared to other common wireless communication technologies and in this section, we outline the main stages of the identification process and highlight factors that affect its performance. In doing this we take a high-level view and point readers interested in the radio frequency aspects of RFID to [13] for a description aimed at a generally knowledgable engineering audience and to [14,15] for the full details.

2.1. Operating principle

Unlike other common types of wireless systems, RFID communication is asymmetric in that one peer – called the reader or interrogator – takes on the role of the transmitter and the other – called the tag or the transponder – the role of the responder. To a certain extent, this is the main reason for the success of RFID: rather than creating its own transmission, the tag is instead modulating or reflecting the electromagnetic waves emitted by the reader to communicate. This technique allows for a somewhat complex reader to be used with a very simple tag, which has small size and can be built at low cost. In many cases, the tag does not even require a battery since the electromagnetic waves emitted by the reader carry enough energy to be harvested and be used as its source of power. In practical implementations, a relatively small number of fixed or mobile readers can be used with very high numbers of tags thus keeping the overall system cost very low.

The 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 feasible although such 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 that occupied the majority of the tag volume. Shrinking electronics in the 90s have been critical in the development of the current generation of tags, which are both significantly more power-efficient and also provide higher storage and computational capability—both as a result of miniaturization.

One particular type of RFID, the so-called active RFID tags, carry 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 (cf. Table 1). 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, at what is often referred to as the sub-carrier frequencies.

Power transmission from the reader to the tag is by magnetic induction – the same principle is 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 and 15693 standards, the near field extends to about 3.5 m from the reader. However, in practice ISO 15693 systems would consistently work at a maximum range of approximately 35 cm using medium size antennas on the reader (radius approximately 30 cm) and credit-card size tags. ISO 14443 system typically offer a reading distance of up to 10 cm using a small antenna (radius approximately 8 cm) and credit-card size tags.

---

1 For a detailed history of RFID see [2].
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 in the 120–136 kHz range but these are loosing rapidly in popularity although they have a unique niche in applications where tag are embedded in living tissue.

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, in this case there is also very considerable loss of power and thus readers have to be extremely sensitive.

Because of the involvement of the far field, tag and reader antennas are also characteristic of their operating frequency and again this fact can be used to identify them. Tag antennas are dipoles, although in many cases there would be several attached to a single chip, and may also be "wiggly"-shaped so that they operate efficiently in different orientations. Readers can also use dipoles but for more effective operation they almost always employ patch antennas instead. Far-field RFID commonly operates in the UHF band between 865–956 MHz but note that the complete range is not available to applications everywhere on the globe and there are further operational restrictions. In practice, this implies that the same equipment would provide significantly different performance depending on its geographic location, roughly 1500 readings/s in the US and 600 readings/s in Europe.

Typical UHF tags are nevertheless able to respond at the complete range of frequencies and it is the responsibility of the reader to make appropriate selections compliant to a particular regulatory framework. UHF allows for longer-range communication compared to near-field tags and for example, modern EPC Gen2 tags can commonly achieve between 3 and 4 m. At this band a dipole tag antenna is approximately 9 cm (a wider tag antenna would be more effective in capturing radiated energy) and a reader patch antenna is roughly 25 by 25 cm. Using larger higher-gain antennas—for example Yagi-Uda or phased array antennas—and power amplification, the range of such a system can reach 10 m.

### 2.2. Readers

An RFID reader consists of the following 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; one or more networking interfaces; and in some cases, additional task and application specific modules for example, digital signal or cryptographic co-processors. The role of the control system is to direct communication with the tag and interact with client applications.

Readers receive one or more scanning plans from client applications or middleware, and implement it by issuing state transition instructions to the tags within range. This process usually has three stages: (i) broadcasting to all tags within range and receiving responses to construct an inventory, (ii) selecting a particular tag as the peer for communication, and (iii) exchanging information with the selected tag.

RFID readers can be simple stand-alone devices that communicate with a host system via a serial interface, but increasingly are complete network computing devices – akin to network routers – that provide advanced processing of streams of RFID observations, and connect to the internet via wired or wireless networks. Such readers are designed to deal with situations where large numbers of tags must be scanned in a brief period of time and when the ranges of several readers overlap and thus must take turns accessing the tags. 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, a fact that considerably complicates the operation of a reader.

---

2 Different countries have allocated slightly different windows for the operation of UHF RFID readers: in Europe RFID can operate between 865–869 MHz, while the US allows 902–928 MHz and Japan 950–956 MHz be used. Other countries have alternative arrangements. There are also differences in the number of signal channels and power output limitations, especially between the Unites States (50 channels and 4W EIRP) and Europe (10 channels and 2W ERP which is roughly equivalent to 3.2W EIRP).

3 A detailed study of the issues and trade-offs involved in the design of high-performance UHF RFID systems can be found in [16].
2.3. Tags

An RFID tag is a far simpler device than the reader and consists of the following three parts:

- 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 good example of modern RFID is the EPC Class 1 Gen 2 tag [16, Chapter 4], which operates at UHF frequencies. The chip has a relatively complex non-volatile memory structure divided in four distinct areas (cf. Fig. 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 (discussed in more detail in the following section) which is the identifier assigned to the tagged entity, and in some cases related metadata. The tag identification (TID) bank contains information about the type of the tag and its manufacturer, including a unique serial number characterizing the tag itself. Finally, the user bank is optional and can be used freely in applications.

