RFID Traceability: A Multilayer Problem

Gildas Avoine and Philippe Oechslin

EPFL
Lausanne, Switzerland

Abstract. RFID tags have very promising applications in many domains (retail, rental, surveillance, medicine to name a few). Unfortunately the use of these tags can have serious implications on the privacy of people carrying tagged items. Serious opposition from consumers has already thwarted several trials of this technology. The main fears associated with the tags is that they may allow other parties to covertly collect information about people or to trace them wherever they go. As long as these privacy issues remain unresolved, it will be impossible to reap the benefits of these new applications. Current solutions to privacy problems are typically limited to the application layer. RFID system have three layers, application, communication and physical. We demonstrate that privacy issues cannot be solved without looking at each layer separately. We also show that current solutions fail to address the multilayer aspect of privacy and as a result fail to protect it. For each layer we describe the main threats and give tentative solutions.

Keywords: RFID, Privacy, Collision Avoidance, Communication Model.

1 Introduction

Often presented as a new technological revolution, Radio Frequency Identification (RFID) systems make possible the identification of objects in an environment, with neither physical nor visual contact. They consist of transponders inserted into the objects, of readers which communicate with the transponders using radio frequencies and usually of a database which contains information on the tagged objects. This technology is not fundamentally new. It has existed for several years, for example for ticketing on public transport, on motorway tollgates or ski-lifts and also for animal identification.

The boom which RFID technology is enjoying today rests essentially on the willingness to develop transponders which are cheap and of reduced size. The direct consequence of this step is a reduction in the capacity of the transponders, that is to say their computation, storage and communication capacities. This willingness is due (in part) to the Auto-ID Center[1] whose purpose is to standardise and promote very cheap RFID technology, by reducing the price

[1] The Auto-ID Center split in 2003, giving rise to the EPCGlobal Inc. [4] and the Auto-ID Labs [1].

A.S. Patrick and M. Yung (Eds.): FC 2005, LNCS 3570, pp. 125–140, 2005.

of transponders to less than 5 cents. Because of their reduced capacities, these transponders, usually called *tags*, bring their share of problems, in particular with regard to privacy issues, whether it be information leakage or traceability.

Firstly, we will present the existing and potential applications related to RFID and we will give a brief description of the technology. Then we will present in Section 2 the privacy threats in RFID systems and we show that contrary to the three well-known cryptographic concepts, i.e., confidentiality, authenticity, and integrity, traceability cannot be ensured in the application layer only, but it must be ensured in each of the layers of the communication model. We will then analyse the traceability threats in each of the three layers of the radio frequency communication model and we will suggest some ideas in order to thwart them. We will go from the application layer in Section 3 to the physical layer in Section 5. We will finally summarise our analysis in Section 6.

1.1 RFID Applications

Advocates of RFID tags call them the super barcodes of the future. Based on a very different technology, identification by radio frequency represents a major innovation in relation to optical identification. In addition to the miniaturisation of the tags which allows them to be implanted within objects, it allows objects to be read en masse, without the need for visual contact. It should also be noted that each tag has a unique identifier: whilst a bar code represents a group of objects, an electronic tag represents a single object.

One area of application for RFID tags is the management of stock and inventories in shops and warehouses. The American mass-marketing giant, Wal-Mart, has recently placed a requirement on its suppliers that they use electronic tags on the palettes and packaging boxes that are delivered to it. This is a progressive policy and, at the beginning, it will only affect suppliers of pharmaceutical products.

The introduction of RFID tags in all articles could also directly benefit the consumer. One obsession of customers is cutting the waiting time at tills, by replacing the shop assistants with an entirely automated device: one would simply pass the contents of the trolley through a reading tunnel. This application will not see the light of day anytime soon, principally for technical reasons, but also for a less frequently thought about reason like fraud. Indeed, the electronic tags can be cloned or rendered ineffective through various processes, which clears the way for malicious activity. Even though barcodes can equally be cloned by a simple photocopy, this type of fraud is thwarted by a human presence when the goods are scanned at the till: in case of doubt, the shop assistant can verify the appropriateness of a product with the description corresponding to the barcode. Some visionaries go even further: the tags could contain information useful in the home, like washing, cooking or storing instructions. Thus maybe the washing machine that asks for confirmation before washing whites with reds or the refrigerator that discovers that a pot of "crème fraîche" stored on its shelves is no longer as fresh as its name suggests may no longer be science fiction?

