Scanning with a Purpose – Supporting the Fair Information Principles in RFID Protocols

Christian Floerkemeier, Roland Schneider, Marc Langheinrich
Institute for Pervasive Computing
ETH Zurich, Switzerland
floerkem@inf.ethz.ch, schneider_roland@student.ethz.ch, langhein@inf.ethz.ch

ABSTRACT

Today's RFID protocols that govern the communication between RFID readers and tags are solely optimized for performance, but fail to address consumer privacy concerns by supporting the fair information practices appropriately. In this paper we propose a feature set that future privacy-aware RFID protocols should include in order to support the fair information principles at the lowest possible level – the air interface between readers and tags – and demonstrate that the performance impact of such an extension would be within acceptable limits. We also outline how this feature set would allow consumer interest groups and privacy-concerned individuals to judge whether an RFID reader deployment complies with the corresponding regulations through the use of a watchdog tag.

INTRODUCTION

When Mark Weiser envisioned computing capabilities everywhere, embedded in the environment in such a way that they can be used without noticing them, he also acknowledged that the invisible nature of the computing devices will make it difficult to know what is controlling what, what is connected to what, and where information is flowing [18]. The intended deployment of RFID-based tracking solutions in today's retail environments epitomizes for many the dangers of such an Orwellian future: Unnoticed by consumers, embedded microchips in our personal devices, clothes, and groceries can unknowingly be triggered to reply with their ID and other information, potentially allowing for a fine-grained yet unobtrusive surveillance mechanism that would pervade large parts of our lives. While industry standard bodies still largely focus on optimizing the communication between RFID readers and tags for speed and cost at the expense of privacy, consumer interest groups consequently advocate the complete ban of RFID tags in the public part of stores [4]. Although the latter approach will naturally protect the privacy of the individual, it falls short of an optimal solution even from a consumer standpoint, since it is not just the retail store that can benefit from the use of RFID tags, but also the consumer. The magic medicine cabinet [17], the magic wardrobe [10], and the often-cited smart fridge are just some of the consumer applications that would benefit from post-point-of-sales item-level RFID tagging.

In this paper we argue for a middle ground, inspired by our everyday lives where we rarely encounter all-or-nothing tradeoffs, but rather engage in meaningful exchanges that conditionally lead us to disclose parts of our personal data to service providers in return for more or less tangible benefits. By incorporating the basic principles of the widely accepted fair information practices at the reader-to-tag protocol level, RFID-system operators will be able to deploy readers that only collect tag data relevant to the actual application, while small personal devices could additionally provide consumers with a detailed look at a reader's owner and its purpose for collecting data, potentially allowing for an explicit consent before any tag information is read out. Future tags might even be able to independently decide whether or not to reply to a reader's query, based on its stated ID, purpose, and target range. Having RFID readers explicitly declare the scope and purpose of the tag data collection, as well as disclosing the identity of their operators, will allow both consumers and regulators to better assess and control the impact of everyday RFID encounters.

The rest of the paper is organized as follows. After briefly restating the fair information principles and their role in today's privacy legislation, we describe some of the most important characteristics of RFID systems and show how the requirements put forth by the fair information principles could be embedded into the reader-to-tag communication of existing RFID standards. We then present an early prototype of a "watchdog" tag, a small personal device that can be used in conjunction with our protocol extensions to further increase the transparency of the identification process. We conclude with a discussion of our approach, giving special regard to its efficiency, as well as outlining future work.

FAIR INFORMATION PRACTICES

The Fair Information Practices (FIP), published by the Organization of Economic Cooperation and Development (OECD) in 1980 [14], are a well established set of guidelines for consumer privacy. They have their roots in a 1973 report of the "United States Department for Health, Education, and Welfare (HEW)" and were drawn up by the OECD to better facilitate the cross-border transfer of customer information as part of trade between its member states. The

eight principles can be summarized as follows:

1. Collection limitation: Data collectors should only collect information that is necessary, and should do so by lawful and fair means, i.e., with the knowledge or consent of the data subject.

2. Data quality: The collected data should be kept up-to-date and stored only as long as it is relevant.

3. Purpose specification: The purpose for which data is collected should be specified (and announced) ahead of the data collection.

4. Use limitation: Personal data should only be used for the stated purpose, except with the data subject's consent or as required by law.

5. Security safeguards: Reasonable security safeguards should protect collected data from unauthorized access, use, modification, or disclosure.

6. Openness: It should be possible for data subjects to learn about the data controller's identity, and how to get in touch with him.

7. Individual participation: Data subjects should be able to query data controllers whether or not their personal information has been stored, and, if possible, challenge (i.e., erase, rectify, or amend) this data.

8. Accountability: Data controllers should be accountable for complying with these principles.

The FIP form the basis for many of today's privacy laws, such as the EU Directive 95/46/EC [7], which provides the framework for the national privacy laws of all EU-member states. For example, article 6 of the Directive requires data collectors to collect only as much information as necessary (also called the *proportionality principle* or the principle of *data minimization*) while article 7 requires them to obtain the unambiguous consent of the data subject before the collection.