Note that although we talk about radio frequency identification, a single Gen2 tag holds several unique codes that have distinct roles and semantics—including the fixed TID and the programmable EPC discussed above. A third identifier commonly in use, the so-called session ID, is used by the reader as the address of the tag during communication. The session ID is roughly equivalent to the MAC address of a typical wireless networking physical layer protocol but in some cases, for example for Gen2 tags, it is only locally unique.

The session ID may be fixed and stored on the tag, as is the case for ISO 14443 Type A tags. Tags that employ this approach can be easily traced by using the session ID as a handler, a fact that raises very considerable privacy protection questions, which we discuss in more detail in Section 4. For this reason most recent tag protocols implement a randomization process whereby tags use a pseudo-random number as their session ID, which is generated on-the-fly, while interrogated by a reader. For example, Gen 2 tags use a 16-bit pseudo-random number generator to obtain a session ID during power up.

2.3.1. Identifiers

The vast majority of applications of interest to pervasive computing, use RFID tags to identify specific entities, including objects, persons, locations, services, or other content. If the scope of auto-identification is limited for example, if it relates to a restricted experiment running for short period within a laboratory environment, a coding scheme can be easily devised with IDs assigned temporarily – and later discarded without penalty – in any way that is convenient for the specific purpose.

If an entity-addressing scheme with wider scope is needed, for example one that operates over a prolonged period of time and involves several authority domains that must co-ordinate, then a far more involved solution is required. In this case, it is necessary to ensure that code uniqueness guarantees are enforced and that associated address assignment and management processes are in place and formally agreed. One such scheme is the ISO 15459 specification on unique
identifiers with provisions on registration (Part 2), common addressing rules (Part 3), transport unit address provisions (Part 1), and item-level tagging for the supply chain (Part 4).

Under this scheme, a guaranteed world-wide unique serial object identifier – the entity ID – is associated with an artefact by its manufacturer at production time. ISO 15459 codes have four parts which in conformance to previous related ISO standards, holds alphanumeric characters rather than digits (see Fig. 2 for an example):

- the data identifier (DI) header,
- the issuing agency code (IAC),
- the company ID, and
- serialized item code.

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. The Netherlands Normalization Institute is identified as being the only registrar authorized to assign IACs. EDIFICE, a European 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 serial entity code. 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 enables a variety of existing product classification schemes be used as entity identifiers. For example, the highly popular EAN-13 barcode standard can be directly employed simply by setting the IAC to 00:EAN. 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 species and the premises where it is held. These codes are incorporated under ISO 15459 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 the domain code (DC) which is the mechanism by which other identification schemes can be incorporated to uID.

Perhaps the most successful entity coding scheme for RFID in terms of adoption is the Electronic Product Code. EPC codes are assigned in blocks to product manufacturers and end-users via GS1 national representatives, which also maintain a variety of associated barcode and e-business standards. 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). SGTINs are similar to the ISO codes discussed above with the notable distinction that are made up of digits only, which allow for more efficient encoding on the tag memory and communication. Similar to ISO and uID, EPC codes provide a hierarchical code structure with the notable addition of a filter value prefix which allows the pre-selection of the entity type for example, whether it is an item, a case, or a pallet.

2.3.2. Sources of RFID read errors

As seen in Section 1, modern applications of RFID require support for higher tag densities and thus for multiple co-located readers with overlapping ranges. This operating environment is particularly challenging and together with the effects of harsher radio frequency conditions lead to significantly higher error rates, which could be up to 20 percent in some cases. The source of such errors varies, with negative reads often caused by interference or due to collisions between tag transmissions. Interference can be the result of an external radio frequency source, for example a second reader or other wireless network, or metallic material in the vicinity. Collisions are often caused when two tags attempt to respond to the same command concurrently and can lead to the complete loss of communication or partial and ghost reads.

Efficient ways to share the air medium and avoid collisions is a particular acute problem for modern RFID as denser 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 an anti-collision and singulation technique that allow tags to be accessed in an ordered way and to pinpoint specific tags.
which can be addressed in isolation [15]. 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 error is the presence of certain 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: Inductive coupling 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. Capacitive coupling 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.

Writeable RFID systems also face the additional problem of indeterminate conditions. These occur when a write command has been issued but not confirmed and the tag is no longer in range of the reader. 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.

Last but not least, the effectiveness of energy harvesting depends on the alignment of tag and reader antennas. Both inductive and capacitive systems completely fail if tags are placed perpendicular to the reader antenna field. When the angle of orientation is less than 90 degrees there is still loss of performance the magnitude of which depends on the particular angle. This problem can be addressed to some extent by 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.