These domestic applications are still experimental and should not detract from the very real applications which already surround us, e.g., the identification of pets by RFID is already a part of our everyday lives. In the European Union, this practice is already obligatory in some countries and will extend across the whole EU in a few years.

1.2 RFID Technology

Very cheap tags, electronic microcircuits equipped with an antenna, have extremely limited computation, storage, and communication capacities, because of the cost and size restrictions imposed by the targeted applications.

They have no microprocessors and are only equipped with a few thousand logic gates at the very most, which makes it a real challenge to integrate encryption or signature algorithms into these devices. This difficulty is reinforced by the fact that the tags are passive, meaning that they do not have their own energy source: they use the power supplied by the magnetic or electric field of the reader. Let us note, however, that promising research is being done at the moment, notably the implementation of AES for RFID tags proposed by Feldhofer, Dominikus and Wolkerstorfe [6].

The storage capacities of RFID tags are also extremely limited. The cheapest devices have only between 64 and 128 bits of ROM memory, which allows the unique identifier of the tag to be stored. Adding EEPROM memory remains an option for more developed applications. Whilst some memory zones can be made remotely inaccessible, the tags are not tamper-resistant, unlike smartcards made for secure applications (credit cards, pay TV, etc.).

The communication distance between tags and readers depends on numerous parameters, in particular the communication frequency. Two principal categories of RFID systems coexist: the systems using the frequency 13.56MHz and the systems using the frequency 860-960MHz, for which the communication distance is greater. In this latter case, the information sent by the reader can in practice be received up to a hundred meters, but the information returned from the tag to the reader reaches a few meters at most. These limits, resulting from standards and regulations, do not mean that the tags cannot be read from a greater distance: non-conforming equipment could exceed these limits, for example by transgressing the laws relating to the maximum authorised power.

2 Privacy Threats

Among the security issues related to RFID technology, one can distinguish those which threaten the functionality of the system from those which pose a threat to the privacy of its users, i.e., by divulging information about the user or allowing the user to be traced. The fact that the tags can be invisible, that they can be read remotely, and that they have a long life (considering that they do not have their own energy source), makes these privacy issues worse. Moreover, the ease of recording and automatically dealing with the logs obtained by RFID

systems contributes to the desire for protection against the undesirable effects of these systems.

Beyond the usual denial of service attacks, threats to the functionality of RFID systems are linked to the falsification of the tags and their concealment. As discussed in the previous section, a cheap tag cannot benefit from protection mechanisms such as those enjoyed by smartcards made for secure applications. Therefore an adversary can obtain the memory content and create a clone of the tag. This operation is obviously simplified if the tag openly transmits all its data, as is the case in the common applications. Although reducing the reading distance reduces the risks of eavesdropping, it is not a satisfactory solution. High gain antennas and use of non conforming power levels may still make it possible to read a tag from greater distances. The possibility of neutralising the tags also prevents the correct functioning of the system. The totally automated trolley reader discussed in Section 1.1 is particularly vulnerable to this kind of attack, since foil or a simple chips packet can be enough to neutralise the tag by forming a Faraday cage.

Below we will concentrate on threats to the privacy of RFID tag carriers. These threats fall into two categories: information leakage and traceability.

2.1 Information Leakage

The disclosure of information arising during the transmission of data by the tag reveals data intrinsic to the object or its environment. For example, tagged pharmaceutical products could reveal data about the health of a person. An employer, insurer or other party could have a particular interest in knowing the state of health of a person that he is close to, and could so obtain sensitive information. The tags are not, however, made to contain or transmit large quantities of data. When a database is present in the system, the tag can send a simple identifier, so that only the people who have access to this database can match the identifier to the corresponding information. This is the principle adopted by systems using barcodes.

2.2 Traceability

The problem of traceability is more complex. Even if a tag only transmits an identifier, this information can be used to trace an object in time and space. If a link can be established between a person and the tags he is carrying, the tracing of objects can become the tracing of a person. An attacker may want to trace a given tag, either deterministically or probabilistically, starting from either an active or passive attack.