It is undisputed that the act of reading out one or more RFID tags constitutes a data collection, meaning that existing privacy laws also apply to the communication between tags and their readers. This has also been recently pointed out by the international community of data protection and privacy commissioners [1]. At the outset, this would mean that RFID readers would need to be openly announced with the help of public signs and placards explaining the purpose and extend of the data collection, as well as the identity of the data collector. While adequate from a legal point of view, presenting the necessary information in such a way easily suffers from being ignored by the consumer, as the ubiquitous privacy policy links on today's Web sites have demonstrated. This is because of two important drawbacks such an out-of-channel solution has: Firstly, data subjects need to actively seek out such information that might otherwise be easily overlooked. Secondly, even when accessible, reading and understanding this information puts an added burden on the consumer, as it is often written in dense legal prose.

Principle	Support
(1a) collection limitation	through selection mask
(1b) consent	with watchdog tag (optional)
(2) data quality	out of scope (use privacy-aware DB)
(3) purpose specification	through purpose declaration
(4) use limitation	out of scope (use privacy-aware DB)
(5) security safeguards	encryption (future work)
(6) openness	through reader and policy ID
(7) participation	out of scope (use privacy-aware DB)
(8) accountability	through reader and policy ID

Table 1. *Support for the FIP in our reader-to-tag air interface.* About half of the principles can be embedded directly at the protocol level.

On the Web, the Platform for Privacy Preferences Project (P3P) aims at alleviating these two drawbacks [5]. Developed under the auspices of the World Wide Web Consortium (W3C), P3P integrates machine readable privacy policies into the browser-to-server protocol, thus allowing the user's Web browser to automatically read the privacy policy of a Web site, compare it with the user's preferences, and subsequently take action on behalf of the consumer (e.g., facilitating or preventing a transfer of personal data, or advising the user in an easily understandable manner). Our goal is to implement a similar mechanism into the protocol between RFID tags and their readers, in order to lessen the burden on the consumer by having her tags (and optionally a personal mobile device carried with her) read and process privacy related information autonomously and support her in this task.

Some of these principles, such as individual participation or data quality, will need support primarily in the storage backend, for example with the help of privacy-aware databases [2, 12]. However, the majority of the principles could be supported directly at the point of data collection, i.e., when the reader interrogates the tags. Table 1 lists the level of technical support for the FIP that our extended reader-to-tag air interface offers. Obviously, most of this support can also be achieved through non-technical means, e.g., a notice about tag-reading taking place could also be simply announced through an easily noticeable sign. However, by incorporating such principles directly into the underlying protocol, both consumers and data collectors can more easily follow them, thus strengthening existing legal protection by providing the means to verify and thus enforce corresponding regulations.

RFID PRIMER

Before describing our planned extensions to existing RFID standards in detail, we give a brief overview on the functioning of an RFID system. RFID systems are composed

of RFID tags, which are attached to the objects to be identified and an RFID reader, which reads from and possibly also writes to the tags. RFID tags consist of a coupling element and a microchip that stores, among other things, data including a tag identification number. The reader forms the radio interface to the tags and typically features some internal storage and processing power in order to provide a high level interface to a host computer system to transmit the captured tag data. Since RFID tags usually do not possess their own power supply, the reader supplies the tags with power through the coupling unit along with data and clock pulses.

While all RFID systems are made up of these two components, a wide variety of different RFID systems exist that address the requirements of individual application scenarios. Finkenzeller [8] provides a comprehensive classification of the various RFID systems commercially available. An overview of RFID systems that also addresses their privacy implications is available in [15]. For the purpose of this paper, the important differentiation features are the memory organization, read range and the methods that an RFID reader employs to detect multiple tags in its read range, the anti-collision algorithm.

In order to identify an individual tag in a group, tags usually store at least a unique ID (UID). One can generally distinguish the EPC approach [6], promoted by the Auto-ID Center (now EPCglobal), where a tag only carries a unique ID, but information about manufacturer and product type are encoded in this identifier, and the approach, where the memory is partitioned into a random serial number identifying the tag and additional memory to store information about the object to which the tag is attached.

Under ideal conditions, modern RFID systems in the UHF band (860-960 MHz) can achieve a read range of up to seven meters, though in reality the range is usually less. For HF and LF-based systems (13.56 MHz and 135 kHz, respectively), this comes down to no more than one or two meters, unless large tag antennas are used. While read range

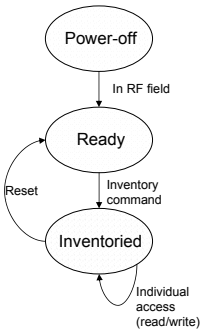

Figure 1. ***Simplified tag state transition diagram.*** **As soon as tags enter the reader's RF field, they move into the "ready" state and reply to the reader's "inventory" command. Once the reader has inventoried tags in its read range, it can access them individually.**

Protocol extension	Init round all	SUID flag	Round size	CRC-5
1 bit	6 bits	1 bit	3 bits	5 bits

Figure 2. ***The inventory command,*** `Init_round_all`**, as specified in ISO 18000-6 Type A. The command frame consists of a field that indicates the number of time slots that are available for a reply (round size), various flags, and a cyclic redundancy check (CRC) to detect transmission errors.**

issues do not play any role in our protocol extension, it is nevertheless an important parameter for any privacy related discussion of RFID systems, as privacy concerns associated with the invisible nature of RFID increase with the achievable read range of an RFID system.