3. RFID systems and software

Due to their distinctive mode of operation, RFID systems have unique requirements on software and systems that can be summarized in terms of the RFID pipeline and stack. This section is focused on these abstractions and we begin with an overview of the different stages of RFID processing, moving from the lower level—where observations are acquired by a reader—towards higher levels of abstraction and application level processing. In this order, RFID data are processed according to the following pattern:

- **Collect observations**: Readers interrogate their vicinity for the presence of tags and subsequently request and retrieve entity IDs and related metadata. Depending on the application, the length of the interrogation cycle may vary considerably.
- **Smooth observation data**: Raw observation data can be incomplete or contain errors due to the restrictions on the communication process. Smoothing 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; and last but not least tags that have not been read must rescanned.
- **Translate observations into events**: Following smoothing, observation data are still not useful to applications, which are interested in higher-level events. This transformation of lower level observations into higher level application events is typically achieved via filtering and aggregation.
- **ID resolution and context retrieval**: Entity IDs must be associated with object descriptions, and related contextual and use data retrieved. This conversion requires access to network services that: (i) map entity IDs to network service locations that can be further queried about object details, and (ii) respond to specific queries related to the current situation, properties and history of the object.
- **Dispatch and processing of event data**: Application level events must be returned to consuming applications for further processing.

Applications control the data flowing through the pipeline by setting event specifications and declaring their interest in them. An complementary control plane operates to provide infrastructure management functions including the maintenance of configuration and status information related to the operating condition of RFID readers and other sensor elements [17].

The different tasks comprising the RFID pipeline are carried out by distinct network segments [18]: 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 entity descriptions and the subsequent querying for associated and context data. Put together, these distinct elements define the RFID stack depicted in Fig. 3.

What is different with RFID, compared with traditional internet client–server style computing, is that a considerable proportion of functionality is shifted from the network core to the edge. Here, edge refers both to the end of the internet – where IP functionality terminates – and also away from the center of the internet graph. Traditionally, internet service provision would be run from a centralised location, for example a data center or a web farm, located near the core of the internet and in physical proximity to the tightly controlled corporate networks—rather than external satellite sites. Instead, RFID systems shift a considerable proportion of system functionality near the edge 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 [19]. In such cases, information processing, content routing, and persistence are all located at the edge, which gains a role of far greater importance.

The principal enabler of this edgeware capability of the RFID stack is the event manager [20], which has the role to:

- Bridge the IP and RFID networks by translating RFID observations into application level events.
- Operate and manage the RFID reader infrastructure as well as related sensor and actuator devices.
- Offer a single well-defined interface to applications.

In many ways the event manager is very similar in scope and function to peerage architectures encountered in overlay networks for example, peer-to-peer [21] and pervasive service delivery networks [22], and dynamic proxy servers in application-level active networks [23]. Seen in this way, RFID infrastructures consist of a network of event management peers, which are transparently accessible through a single interface, and coordinate their operation via super-peers at the network core. From another point of view, the event manager provides RFID-specific context translation, a common feature in pervasive computing systems [24].

3.1. The RFID event manager

To convert observations into events, the event manager employs a scan and tag querying plan implemented at the device abstraction layer and executed by the readers. A scan cycle specification specifies the frequency of data acquisition, how many attempts are made, triggering conditions, and so on. Predicating a scan cycle on triggers set on external sensor events for example, when a motion sensor detects movement, is of particular interest as it can facilitate more efficient system operation.

A scan specification can also include the locations of data collection in terms of the specific readers that are employed. This task is facilitated by the device abstraction layer which provides mechanisms for the discovery of reader capabilities, for example supported functionality, attached components, software versions and so forth. Certain types of readers can also be instructed to carry out pre-processing of the observation data in a way that is appropriate in assisting the execution of a particular plan. Examples of RFID device abstraction interfaces are the Reader Protocol, which is part of the EPCglobal family of standards, the Ubiquitous Communicator API [25], and the generic API of WinRFID [26]. Particular reader manufacturers have also developed proprietary abstract device interfaces however these are less useful as they can only be used with readers from specific vendors.

Filtering and aggregation processing by the event manager aims to identify specific patterns and summarize observation data respectively [27]. Filters work by applying include or exclude regular expressions 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 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 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 [28].

The event manager provides application programming interfaces for event discovery, subscription and reporting [29, 30]. This allows client applications to find what events are available and define new ones, subscribe to those of interest.
Fig. 4. The RFID event manager.

and receive reports with results. Events are defined over 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. Fig. 4).

While there seems to be some consensus about the desired functionality of the application-level event interface, the practical implementation of the event manager can take several alternative forms. These alternatives are not mutually exclusive but should be seen as different tradeoffs between levels of functionality and performance guarantees [31, 32]. 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 between 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 [20] 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 aggregation 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, accessibility, system management and programmability of the edge services using the Websphere platform [18]. Taking a network-centric view, 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 filters 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-encoded RFID observations generated by a networked reader and apply event generation rules.

3.2. RFID network services

To provide full functionality, the upper three layers of the RFID stack of Fig. 3, require access to discovery and repository management services. Discovery services resolve captured entity IDs into network service locations where repository services reside. Repository services in turn can be further queried via standard service profiles to obtain historical and other meta-data related to a particular ID.

Discovery Services. Mapping entity 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 [33] to provide associations of EPC codes to Universal Resource Descriptors. Under ONS, the serial item segment of an EPC code is removed, and the remaining 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 replies on the eTP authentication facility which is native to the proprietary TRON operating system and cannot be considered a general purpose solution.