A simple approach for dealing with the problem of privacy is to prevent the readers from receiving data coming from the tags. Besides the difficulty of putting these techniques into practice, they have the pernicious side-effect that they can also be used by an adversary to harm the system. The first technique arising from the need to ensure privacy is to *kill the tag*. The technique is effective but has the major inconvenience that the tag can no longer be used. A less radical method consists of *preventing the tag from hearing the request* by enclosing the tag in

a Faraday cage as we have already mentioned. This solution is only suitable for a few precise applications, e.g., money wallets, but is not for general use: animal identification is an example of an application which could not benefit from this technique. The third technique consists of *preventing the reader from understanding the reply.* The best illustration of this technique is surely the *blocker tag* [14] which aims to prevent a reader from determining which tags are present in its environment. Broadly, the blocker tag relies on the tree walking protocol (see Section 4.2) and simulates the full spectrum of possible identifiers.

Another approach is to design protocols which can identify the tags without compromising the privacy of their carriers. In spite of the huge interest that RFID technology has caused (and the fears of consumers), relatively few people have worked on such protocols. Sarma, Weis and Engels were the first to take a step in this direction [19]. Other protocols were then put forward, in particular by Juels *et al.* [13], Ohkubo *et al.* [16], Henrici and Müller [7], Feldhofer *et al.* [6], Molnar and Wagner [15], etc. Up until now, little work has been done to prove the security or to exhibit weaknesses of the proposed protocols. Only Avoine [2] and Saito *et al.* [18] paved the way by showing weaknesses in some existing schemes.

Unfortunately, we will show in Section 2.3 that even if an identification protocol is proven to ensure the privacy in the classical adversarial models, this does not mean that the protocol truly ensures privacy in practice.

2.3 Relationship Between Traceability and Layers

The three main concepts that are considered in cryptography are confidentiality, integrity and authentication. To analyse these concepts, a model of the adversary is defined, that is, the actions that the adversary may carry out on the entities and their communication channels in order to compromise confidentiality, integrity or authentication. This model is usually defined in theoretic notions like tamperproofness of the entities or timeliness of the channels without considering the exact nature of the underlying physical architecture.

The communication channels are usually devised using a layered approach (as in the OSI model [12]). By implementing a corresponding protocol at a given layer, confidentiality, integrity or authentication can be guaranteed independently from the characteristics of the lower layers. With regard to traceability, the problem is very different. Each layer can reveal information which can be used to trace a tag and we have to prove that the system is tracing-resistant at each layer. Thus, a protocol that is safe with regard to traceability in a classic adversary model may not be safe in practice. This is the case for all RFID protocols that have been described in the literature, since lower layers are never taken into consideration. It is thus of paramount importance to investigate traceability issues at each layer of the communication model. Below we refer to the model in Fig. 1 which is compatible with the ISO standard 18000-1 [10]. It is made of three layers, the application, the communication and the physical layer.

– The *application layer* handles the information defined by the user. This could be information about the tagged object (e.g., the title of a book) or more probably an identifier allowing the reader to extract the corresponding

information from a database. To protect an identifier, an application protocol may transform the data before it is transmitted or deliver the information only if certain conditions are met.
- The *communication layer* defines the way in which the readers and tags can communicate. Collision avoidance protocols are found in this layer as well as an identifier that makes it possible to single out a specific tag for communication with a reader (this identifier does not have to be the same as the one in the application layer).
- The *physical layer* defines the physical air interface, that is to say, the frequency, modulation of transmission, data encoding, timings and so on.

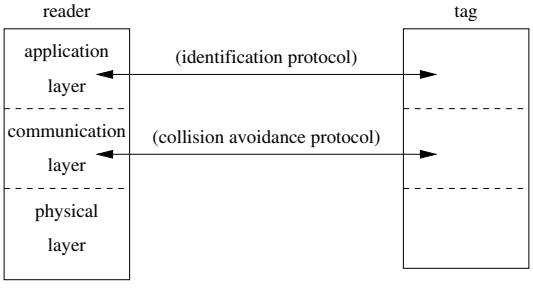

Fig. 1. Communication model

Up to now, very little of the work done has addressed this problem. We can cite Juels *et al.* [14], Molnar and Wagner [15], and Weis [21]. In the following sections we will analyse privacy issues (traceability) at each of the three layers of the communication model.