Once the tag is within the read range of an RFID reader, the tag is powered and is ready to communicate with the reader (cf. figure 1). When multiple tags respond simultaneously to a request from the reader, their signals can interfere with each other, resulting in a failed transmission. In order to inventory all tags within the read range, an anti-collision algorithm that controls access to the shared radio channel is employed by the reader.

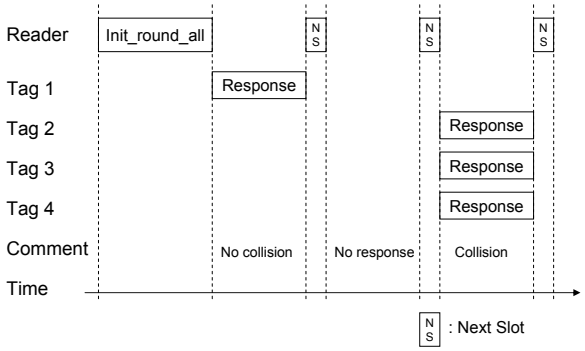

Figure 3. ***The inventory process,*** **as specified in ISO 18000-6 Type A. The reader initiates a round of tag replies by issuing an** `Init_round_all` **command. Energized tags respond by selecting one of the available time slots at random to transmit their ID .**

Figures 2 and 3 show examples of the inventory command (`Init_round_all`) and process, respectively, as defined in the ISO-Standard 18000 Part 6 Type A [9] (which is the standard we are basing our protocol extension on). This standard uses a probabilistic anti-collision protocol scheme, meaning that tags respond at randomly generated times, e.g., based on the Aloha scheme [8]. Deterministic algorithms, in contrast, use typically a binary tree-walking scheme to traverse the set of all possible tag numbers.

SUPPORTING THE FIP IN EXISTING RFID STANDARDS

In this section we outline how existing RFID standards can be modified to satisfy the principles of *collection limitation*, *purpose specification*, *openness* and *accountability*. The extensions are illustrated using the

ISO-Standard 18000 Part 6 Type A as an example, though they can equally well be applied to other RFID standards.

Openness through reader and policy identification

None of today's RFID standards allow tags to identify the reader they are communicating with. The anonymous broadcast by the reader is certainly desirable from a performance point of view, since the reader's goal is to identify as many tags by their UID as possible in a certain period of time. The transmission of any additional data such as the identification number of the reader will thus reduce the speed at which tags can be detected. Without knowledge about the device that is collecting data, it is, however, impossible to satisfy the principles of *openness* and *accountability*. In order to address these FIP requirements also at the air interface, we include a unique reader policy ID (RPID) into the reader's inventory command, which both uniquely identifies the reader and its operator, as well as the policy in place. Having an explicit reference to the policy allows us to provide additional information about a policy over a separate channel and also facilitates dispute resolution by allowing customers to directly identify the policy used.

The RPID itself is encoded in a three-tier format, specifying the following three fields: the data collector ID, the policy ID, and the reader ID (cf. figure 4). With this structure, our solution follows closely the well-established EPC format and its general identifier encoding (GID-96) [6]. Even though we are not identifying products, but data collectors and their policies, this symmetry could potentially benefit the administration of the data collector IDs, as their identical format would allow data collectors to reuse their existing "General Manager Number" [6] of their EPCs (data collectors that do not already have such a number could acquire it in a similar fashion as they would for obtaining an EPC identifier). Moreover, the existing ONS architecture [13] that provides a look-up functionality for captured EPCs could transparently be used to resolve our reader policy references as well.

The policy ID follows directly after the data collector ID, giving data collectors a 24 bit value for identifying policies. Data collectors are free to substructure this value in any way they like, as they can do for the last value, the actual reader device ID, which comprises 36 bits. Useful substructures would be a division across country, region, city, or store, thus simplifying both policy publishing and reader localization from this ID. In our prototype, we use the policy ID to acquire more detailed policy information over wireless LAN, while the reader ID is resolved to its designated approximate location, in order to allow the (manual) detection of reader ID spoofs (e.g., a reader of a retail outlet on 5th Ave. suddenly appearing ten blocks south of this address).

Figure 4 shows a summary of our reader and policy identification code, and illustrates its usage again using the inventory command of the ISO 18000 Part 6 Type A protocol as an example.