Moreover, both ONS and uID RS are limited by the fact that they only retain a single service location related to each entity ID for example, the URI corresponding to the manufacturer of a particular artifact. This is hardly enough in many cases: in addition to the specifics of the current situation of the object, many pervasive computing applications need to gain access to historical and 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: Entity IDs specifically for objects are assigned at production time from the address space controlled by their manufacturer, while the

---

4 Most current EPC-enabled readers would support this functionality.
artifact itself changes ownership several times during its lifetime. As a result, a naive resolution of the entity ID would point only to its initial custodian but the full entity history would be fragmented over different service locations, which would be untraceable. To support such context histories, the resolution process can alternatively point to a secondary discovery service instead, which maintains the complete record of successive custodians. This approach is envisioned to become part of the EPC Discovery Service at some future time.

**Repository Services.** RFID repository services record and maintain historical information related to specific entities collected by their custodians. Conceptually, such services are equivalent to a federated distributed database that is universally accessible through well-specified interfaces. Repository services have been specified by both EPCglobal and UID, as the EPC Information Service and the Ubiquitous Product Information Service respectively. Both define a set of web services that can be used to access entity-specific data repositories. And both, provide methods to record, retrieve and modify event information for specific entity IDs.

What does stand out when considering RFID repository services however is the massive size and complexity the databases involved, which – if successfully implemented – would be unprecedented. This task is further complicated by the complex network of trust domains, roles, and relationships, which requires careful management and enforced conformance to diverse access policies and regulations. Yet, at the present time these challenges are inadequately understood as neither system has significant deployments and as a consequence practical experience of their performance is very limited. Without doubt, further research in this area is required.

One feature of such repository services that merits further discussion today is the so-called containment profile. This mechanism is necessary to form a single entity by combining individual components, and be able to reference it directly. Consider the case of an automobile for example: it is made up of thousands of individual parts – mostly acquired from external suppliers – which at a certain point in time come together to make a single entity. Over the lifetime of a particular vehicle, these components will change as a result of maintenance, upgrades or changing use. Most applications require that the car as a whole can be identified, but in some cases it would be necessary to also refer to individual components. The containment profile has been introduced to address exactly such time-dependent processes, and is used within the EPC Information Service to group together entities into assemblies with their own unique EPC code.

### 3.3. Mobile service and content discovery

RFID can also be used to tag mobile services and media content, which require extra functionality from this technology. For example, RFID tags can provide a mechanism for mobile devices to bootstrap communication taking a localized rather than network approach, and agree on a communication channel, media format and so on. These additional requirements are addressed by the so-called Near-Field Communication (NFC) technology, introduced by Philips and Sony in 2002 and established as a cross-industry forum in 2004 [34]. Although initially NFC was seen solely as a technology operating between two sophisticated mobile devices, it has been recently extended to include interactions with RFID tags and readers and with specific provisions for consumer applications such as ticketing and payments [35].

NFC-enabled mobile devices can interact directly with ISO 14443 RFID tags and retrieve service descriptions and access details for content. NFC smart phones in particular are capable of acting as both reader and tag and at the same time provide internet connectivity over a cellular or other wireless network. As a result, Uniform Resource Locators are embedded in tags to provide redirection to the associated resource. The transfer of URL data across NFC devices is facilitated by the NFC URI RTD specification, which defines how URIs can be stored in a tag using the NFC Record Type Definition (RTD) format; how they are assembled; and how exchanged between devices using the NFC Data Exchange Format (NDEF) specification.

Through NFC, tags become physical hyperlinks that can be further associated with actions. This physical hyperlinking feature is greatly enhanced by the increased resource availability on most NFC phones, for example their support for embedded Java virtual machines, which allows the automatic launch and execution of applications relating to the interpretation of the collected URL. For example, in [36–38] this facility of NFC is used to transform smart phones into a tangible interface for pervasive computing applications that automatically identify and retrieve information relating to particular objects or locations.

The selection of near-field RFID technology as the basis for NFC precludes the use of the increasingly common EPC Gen2 tags. To address this limitation, the Mobile RFID protocol [39] relays identifiers captured from UHF tags to an associated Object Description Service (ODS), which resolves the code into a URI, where services or content can be accessed.

### 4. 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 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.

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 [40] for an overview and to [9] for a detailed discussion of implications for privacy. The comprehensive survey of security techniques in [41] 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 by [42].

4.1. 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 [9, 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 entity ID but the technique can be successful using the session or tag ID only. Of course, poor access control can lead to the leak of the entity ID, which can be used to identify a person through their ownership of specific artifacts for instance. 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 [40]. 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 [43]. 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 [44]. 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 [45].

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 [46,47], direct marketing [48] and customer relationship management [49,50]. 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 [51–53], to help the blind navigate home environments [54], or the elderly suffering from Alzheimer’s disease [55]. 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 [56].

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 entity ID. A potential solution is to either suppress access to it altogether at the point of sale for example using hashing [57], or allow consumers to change the identifier and control the granularity of data that becomes available to interrogators [58]. 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 [59], 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 [60], or a physically un-clonable function [61] 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 [62].