3 Traceability at the Application Layer

3.1 Identification Protocols

We explained in Section 2.3 that the traceability issue has to be considered in each of the three layers of the communication model. Up until now, only the application layer has been extensively studied (e.g., [7, 13, 15, 16, 19]). Broadly, all these works are based on the fact that the information sent by the tag to the reader changes at each identification. This information is either the identifier of the tag or an encrypted value of it. What differentiates the existing protocols is the way in which this information is refreshed between two identifications. Usually, during this process, the reader supplies the tag with the next value to send (new identifier or new ciphertext) or data allowing the tag to carry out the refreshment by itself. So, we can represent many of the RFID protocols by a 3-round protocol whose exchanged messages contain respectively the request, the identifier of the tag, and data to refresh the identifier.

Therefore, in order to avoid traceability, the information sent by the tag needs to be indistinguishable (by an adversary) from a *random* value and to be *used only once*. If the reader is involved in the refreshment process, it can voluntarily send information which is not indistinguishable from a random value. We characterise the RFID protocols according to whether the reader is or is not involved in the refreshment process.

In the first case, the tag must be able to generate new information by itself. For example, Ohkubo *et al.* [16] propose an RFID protocol where the tag can refresh its identifier by itself by using two hash functions. Obviously, the identifier is used only once since the tag changes it by itself as soon as an identification is completed. Whilst this scheme is proven secure from the point of view of privacy, it suffers from scalability issues. Avoine and Oechslin [3] however have shown that complexity can be significantly reduced using a time-memory trade-off.

In the case where the reader is involved in the regeneration of the information, we need to be sure that this information is indistinguishable (by an attacker) from a random value, but also that this information is used only once. Many of the existing protocols suffer from these two problems. This shows the difficulty of defining tracing-resistant RFID protocols if the tag depends on the reader for generating such values. To illustrate our point, we present below an attack against the protocol of Henrici and Müller [7].

3.2 Case Study: Protocol of Henrici and Müller

The principle of the protocol is as follows: after the personalisation phase, the tag contains its current identifier (ID), the current session number i and the last successful session number i^*. When the system is launched, the database contains a list of entries, one for each tag it manages. Each entry contains the same data as is stored in the tag, augmented by a hash value of ID, $h(\mathrm{ID})$, which constitutes the database primary key and other additional data. ID and i are set up with random values and i^* equals i. The identification process is as follows (see Fig. 2):

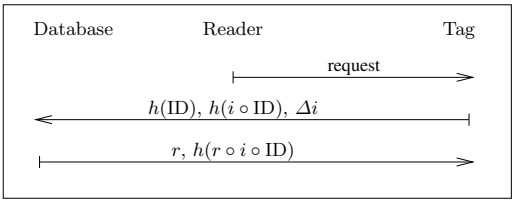

Fig. 2. Protocol of Henrici and Müller

1. The reader sends a request to the tag.
2. The tag increases its current session number by one. It then sends back $h(\mathrm{ID})$, $h(i \circ \mathrm{ID})$ and $\Delta i := i - i^*$ to the reader which forwards the values to the

database. Here $\circ$ is a "suitable conjunction function"; "A simple exclusive-or function is adequate for the purpose" [7]. $h(\mathrm{ID})$ allows the database to recover the identity of the tag in its data; $h(i \circ \mathrm{ID})$ aims at thwarting replay attacks and Δi is used by the database to recover i and therefore to compute $h(i \circ \mathrm{ID})$.

3. The database checks the validity of these values according to its recorded data. If all is fine, it sends a random number r and the value $h(r \circ i \circ \mathrm{ID})$ to the tag, through the reader.
4. Since the tag knows i and ID and receives r, it can check whether or not $h(r \circ i \circ \mathrm{ID})$ is correct. If this is case, the tag calculates its new identifier $\mathrm{ID}' := r \circ \mathrm{ID}$ and $i^* := i$, which is used in the next identification. Otherwise it does not calculate ID'.

Note that due to resilience considerations, an entry is not erased when the database has replied to the tag, but a copy is kept until the next correct session: if the third step fails, the database will still be able to identify the tag the next time with the "old" entry. Thus two entries per tag are used in turn.

Attack Based on the Non-randomness of the Sent Information. The first attack consists of tracking a tag, in a probabilistic way, taking advantage of the side channel supplied by Δi. Indeed, since the tag increases its value i every time it receives a request (Step 2), even if the identification finally fails, while i^* is updated only when the identification succeeds (Step 4), an attacker may interrogate the tag several times to abnormally increase i and therefore Δi. Thanks to the fact that this value is sent in the message from the tag to the reader, the attacker is then able to (probabilistically) recognise his target later according to this value: if the adversary later interrogates a tag that sends an abnormally high Δi, he concludes that this is his target.