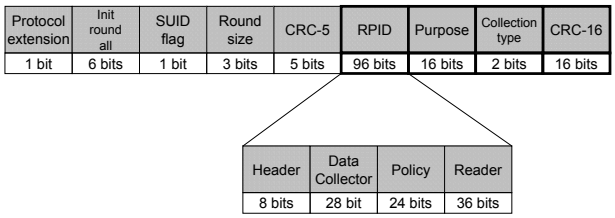

Figure 4. ***The modified inventory command,*** `Init_round_all`**, of ISO 18000-6 Type A featuring an additional field for the reader policy identifier, the purpose declaration, collection type, and an additional checksum (CRC).**

Purpose specification in the inventory command

The FIP require that the purpose for which personal data is collected should be specified no later than at the time of data collection. P3P addresses this issue by providing a list of 12 abstract purpose types that describe why data is being collected relevant to the specific web site that the policy describes [5]. Although RFID needs to be treated slightly different in the sense that in most cases the user will be unaware of the data collection taking place, as well as of the actual data being collected, many of the P3P purpose definitions can be equally well applied to the RFID domain.

Contrary to Web services, however, some purposes such as *admin* or *current* are much more difficult to assess in an RFID environment. For example, the current purpose is usually implicitly defined by the Web interaction the user is currently experiencing, e.g., the shopping cart checkout in a Web shop, while administration is usually defined by keeping Web server log files. In an RFID context, however, many different "current" or "admin" purposes can be envisioned: A smart shelf might issue read commands for inventory purposes (in a supermarket) or for asset tracking (e.g., for multimedia equipment that employees can check out from a central magazine), both of which could be called administrative purposes. "Current" purposes can equally vary, from a payment purpose at a self check-out station, to a repair and return purpose at a customer information station.

Consequently, we have expanded some of the existing P3P purposes while dropping others, in order to better reflect the more implicit interactions present in RFID systems. Table 2 lists the 14 purposes we identified as useful declarations in this context, even though additional purposes might become necessary in the future. Our above list is therefore only an initial suggestion that should be repeatedly validated by real-world prototypes, and subsequently standardized by an appropriate standardization body.

Apart from the "profiling" purpose, all purposes are encoded

as single bit values that can be arbitrarily combined in our 16 bit number, indicating that data are collected for multiple purposes. The profiling purpose uses three bits to encode one of five possible profiling purpose types that are mutually exclusive (see table 3).

For example, a smart shelf application that monitors its contents for out-of-stock warnings, as well as provide data for anonymous in-store movement information (e.g., to see where consumers spend most of their time), would need to declare both the "inventory" and the "pseudo-analysis"-profiling purposes. A corresponding smart shopping cart that would provide customers with shopping suggestions, based on its contents, would declare "pseudo-decision"-profiling. And a self-checkout station that allows customers to wirelessly pay for their goods, while also associating the purchased items with the customer's loyalty card, would consequently declare the "payment," "anti-theft," and "individual-decision"-profiling purposes.

Use limitation through collection types

The principle of RFID reader-to-tag interactions (i.e., readers issuing an inventory command and tags replying with their IDs) makes it difficult to create privacy-friendly monitoring applications even if no identifying tag information needs to be collected as part of the envisioned application. Imagine an RFID system that tries to keep track of the number of people on a certain station platform, in order to avoid overcrowding. Even though RFID tags entering and exiting the area might reply to reader commands with their IDs, the application only needs to keep track of individual tags (e.g., an RFID-based train pass) without having to actually know their specific ID. Additionally, even when identifying information is collected, consumers will typically become much more concerned if this information is not only used locally, but also correlated across multiple readers in order to track an item's (or a person's) movements over time.

To allow data collectors to differentiate between the various collection needs, i.e., whether or not they actually require the serial number of individual tags, or whether they intend to track multiple occurrences of the same tag across different location, we additionally define four distinct collection practices that must be declared as part of a reader's inventory command:

1. *Anonymous Monitoring:* Collecting state information about the items in the vicinity of a particular location, without the need to actual identify tags by their unique serial number. Examples would be simple sensor applications (e.g., an automatic door opener) or counting tasks (e.g., monitoring the number of items in a certain area).
2. *Local Identification:* Tag IDs are collected in order to provide a localized service, e.g., a smart medicine cabinet or smart fridge that monitors its contents. Although unique IDs are collected (e.g., for resolving them to human readable descriptions), the application does not require (nor attempt) the correlation of events across different locations.
3. *Item Tracking:* Collecting information about the location of an item for the purpose of monitoring its movements. Note that this potentially enables tracking people through constellations. However, in order to differentiate between these different intentions, the separate "tracking person" declaration should be used, if people are tracked by the items they carry.
4. *Person Tracking:* Collecting information about the location of a person. Note that although item-level tracking can potentially reveal the location of a person, data collectors will only need to declare this, if they actually collect