Of course, any technological approach is only part of the solution [63]. 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 [64].

4.2. 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 [65]. Anyone with a reader capable of interrogating the tags has been allowed access [9, 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 [66].

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 [67]. 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 [68], 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 entity 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 [60] 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 [59] 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 [69] 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 [70,71]. 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 [72]. Finally, physically tampering with a reader can provide useful information to reverse engineer a tag that can be used to gain unauthorized access [73].

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 [74]. Software errors in RFID middleware can also be used to compromise the system using standard techniques for example exploiting buffer overruns [75]. 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.

5. Enabling advanced RFID capabilities

In this section we will look at enhancements of basic RFID technology that are expected to play a significant role in the medium term. These developments have considerable performance and cost-reduction implications, and may allow the application of RFID in new domains and applications.
5.1. Materials and packaging

One of the remaining obstacles for the general adoption of RFID is the cost of the individual tag. For higher-value objects, item-level tagging is cost efficient and has already been introduced to full-scale applications. But extensive item-level tagging of objects and metropolitan-scale tagging of locations will not become feasible unless the cost of tags falls below the tipping-point of a few pennies per item. This reduction requires innovation in production engineering and materials. In the shorter term, the largest gains are expected from two techniques that are now entering maturity: printable electronics and mass tag-assembly processes.

The main ingredient for printable electronics is conductive ink, which contains powdered silver and carbon. Conductive ink also has capacitive coupling capability and can be used to print tag antennas on a variety of materials including paper and self-adhesive thin film. A large proportion of the current generation of UHF EPC Gen2 tags carry conductive ink antennas. Savings due to the use of conductive ink are due to reduced material and manufacturing costs. This is a consequence of the fact that it is only necessary to apply the exact amount of ink to print the antenna, which is far more efficient than the construction of a metal antenna using copper or aluminum. It is also a simpler process since a printer similar to standard ink-jet technology can be used to create the antennas.

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

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

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

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

5.2. Augmented RFID sensing

RFID is effective in identifying the presence of tagged entities but cannot provide any information about their situation, for example it cannot measure the temperature of their immediate environment. There are many applications that would benefit from additional information about the setting of an identification event. For example, commodities must often be temperature-controlled, as is the case of the so-called cold supply chain. Temperature sensors embedded in RFID tags used in this setting could be used to indicate whether a certain threshold has been exceeded thus improving food safety.

Unfortunately, even very simple temperature sensors currently require some form of battery to operate, and although such augmented active RFID tags have been readily available for some time, this is not the case for passive tags. Nevertheless, recent proposals have identified ways in which temperature sensors can operate using passive tag technology [76], though actual prototypes are not yet in production. Other types of sensors based on simple micro-fluidic binary devices are also in development and in the near future it is expected that they could be used to identify pressure levels and simple forms of faults in materials [107].

5.3. Surface acoustic wave technology

In addition to the LF, HF, and UHF frequencies, there are also a small number of tags that use microwaves for communication. These tags usually operate in the 2.45 GHz band, which is also available to industrial, scientific and medical applications. This frequency range has distinct advantages in that the antenna of the tag is very short and as a consequence it is possible to build tags that have very small footprint and can thus be embedded into a wider range of artifacts. This frequency also offers the opportunity to support higher data transmission rates and thus higher-performance systems.

However, microwaves also have limitations, in particular the fact that metals and liquids absorb the majority of the transmitted energy and hence severely affect the operation of the system. In practice, this means that the majority of RFID
systems in production using this band are electronic article surveillance (EAS) tags, which are mostly used on fabrics. Similar to UHF RFID, microwave systems use backscatter for communication and, due to various performance considerations, they are often semi-passive rather than fully passive tags. Semi-passive tags carry a battery which is used only to provide energy to the chip but not for communication.

Recently there has been a lot of interest in so-called surface acoustic wave (SAW) technology, which uses the piezoelectric effect to produce fully passive tags in the microwave range. In this case, energy is not harvested from the reader signal directly through a coupling effect but generated from ionic crystals embedded in the tag. The component of a SAW tag that enables this functionality is the interdigital transducer (IDT), a dual comb-like electrode structure that can convert microwaves into SAWs and vice versa. The tag also contains several standard electrodes arranged into a unique pattern specific to each tag (see Fig. 5).

When a tag receives a microwave pulse generated by a reader, its antenna captures the wave and supplies it to the IDT. The IDT in turn converts the electromagnetic wave into a SAW at the same frequency, which is transmitted through the tag body. A small proportion of the SAW is reflected by the electrodes, in a way that is characteristic of their size and arrangement. The reflection is captured by the IDT, which converts it back into microwave that is transmitted by the antenna. The resulting transmission is received by the reader, which detects the reflection cross-section. Effectively, the spatial characteristics of the electrode layout directly define the tag’s transmission pattern, and specific configurations can be used to encode sequences of binary digits.