Attack Based on Refreshment Avoidance. Another attack consists of corrupting the hash value sent from the reader to the tag. When this value is not correct, "the message is discarded and no further action is taken" [7], so the tag does not refresh its identifier. Note, however, that it is difficult to modify this message due to the fact that the communication channel is wireless. We therefore propose a practical variant of this attack: when a reader interrogates a tag, the attacker interrogates this tag as well before the reader carries out the third step. Receiving the request from the attacker, the tag increases i. Consequently, the hash value sent by the reader seems to be incorrect since i has now changed. More generally, an attacker can always trace a tag between two correct identifications. In other words, this attack is possible because the signal to refresh the identifier comes from the outside of the tag, i.e., the reader.

Attack Based on Database Desynchronisation. A more subtle and definitive attack consists of desynchronising the tag and the database. In order to do this, when a reader queries a tag, the attacker performs the third step of the identification before the reader does it. The random value r sent in the third step by the attacker is the neutral element of the operation $\circ$. Typically, if $\circ$ is the exclusive-or operation (according to [7]), the attacker replaces r by

the null bit-string and replaces $h(r \circ i \circ \mathrm{ID})$ by $h(i \circ \mathrm{ID})$ obtained by eavesdropping the second message of the current identification. We have trivially $h(\mathbf{0} \oplus i \oplus \mathrm{ID}) = h(i \oplus \mathrm{ID})$. Hence, the tag does not detect the attack and computes its new identity $\mathrm{ID}' = \mathbf{0} \oplus \mathrm{ID}$ (which is equal to its "old" identity) and it updates i^*. Therefore, in the next identification, the tag and the database will be desynchronised, since the tag computes the hash value using the "old" ID and the "new" i^* whereas the database checks the hash value with the "old" ID and the "old" i^*: the test fails and the received message is discarded. Consequently, the database will never send the signal to refresh the tag's identifier and the tag is definitively traceable.

4 Traceability at the Communication Layer

4.1 Singulation Protocols

With several entities communicating on a same channel, we need to define some rules to avoid collisions and therefore to avoid information loss. This arises in RFID systems because when a reader sends a request, all the tags in its field reply simultaneously, causing collisions. The required rules are known as the *collision avoidance* protocol. The tags' computational power is very limited and they are unable to communicate with each other. Therefore, the readers must deal with the collision avoidance themselves, without the help of tags. Usually, they consist of querying the tags until all identifiers are obtained. We say that the reader performs the *singulation* of the tags because it can then request them selectively, without collision, by indicating the identifier of the queried tag in its request.

The collision avoidance protocols which are used in the current RFID systems are often (non-open source) proprietary algorithms. Therefore, obtaining information on them is difficult. Currently, several open standards appear and they are used more and more instead of proprietary solutions. We distinguish the EPC[2] family [4] from the ISO family [9]. Whether they are is EPC or ISO, there are several collision avoidance protocols. Choosing one of them depends (in part) on the used frequency. EPC proposes standards for the most used frequency, i.e., 13.56MHz and 860-930MHz. ISO proposes standards from 18000-1 to 18000-6 where 18000-3 corresponds to the frequency 13.56MHz, and 18000-6 corresponds to the frequency 860-960MHz. We have two main classes of collision avoidance protocols: the deterministic protocols and the probabilistic protocols. Usually, we use the probabilistic protocols for systems using the frequency 13.56MHz, and the deterministic protocols for systems using the frequency 860-960MHz because they are more efficient in this case. After describing both the deterministic and the probabilistic collision avoidance protocols in Section 4.2 and 4.3, we will then analyse the traceability issues of these protocols.

[2] Electronic Product Code.

4.2 Deterministic Protocols

Deterministic protocols rely on the fact that each tag has a unique identifier. If we want the singulation process to succeed, the identifiers must stay unchanged until the end of the process. In the current tags, the identifiers are set by the manufacturer of the tag and written in the tag's ROM. In the usual RFID systems, there is no exchange after the singulation because the reader has obtained the expected information, i.e., the identifiers of the tags which are in its field. Below, we use *singulation identifier* to denote such an identifier, or more simply *identifier* where there is no ambiguity with the identifier of the application layer. We give an example of deterministic collision avoidance protocol called *tree walking*.