Type (Pos)	Description
access control (0)	Tag IDs are scanned for the purpose of access control, e.g., by identifying a pass holder or by authorizing the validity of an access key.
anti-counterfitting (1)	Readers read out data stored on the tags to assert the genuineness of a merchandise.
anti-theft (2)	Readers scan for tags that are attached to items that have not been paid for.
asset management (3)	Contrary to inventory purposes, tags are read to provide a picture of the whereabouts of assets, instead of monitoring changing stock quantities.
contact (4)	Tag contents are read out in order to determine a contact channel to the customer, e.g., a mobile phone number or email address.
current (5)	Tags are read to provide a service that was explicitly desired by the individual, e.g., when placing shopping items on a kiosk in order to calculate totals, or for disabling (killing) tags.
development (6)	This purpose should be used during system testing and development only.
emergency services (7)	The system is monitoring tags in order to provide rescue workers with occupancy information.
inventory (8)	A shelf monitoring its contents, e.g., in order to provide out-of-stock notices to a central system.
legal (9)	Law enforcement or other legal obligations require the system owner to read out tag IDs. Additional information on the legal grounds should be made available to the customer.
payment (10)	The current action involves payment, e.g., at checkout when tag IDs are read for billing purposes.
profiling (11-13)	Data is collected for profiling or ad-hoc personalization. See table 3 for individual values.
repairs and returns (14)	Warranty and manufacturing details are read out in order to facilitate or speed up a repair or return process.
other (15)	None of the above purposes fits. Further information should be accessible, e.g., in form of a sign or explicit contractual agreement.

Table 2. *RFID purposes declarations.* Data collectors can combine 15 different purpose declarations for RFID reader queries.

Type (Bits)	Description
ad-hoc-tailoring (011)	This applies to immediate and anonymous tailoring, e.g., providing shopping recommendations based on the current content of a shopping basket, or suggesting accessories based on the clothing the customer has taken into the dressing rooms.
pseudo-analysis (100)	The collected data are used to learn about the interests or other characteristics of individuals. This may help to reveal the interests of visitors to different areas of a store. For example a store's shelves could be newly arranged based on the collected aggregated data.
pseudo-decision (101)	This information will be used to make customization decisions based on the interests of individuals, without actually identifying them. For example, a shop could suggest items to a customer based on his or her previous visits (without actually identifying that person).
individual-analysis (110)	The data collected is used in combination with identified data of an individual, allowing a profile of a certain customer to be generated. This could help to reveal the interests of visitors based on their age, social situation, or other relevant demographic data. Identification could occur in combination with a consumer or credit card.
individual-decision (111)	The information is used to determine individual preferences and to link them with identified data. This profile allows personalized suggestions, based on the individual's interests collected from previous visits, combined with personal information, e.g., from a consumer loyalty card.

Table 3. ***Profiling purposes.*** **Profiling purposes are mutually exclusive, as profiling types lower in the table (i.e., with higher bit-codes) can potentially include all of the above types.**

RFID tag information for this purpose. It is up to legal frameworks to force data collectors to anonymize item-tracking data so that it cannot be used for person tracking.

Together with a corresponding purpose, collection declarations further facilitate the accurate assessment of any RFID scan event. This does not only help data subjects to better understand the *intentions* behind a data collection, but can also be used to selectively allow tags to remain *anonymous* whenever possible. Anonymous replies are already part of some RFID protocols, e.g., ISO 18000 Part 6 Type A, though the reason for using them is usually, again, efficiency, not data privacy. To detect collisions, a 64 bit or longer unique ID is usually not needed and just decreases the number of individual tags that can be successfully detected per unit of time. The anti-collision routine can thus first use the tag's random short identifier to single it out from the set of present tags, before requesting additional data, which might include the unique, but static serial number. We propose that this kind of an anti-collision protocol could become the default, whenever "anonymous monitoring" intentions are declared, thus explicitly providing tag anonymity and unlinkability.

Even without any specific support in the tags themselves: declaring, say, "local identification" would still provide the data subject with the additional level of assurance that her movements would not be tracked across different locations (though this might not preclude the keeping of log files that could be later combined, e.g., as part of a criminal investigation). Again, none of these declarations are a proof that the data collector stating them is actually following them. However, as with the purpose declarations, any explicit privacy policy declaration provides a lever to threaten wrongdoers with legal actions – just as it is the case with today's printed policies.