One notable characteristic of this approach is that due to the overhead incurred by the involvement of the IDT and the fact that SAWs travel at much slower speeds through the tag material than microwaves in the air, the response from the tag has considerable lag—of the order of milliseconds. This fact can be exploited to address the main problem of effective operation of RFID in these frequencies that is, noise generated by the reflection of microwaves on metal surfaces in the environment. Such reflections, even those due to surfaces located hundreds of meters away, would return to the reader much faster than the response of the tag and thus can be clearly identified and isolated from its transmission.

The amount of data that can be encoded in a SAW tag depends on the number of electrodes that are embedded in it and the spatial density with which they can be placed next to each other. This number is relatively low for the current generation of SAW technology, but it is expected to increase in the coming years. Another practical limitation of SAW is that due to regulatory restrictions on transmissions in the microwave band, the pulse generated by the reader, and thus the effective communication range of the system, is restricted and in many cases it would only be a few meters. More details on SAW and the details of its principle of operation can be found in [15].

5.4.Writable tags

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

One such application brings together RFID and robotics in a system that adopts a mode of operation inspired by ant colonies [77]. Robotic agents roam space and use RFID tags to exchange messages and notify each other about the discoveries they have made, thus enabling a persistent, location-specific, delay-tolerant mode of communication. Using standard bio-inspired mathematical techniques, the robots can achieve shared objectives by working collectively in a decentralized manner that would have been impossible without the use of significant infrastructure dedicated to coordination. A similar exchange of messages using tags to hold the information is of course also possible between humans, though this mode of communication has yet to be explored to any depth [78].

Another role for RFID enabled by the availability of writable tags is as the transport layer for routing internet traffic. This possibility has been explored in [79], which proposes the use of spare capacity typically available in access-control tags to
transport IP data between readers. The concept and the techniques proposed are very similar to those used in ad-hoc and delay-tolerant networks. Although obviously not suitable for general-purpose networking, this approach can provide an appropriate mechanism for the distribution of new software and policies to access control readers thus removing the need for the installation of network infrastructure dedicated to this task.

This system views RFID tags solely as a transport mechanism, and in most cases the communication endpoint would not be a tag. A proposal for a system that supports addressing specific tags directly from the internet is presented in [80] as part of an object-tracking solution. This proposal introduces a gateway device that is responsible for the translation of IP to RFID network traffic including specifications for the encoding of IPv6 packets into RFID transmissions to the tags. It also introduces a discovery service that translates IP addresses into RFID entity identifiers in a manner akin to the operation of home agents described in the Mobile IP specification.

5.5. Localization

RFID tags can and have been used for proximity-based location sensing. The basic method is to employ tags fixed at well-known locations, so that a reader within reading range of a tag would be assumed to be co-located—the accuracy of the location estimate would of course depend on the particular RFID technology used and its effective read range. This technique has been used by many systems, notably [81] reports on a metropolitan scale deployment of uID technology as a means to provide location-aware adaptation for a tour guide.

However, such proximity-based location sensing does not provide particularly high-accuracy and recent research studies have attempted to improve on the performance of this approach. One approach advocated by [82] is to use triangulation using signal strength measurements. This approach has been relatively successful in other types of wireless communication systems in improving location estimation compared to in-cell techniques in particular. This technique has been used to good effect especially for WLANs with several systems commercially available. The accuracy of such a system depends on the precision with which it is possible to estimate the distance between a tag and a reader by examining signal attenuation. However, passive RFID depends on the reader to provide the carrier wave for all communication, which means that tag responses are extremely low-power. As a consequence tag signals are very sensitive to interference and system performance suffers overall. The results reported in [82] seem to indicate that this approach does not produce significantly improved approximations compared to simple proximity sensing, while at the same time requires considerably more readers, and thus it does not appear viable.

Another alternative to improve location estimation is through the use of proximity to many different tags at a time. This scenario implies the wider availability of RFID tagging so that at any one time a reader would be within range to many tags. This approach can be implemented in different ways, for example [83] uses a mobile robot and cross-calibration using WLAN signal strength measurements to improve the accuracy of both systems. However, the viability of this approach depends heavily on the density of tags and despite early promising results the method requires further validation.

Finally, RFID can be used in combination with vision technology to improve localization, for example [70] adopts this approach within a mobile robotic system to improve unknown terrain mapping. Indeed, the so-called simultaneous localization and mapping techniques (SLAM), which are commonly used in robotics and augmented reality systems, can benefit significantly from the availability of predetermined reference points – or markers – from which the SLAM algorithm can extrapolate more accurate localization information. RFID tags appear to be well-suited to become wireless markers and for this reason there is increasing interest in their use in this role.

6. Prototyping pervasive computing applications with RFID

Despite its 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 experiments to support automatic-identification for different types of tagged entities. In this section we consider some of these applications and highlight the lessons learnt and the challenges involved in their implementation.

6.1. Universal item-level tagging

Due to the efforts of ISO, EPCglobal, uID and many others, RFID can today address the main technical challenges of universal item-level tagging. Moreover, the rapid implementation of RFID technology at the container level in the supply chain and the resulting economies of scale, bring the cost of tags closer to the crossover point where tagging individual items also becomes financially viable. Though it is not uncommon for some higher-value products to be tagged, from a pervasive computing perspective interesting applications would be enabled only when lower-value objects can also be automatically-identified.