Suppose tags have a unique identifier of bit-length ℓ. All the possible identifiers can be visualised by a binary tree of depth ℓ. A node at depth d in this tree can be uniquely identified by a binary prefix $b_1 b_2 ... b_d$. The reader starts at the root of the tree and performs a recursive depth-first search. So, at node $b_1 b_2 ... b_d$, the reader queries all tags whose serial numbers bear this prefix, the others remain silent. The tags reply with the $d+1$-st bit of their serial number. If there is a collision, the reader restarts from the child of the prefix. When the algorithm reaches a leaf, it has detected a tag. The full output of the algorithm is a list of all tags within its field.

4.3 Probabilistic Protocols

The probabilistic protocols are usually based on a time-division multiple access protocol, called *Aloha*. We describe one of the variants of Aloha, namely the slotted Aloha. In the slotted Aloha, the access to the communication channel is split into time slots. In general, the number of slots is chosen randomly by the reader which informs the tags that they will have n slots to answer to its singulation request. Each tag randomly chooses one slot among the n and responds to the reader when its slot arrives. If n is not sufficiently large with regard to the number of tags which are present, then some collisions occur. In order to recover the missing information, the reader interrogates the tags one more time. It can mute the tags which have not brought out collisions (*switched-off* technique) by indicating their identifiers or the time slots during which they transmitted. Also, according to the number of collisions, it can choose a more appropriate n.

Note that although all the usual tags have a (unique) singulation identifier, this condition is not fundamentally required for Aloha, but is desirable for efficiency reasons [11]. Without using these identifiers, the exchange of information of the application layer is carried out during the singulation because the reader cannot communicate anymore with the tag when the singulation process is completed. Note also that the singulation seems *atomic* from the tag's view: whilst a tag must reply to the reader several times when the tree walking is used, the tag can answer only once when no collision occurs with the Aloha protocol. In the case where the response brings out a collision, the reader restarts a new singulation process with possibly a larger n. On the other hand, if the switched-off technique is used, then the protocol is not atomic anymore.

4.4 Threats Due to an Uncompleted Singulation Session

It is clear that deterministic collision avoidance protocols relying on the static identifiers give an adversary an easy way to track the tags. To avoid traceability, the identifiers would need to be dynamic. However if the identifier is modified during the singulation process, singulation becomes impossible. So we introduce the concept of *singulation session* as being the set of exchanges between a reader and a tag which are needed to singulate the latter. When the session does not finish, due to failures or attacks, we say that the session stays *open*.

Since the singulation identifier cannot be changed during a session, the idea, to avoid traceability, is to use an identifier which is different for each session. The fact that the tag can be tracked during a session is not really a problem due to the shortness of such a session. In practice, the notion of singulation session already informally exists because the readers usually send a signal at the beginning and end of a singulation. Unfortunately, there is no reason to trust the readers to correctly accomplish this task. In particular, a malicious reader can voluntarily keep a session open to track the tag thanks to the unchanged identifier. This attack cannot be avoided when the signals come from the reader and not from the tag itself.

Contrary to what we usually think, using a probabilistic protocol based on Aloha does not directly solve the traceability problem at the communication layer. Because, apart from the (inefficient) Aloha-based protocols which do not use the switched-off technique, the concept of singulation session is also needed with probabilistic singulation protocols. Indeed, after having queried the tags, the reader sends an acknowledgement (either to each tag or to all the tags) to indicate which tags should retransmit (either the reader acknowledges the identifiers of the tags it has successfully read, or it indicates the numbers of the slots where a collision occurred). In the case where the identifiers are used, the fact that a singulation session stays open allows an adversary to track the tags. In the case where the acknowledgement does not contain the identifiers but contains instead the numbers of the slots where a collision occurred, then an attack relying on these slots is also possible, as follows: an adversary who is in the presence of a targeted tag sends it a (first) singulation request with the number of potential slots n. Assume the tag answers during the randomly chosen slot s_{target}. The tag being alone, the reader can easily link s_{target} to the targeted tag. The reader keeps the session opened. Later, when the adversary meets a set of tags potentially containing its target, it interrogates the tags again, indicating that only tags which transmitted during s_{target} must retransmit: if a tag retransmits, there is a high probability, depending on n and the number of tags in the set, that it is the target of the adversary since another tag will respond to the (2nd) singulation request during s_{target} if, and only if, its last session stayed opened and it transmitted during s_{target}.