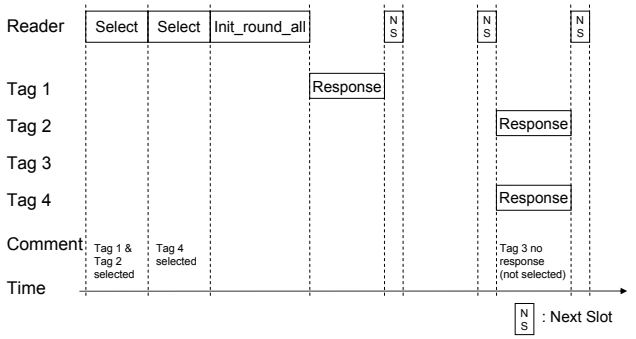

Figure 5. ***The modified inventory process.*** **The reader first selects a tag population, before initiating a round of tag replies by issuing the modified `Init_round_all` command. Previously selected tags (Tag 1, 2 and 4) respond in a randomly chosen slot.**

Keeping with the examples from the previous section, the smart shelf tracking inventory and performing anonymous movement analysis of customers within the store would thus need to declare a collection practice of "person tracking", even though these traces are anonymous (pseudo-analysis). The smart shopping cart would use "local identification", as it would use the identity of the items in the cart to locally decide what other products to suggest to the user. Note that it does not matter whether this decision process is actually done on the shopping cart itself or wirelessly via a remote system, as long as the tracked tags are not correlated to other carts or shelves. A smart check-out station would need to declare "person tracking" again, in case a consumer loyalty card is scanned at the point of sale.

Collection limitation by appropriate tag selection

The first of the fair information principles requires data collectors to limit the amount of data they collect to what is absolutely necessary (today, the EU directive makes this a legal requirement in most European countries). Consequently, rather than asking *any* tag present to respond to a reader query and then filtering out the tags of interest on the application level, we want readers to limit their initial query to target only relevant tags in the first place, thus realizing the collection limitation principle already at the protocol level.

As an example of how this would work in practice, let us look at the frequently considered usage scenario of a super-

market smart shelf, whose purpose is to detect whether it is stocked with sufficient supplies of a particular item. Instead of issuing indiscriminate read commands, which might also pick up tags in the clothing of nearby shoppers, the shelf reader will target only tags of products stacked on the shelf, such as a particular brand of razor blades. Optionally, the shelf reader could occasionally run a separate request that targets *all* of the supermarket's products in order to detect misplaced items.

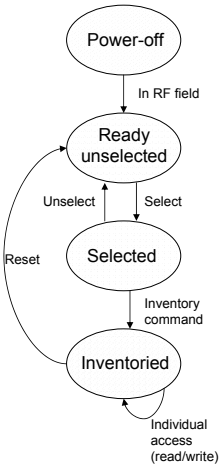

Figure 6. ***Modified tag state transition diagram.*** **After getting energized the tag enters the ready unselected state. The tag moves into the selected state, once it receives a matching "select" command. Only selected tags will respond to an "inventory" command by the reader.**

To implement this functionality in our reader-to-tag-protocol, we make use of a similar mechanism that is typically used to singularize a particular tag from a set of tags in range (e.g., the Group-Select and Group-Unselect commands in ISO 18000 Part 6 Type B). However, instead of using a selection mask to facilitate and potentially speed up the inventory process, we are using selection masks to restrict tag ID collection by the reader to relevant tags for privacy reasons.

Once tags appear in the range of a reader and get energized, they initially begin in an "unselected" state. Unselected tags will need to be explicitly selected before replying to any inventory, read or write command from the reader. Tags become selected only after receiving a select mask that matches their data in memory. Readers thus begin any command cycle with one or more select commands that first determine the tag population that is the target of the query (see figure 5). Once selected tags have been "inventoried", readers can issue actual access commands (see figure 6).

The `Select` command contains the following parameters (as shown in figure 7):

- *Pointer, length, and mask (PLM).* Pointer and length address a certain tag memory range. The mask, which must be "length" Bits long, contains a bit string that the tag must compare against the contents of the specified memory location.
- *Selection type.* The selection type indicates whether tags that match the PLM should enter the selected state or return to the ready, but unselected state.

Note that an appropriate selection of tags that fulfills the requirement of the collection limitation principle will only be feasible if the tag IDs follow a known structure that allows for a certain grouping, e.g., a common prefix for a certain product from a particular manufacturer. This is the case in the currently favoured EPC system, where ID ranges are grouped by manufacturer ID and product type. If there is no such information encoded in the identifier, it needs to be available in the remaining portion of the tag memory and accessible during the selection process, as random tag IDs would be difficult to efficiently select.

Protocol extension	Select	State flag	Pointer	Mask length	Mask value	CRC16
1 bit	6 bit	1 bit	8 bits	8 bits	variable	16 bits

Figure 7. ***The new*** `Select` ***command*** **enables readers to select a subset of tags within the read range. The state flag indicates whether a tag with a matching mask should enter or leave the selected state.**

In the following section, we show how the feature set outlined in this section – i.e., the reader policy ID, the purpose and collection type declaration, and the selection mask – can significantly increase the transparency in today's RFID scenarios.

WATCHDOG TAG

In order to make full use of the additional information now present in the reader protocol, we use a so-called "watchdog tag" to provide transparency to the otherwise invisible tag detection process. Simply speaking, the watchdog tag is a sophisticated version of an ordinary tag, as it features an additional battery, a small screen, and potentially even a long-range communication channel. The watchdog tag's main task is to decode the commands transmitted by a reader, and make them available on the screen of the device for inspection by the user, as shown in figure 8, or to log all data transfers and provide consumers with detailed summaries whenever needed. While the watchdog tag could be carried by the user as a separate device, its functionality could also be integrated into a mobile phone, allowing it to leverage the existing display, battery, memory capacity and long-range communication features of the phone.