Applications that exploit such universal item-level tagging have already been anticipated in retail, for example [5,84,106] consider specific consumer-facing services. It is of course natural that the first such applications would be in this sector due to its intimate involvement in the supply chain. This is a predictable application of RFID and in previous sections we have
referred to work carried out in the spirit of u-commerce. In this context, a central question appears to be that of trust that is, whether the technology can be deployed in such a way that it is both worthwhile for the retailer and respectful of the privacy of the consumer [43]. Developing strategies that can achieve a balance between these conflicting objectives appears to be an intractable goal at present. Its successful resolution would require a far better understanding of the various aspects of the economics of privacy [47] as well as how consumers perceive pervasive computing services, especially with regard to the inevitable loss of privacy [46] and the availability of effective controls over the technology [44].

Nevertheless, the most interesting opportunities enabled by universal item-level tagging are related to new ways of interacting with information through the use of tangible interfaces. Tangibles allow interaction with a digital system through direct manipulation of physical artifacts that can be enabled by the use of RFID. For example, in the case of the periscope, a device that allows children to explore the Ambient Wood [85], Petri dishes containing a variety of micro-organisms were tagged. Moving a dish close to the periscope would trigger the display of information related to the specific species and its habitat. It is thus possible to interact simultaneously both with the actual organism and its recorded data.

RFID-based tangible interfaces have also been explored in [86], where tags were used to augment toys employed in educational games. The emphasis of this work is on the ability of tangibles – unlike traditional computer games – to be shared between several children, a capability which allows for simultaneous and co-located interactions and furthermore it does not hinder social interaction between members of the play-group.

6.2. Data management

Item-level tagging can produce massive data sets especially in the case of supply-chain applications, where products move regularly and in large quantities. Although there are several issues related to the management of this data [87], two in particular stand out: (i) processing raw RFID observations and converting them into application-level events, and (ii) supporting efficient querying of RFID information services across federated sources.

Relating to the first problem, the current generation of middleware often employ a sliding window approach in applying temporal smoothing filters to the incoming observation streams. This approach is limited in that it attempts at the same time to ensure completeness and time-accurate capture of tag dynamics, objectives that are in direct conflict with each other. As a result, the choice of smaller or larger aggregation windows is not an adequate solution to the smoothing problem, as it skews the results toward one or the other direction. An alternative approach is proposed in [88], whereby incoming streams are treated as a statistical sampling of tags and employ advanced statistics to drive adaptive smoothing. Specifically, binomial sampling and \( \pi \)-estimators are proposed as the driver of an adaptive automatic selection of the appropriate window size, which produces superior results. This technique is encapsulated in the so-called SMURF mechanism, which also provides a fairly complete specification for the definition of complex event queries.

Although the expressivity of SMURF is superior to other RFID middleware, it is still limited in that it supports only one level of mapping, from observations to application-level events. This has been considerably extended in [89] with event specifications that support flexible use of negation, parameterized predicates, and sliding windows, while maintaining a relatively compact language specification. The proposed approach appears to indicate that it is possible to implement this technique with good performance compared with high-performance dataflow processors. This is promising work but requires to be extended with the facility to define hierarchies of complex or composite events of ever-increasing complexity in the same way as real-world semantics compose complex hierarchies of meaning.

The second of the two main RFID data management problems is far harder and little progress have been made towards its solution [28]. There are several aspects of the problem that contribute towards its complexity in addition to the large quantities of data involved. Firstly, the fact that RFID lineage tracing involves a chain of distinct data custodians which have few incentives to share a complete data trace. As a result, it is unlikely that established data warehousing techniques would work well in this context as they involve the use of materialized views which are continuously updated from the underlying data sources. Second, individual custodians would employ distinct database schemata which must be integrated, a task that is tractable for example using the Both-as-View framework, but adds considerable overhead to the task. A further complication is due to the so-called containment relationships whereby assemblies of individual entities are replaced by a single identifier at a certain point in time. Such relationships are time-dependent also in that they expire at a future time and thus queries about constituent entities must be expanded to an extended spatiotemporal scope.

6.3. Application-level programming models

In Section 3, we discussed the system components that are required to create and program complete RFID infrastructures. 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 [32]. 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 [17] is to introduce domain-

\[ \text{RFID is only one of several alternative technologies that can be used for the construction of tangible interfaces.} \]
specific models and associated frameworks that provide system-level abstractions and use those to develop applications. An alternative is proposed in [30] 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 application programming interfaces specifically to support location tracking at a high level in a manner similar to the approach adopted in cellular networks is proposed in [90].

However, such high-level approaches do not always cater well to the peculiarities of specific RFID systems but are forced to abstract to the lowest common denominator thus missing opportunities for improved performance. One example of this is provided by the different capacities of near and far field tags: assuming that only the lower memory capacity of UHF is available leads to failure to use local storage on HF tags that can very considerably improve system robustness [5].

6.4. Context awareness

RFID can also be used to support context awareness by identifying proximity relationships between identifiable entities [91]. For example, when specific objects are located close to a user, the system can automatically fetch the information required to adapt and tailor interactions to the particular situation [92–94]. Typical examples of such adaptation using RFID include medical facilities for example, hospital beds that fit the needs of particular patients with particular ailments [95]; 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 [96]; 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 [55].