Without the privacy features in the protocol, the watchdog tag would only be able to inform the user that some anonymous reader is scanning for tags in a certain vicinity. Due to the privacy features introduced in the RFID protocol, this notice can now include the operator's ID, the purpose and type

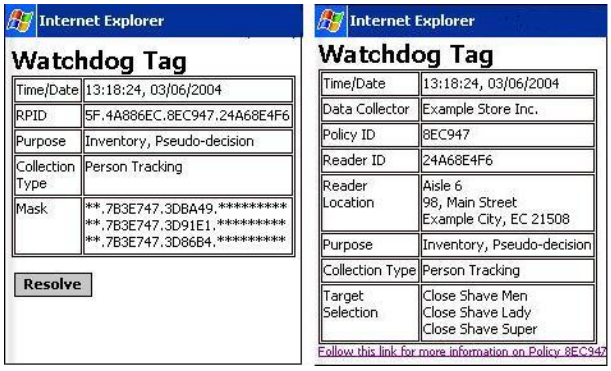

Figure 8. ***Watchdog Tag screen shots.*** **The screen shot on the left shows data collected by the watchdog tag over the RFID channel. If a separate communication channel is available these raw data can be resolved to a more expressive, human readable format as shown in the screen shot on the right.**

of data collection, and the target range of tags. If a separate long range communication channel is available (e.g., wireless LAN or GSM), the watchdog tag can additionally translate the data transmitted over the RFID channel into a more expressive format, as shown in figure 8. Of course, any such lookup would require an appropriate backend infrastructure, e.g., the ONS architecture developed by the Auto-ID Center [13]. In addition, providing the reader location in a human readable format allows for a simple, manual detection of reader ID spoofs. More sophisticated watchdog tags featuring an integrated location system could potentially detect reader ID spoofing automatically.

The above screen shots were taken from our initial watchdog prototype, which serves as our design test bed for our protocol extension. Built on top of a standard Windows CE PDA, it uses the built-in wireless LAN to retrieve human readable descriptions. While we are currently working on a separate antenna design that allows us to interface our PDA directly with the RFID reader's communication channel, we so far have been simulating the complete RFID protocol over the wireless LAN as well (with a PC posing as a virtual RFID reader).

DISCUSSION AND FUTURE WORK

Even with our proposed protocol extensions, unauthorized read attempts by readers not conforming to our specification will still be possible. While consumers carrying a watchdog tag might be able to actively jam or block the tag-to-reader communication [11], for example based on user preferences regarding the reader's ID (e.g., following an online lookup), the average consumer would still need to resort to explicitly disabling her tags in order to completely prevent misuse. However, even without any additional devices, the required selection mechanism at the protocol level supports the core principle of *collection limitation*, while the compulsory identification string facilitates the principles of *openness* and *accountability*, thus providing the same level of protection as today's compulsory forms, signs, and placards announcing the privacy policy of the data collector. While they might be ignored in the routine of our everyday, their presence forms an important legal lever once a dispute over the proper use of personal data arises.

Our proposed protocol extensions are easily realized even with today's readers, as they only require updates to the reader's firmware, since the physical layer remains unaltered. While tags would require changes to their logic, these should be straightforward to implement, as the physical layer is not affected and only slight alterations to the medium access layer and the command set would be necessary. Our extensions do, however, affect the performance of an RFID system. The addition of the RPID, purpose code and collection type require the additional transmission of 130 bits. At a data transfer rate of 30 kBit/s, typical for reader-to-tag signalling of systems operating in the UHF band, it prolongs the execution time of any command by 4.3 ms. This delay is thus comparable to the time it takes for a single tag to reply with its ID, assuming symmetrical data transfer rates. In modern RFID systems that typically read several dozens, if not hundreds of tags at a time, loosing a single tag slot thus seems negligible. For an RFID system that features a slow data transfer rate, e.g., 1.6 kBit/s as specified in ISO 15693 (HF), the delay is more significant, approximately 80 ms. However, in many situations such a delay would be outweighted by the shortened reply times, as the `Select` command allows the reader to ignore tag IDs that are of no interest to the application in the first place. Newly arriving tags in the read range will have to wait for the next select command before they can be inventoried by a reader.

Future tags might also be able to incorporate basic cryptographic functionalities, thus facilitating a national or even supra-national (e.g., EU-wide) certification system for IDs, as well as allowing tags to thwart an imposter's attempt to "steal" the identification string of a valid reader (thus supporting the FIP principle *security*). To this end, companies would need to register their identification strings with the corresponding authorities, which would use their private keys to sign the submitted ID. Tags would be preprogrammed with the certification agencies public key and could therefore verify the validity of the registration in real-time. In order to prevent replay attacks from rogue readers, not only the ID of a reader, but also the public key of its owner would be signed by the agency (and subsequently transmitted to the tags), which would use this public key for all subsequent communication with the reader. Unauthorized readers would also need the real owner's private key in order to decipher tag IDs. Even though certificate revocation will not work with this scheme, the damage due to unrevokable certificates seems negligible, given the ability of consumer interest groups or concerned citizens to use watchdog

tags with online lookup capabilities to detect misuse. Also, certified reader IDs could allow tags to implement the resurrecting duckling model proposed by Stajano [16], where tags would only respond to a "mother" reader, but ignore requests from all others. Instead of killing tags at checkout [3], stores would transfer their "mother" rights to the customer's reader, thus allowing for a safe post-sales RFID usage. Additionally, such "mother" readers could inhibit replies by "its" tags for non-desired purposes and intentions by unknown readers by programming the tags accordingly.

CONCLUSION

The work presented in this paper helps to build future privacy-aware RFID standards that are not only optimized for performance and low cost, but also satisfy the fair information principles. The key idea of our approach is to augment the communication protocol between RFID readers and tags with a feature set that identifies the reader to provide *openness* and *accountability*, enables RFID operators to disclose a *purpose specification* and collection type, and supports a selection mechanism to facilitate the principle of *collection limitation*. In concert with a watchdog tag or a similar device, selective jamming can support the principle of explicit *consent*, while the integration of readers into an overarching privacy-infrastructure such as "pawS" [12] would allow the enforcement of the *use limitation*, *data quality*, and *participation* principles. Its simplicity provides for a readily available, practical solution to many of today's RFID privacy concerns, while the possible integration of the watchdog tag functionality into future mobile phones might even make the detection of an RFID reader, its policy, and location in the future as easy as detecting the signal strength and operator IDs on a mobile phone today.

REFERENCES

1. Resolution on Radio Frequency Identification. 25th International Conference of Data Protection and Privacy Commissioners, November 2003.
2. Rakesh Agrawal, Jerry Kiernan, Ramakrishnan Srikant, and Yirong Xu. Implementing P3P Using Database Technology. In *Proceedings of the IEEE 19th International Conference on Data Engineering*, pages 595–606, Bangalor, India, March 2003. Computer Society, IEEE Press.
3. Auto-ID Center. *Draft protocol specification for a 900 MHz Class 0 Radio Frequency Identification Tag*, May 2003.
4. Privacy Rights Clearinghouse. Position Statement on the Use of RFID on Consumer Products. www.privacyrights.org/ar/rfidposition.htm/.
5. Lorrie Cranor, Marc Langheinrich, Massimo Marchiori, and Joseph Reagle. The Platform for Privacy Preferences 1.0 (P3P1.0) Specification. W3C Candidate Recommendation, HTML Version at `www.w3.org/TR/P3P/`, December 2000.
6. EPCglobal. EPC Tag Data Specification 1.1, November 2003.
7. European Commission. Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data, November 1995.
8. Klaus Finkenzeller. *RFID Handbook: Fundamentals and Applications in Contactless Smart Cards and Identification*. John Wiley & Sons, Ltd, 2003.
9. International Organization for Standardization. ISO/IEC 18000: Information technology automatic identification and data capture techniques - Radio frequency identification for item management air interface, 2003.
10. A.V. Gershman and A. Fano. *A wireless world: The Internet sheds its chains*. www.accenture.com/.
11. Ari Juels and Ronald L. Rivest. The Blocker Tag: Selective Blocking of RFID Tags for Consumer Privacy. In *10th Annual ACM CCS 2003*, May 2003.
12. Marc Langheinrich. A Privacy Awareness System for Ubiquitous Computing Environments. In L.E. Holmquist G. Borriello, editor, *4th International Conference on Ubiquitous Computing (UbiComp2002)*, pages 237–245, Springer-Verlag LNCS 2498, September 2002.
13. Michael Mealling. *Auto-ID Object Name Service (ONS) 1.0*, August 2003.
14. Organisation for Economic Co-operation and Development (OECD). Recommendation of the Council concerning Guidelines governing the Protection of Privacy and Transborder Flows of Personal Data, September 1980.
15. Sanjay E. Sarma, Stephen A. Weis, and Daniel W. Engels. RFID Systems and Security and Privacy Implications. In *Workshop on Cryptographic Hardware and Embedded Systems*, pages 454–470. Lecture Notes in Computer Science, 2002.
16. Frank Stajano. *Security for ubiquitous computing*. John Wiley & Sons, Ltd, 2002.
17. D. Wan. Magic medicine cabinet: A situated portal for consumer healthcare. In *Proceedings of the International Symposium on Handheld and Ubiquitous Computing, Karlsruhe, Germany*, September 1999.
18. M. Weiser, R. Gold, and J.S. Brown. The origins of ubiquitous computing research at PARC in the late 1980s. In *IBM Systems Journal*, pages 693–696, 1999.