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 [97] 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.

6.5. 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 the object itself as a tangible interface to support collaborative work [98].

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 for instance. In the specific case study reported in [78], it is not the object itself that is being manipulated but rather the RFID reader embedded in an NFC-enabled mobile phone is used to facilitate this exchange of instructions.

Another way that RFID can enable social interactions is in coordination with public displays [99]. 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 [100]. Such ideas have been further explored to bring together combined physical and digital co-incidence that is, to investigate concurrent proximity of authorship and location again within an academic conference environment [101]. In this case, public displays are also used to explore common links to the same social groups as the enabler of further interaction between conference attendees.

6.6. Proactive privacy protection

In Section 4 we looked at the multiple and complex issues of privacy protection and secure operation of RFID-based systems. Although there is currently considerable effort invested in the development of new mechanisms and techniques to address these issues, and in the long run it is likely that RFID would be a trustworthy technology, there are more than adequate reasons to justify a more proactive approach in the short-term. Moreover, a number of these methods place control
of the technology firmly on the side of the operator and it is difficult for the user to confirm if a system complies with its stated behavior or not.

To address the issues inherent in relying on public RFID readers to enforce appropriate and acceptable privacy policies, users can choose to protect their privacy by operating privacy-preserving devices proactively. An example of such a device is the so-called Watchdog Tag [102] which can monitor read attempts and collect reader information. Similarly, the RFID Enhancer Proxy [103] mediates between readers and consumer tags to enforce privacy-preserving communication and the Blocker Tag employs active jamming to completely prevent communication [104]. These devices are characteristic examples of an increasing variety of programmable computers that can observe reader and tag communications and interject a variety of responses or interference, so as to disrupt the normal operation of a system, thus restricting the readability of tags to specific peers only.

Other approaches aim to completely prevent tags from being read. For example a common technique achieves tag-cloaking through the use of a Faraday cage, which can be easily constructed around the tag using by for example duct tape or tin foil. A somewhat more sophisticated way to achieve the same effect, has been employed in the more recent versions of the US e-passport whereby a thin wire net has been embedded in the cover of the document. Alternatively, tags can incorporate low-level circuitry that can approximately calculate the distance to a reader, and choose to only respond to requests from very proximate readers which can be inspected visually [70]. With sophisticated field-programmable gate arrays increasingly used for the development of more intelligent systems it is not improbable that in the near future such devices may come to common use.

7. Conclusions

This survey has reviewed developments towards establishing RFID as the cost-effective technical solution for the development of first-generation, open, shared, global pervasive computing infrastructures. RFID is an effective automatic identification technology for a variety of entities including physical, manufactured and handmade objects; humans and other species; locations; and increasingly media content and mobile services. Because of the limited capabilities 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 commercial applications 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 against more recent technologies.

Further development of RFID technology depends on successfully addressing two fundamental concerns: economic and environmental sustainability, and acceptable privacy protection. Despite the massive cost reductions 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 – which seems to be unlikely in the sort term due to market, intellectual property and manufacturing reasons. Universal deployment of RFID also depends on robust business cases that satisfy all paries involved and are far from forthcoming at the current state of development.

Perhaps more critically, RFID manufacturers are yet to address 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 addressed effectively before RFID becomes pervasive. Support for sustainable practices should be a central consideration in the immediate future.

Universal tagging also has very considerable privacy implications that are far from being adequately addressed by any of the proposals reviewed in this survey. It is unlikely that these concerns can be addressed by new technology alone and to achieve this goal it will be necessary to make appropriate provisions within the 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

George Roussos holds a B.Sc. in Mathematics from University of Athens, Greece, an M.Sc. in Numerical Analysis and Computing from University of Manchester Institute of Science and Technology, UK, and a Ph.D. from Imperial College of Science Technology and Medicine, University of London, UK. Before joining Birkbeck College as a lecturer he worked as the Research and Development Manager for a multinational information technology corporation in Athens, Greece, where he was responsible for the strategic development of new IT products in the areas of knowledge management and mobile internet; as an Internet security officer for the Ministry of Defence, Athens, where he designed the Hellenic armed forces Internet exchange and domain name systems; and as an Internet security officer for a multinational information technology company in Athens, Greece, where he was responsible for the strategic development of new IT products in the areas of knowledge management and mobile internet; as an Internet security officer for the Ministry of Defence, Athens, where he designed the Hellenic armed forces Internet exchange and domain name systems; and as a research fellow for Imperial College, London, where he conducted research in distributed systems. He is currently investigating the effects of social activity on system architectures, and exploring mechanisms to support navigation and findability. He is a member of the ACM, SIGMOBILE, the IEEE, the IEEE Communications and the IEEE Computer Society.
Vassilis Kostakos holds a Ph.D. in Human-Computer Interaction. He is an Assistant Professor at the University of Madeira, and an Adjunct Assistant Professor at Carnegie Mellon University. Previously he was a Research Associate at the University of Bath. He designs Pervasive Systems. Despite his ups and downs, he still plays basketball. He now lives on a tiny island in the Atlantic. Last but not least, he is a Maniot and a downshifter.