Whether we consider deterministic or probabilistic protocols, it is fundamental that singulation sessions cannot stay open. The tag needs to be able to detect such sessions and to close them by itself. In other words, the signal needs to be internal to the tag.

Consequently, we suggest using an internal timeout to abort singulation sessions with abnormal duration. Thus, the tag starts the timeout when the singulation session begins (i.e., when it receives the first request of a singulation session). When the timeout expires, the current session is considered as aborted.

Implementation of such a timeout strongly depends on the practical system, e.g., the timeout could be a capacitor. When the tag receives the first request of a singulation session, it generates a fresh identifier and loads its capacitor. Then, each time it is queried (such that the request is not the first one of a session), it checks whether its capacitor is empty. If this is the case, the tag erases its identifier and does not answer until the next "first" request. If it is not the case, it follows the protocol. Note that the duration of the capacitor may be less than the duration of a singulation session if this capacity is reloaded periodically and the number of reloads is counted.

4.5 Threats Due to Lack of Randomness

Changing the identifier of the tag is essential but does not suffice because these identifiers need to be perfectly random not to supply an adversary with a source of additional information. The use of a cryptographically secure pseudo-random number generator (PRNG), initialised with a different value for every tag, is indispensable for avoiding traceability. Of course, singulation must rely only on this random identifier without requiring other characteristic data of the tag.

In the tree walking case, [5] proposes for instance using short singulation identifiers which are refreshed for each new singulation using a PRNG. The used identifiers are short for efficiency reasons since there are usually only few tags in a given field. However, if the number of tags in the field is large, the reader can impose the use of additional static identifiers, available in the tag, set by the manufacturer! The benefit of using PRNG is therefore totally null and void.

In the case of Aloha, if the singulation identifiers do not appear in the acknowledgement sent by the readers, they do not directly bring information to an adversary. On the other hand, they supply much information through a side channel if we analyse how the slot is chosen by the tag. If this is randomly picked, it will not supply useful information to the adversary, but a non uniform distribution can open the door to attacks. Unfortunately this is the case with the current existing standards and protocols.

In order to illustrate our point, we can analyse the collision avoidance protocol proposed by Philips for its tag ICode1 Label IC [17] using the 13.56MHz frequency. It contains a 64 bit identifier of which only 32 are used for the singulation process, denoted by $b_1 ... b_{32}$. Although the tag does not have a PRNG, the implemented collision avoidance protocol is probabilistic. The choice of the time slot depends on the identifier of the tag and data sent by the reader. When the reader queries a tag, it sends a request containing: the number of slots n which the tags can use, where $n \in \{2^0, 2^1, ..., 2^8\}$, and a value $h \in 0, ..., 25$ called *hash value*. The selection of the time slot s_i is done as follows:

$$s_i := \text{CRC8}(b_{h+1}...b_{h+8} \oplus prev) \oplus n$$

where CRC8 is a *Cyclic Redundancy Check* with generator polynomial $x^8 + x^4 + x^3 + x^2 + 1$ and where *prev* is the output of the previous CRC8, initialised with `0x01` when the tag enters the field of a reader. Hence, an adversary can easily track a tag according to the slot chosen by the tag, if he always sends the same values h and n. One way to proceed is as follows.

An adversary sends to his (isolated) targeted tag a request with the number of slots n and the hash value h . The tag responds during slot s_{target}. When he meets a set of m tags, the adversary wants to know if his target is here. In order to do this, he sends a singulation request containing the same n and h. If no tag responds during s_{target} then the target is not included in the set of tags. However, the conditional probability that the tag is in the set given that at least one tag answers during slot s_{target} is

$$P(n, m, p) = \frac{p}{p + (1-p)(1 - (\frac{n-1}{n})^m)},$$

where p is the probability that the target is in the set[3].

Consequently, choosing the identifier, in the case of the three walking-based protocols, and choosing the time slot, in the case of the Aloha-based protocols, must be done using a cryptographically secure PRNG. Otherwise, an adversary may take advantage of the distorted distribution in order to track his target in a probabilistic way, or worse, to recover its identifiers as with the ICode1 tag.

5 Traceability at the Physical Layer

The physical signals exchanged between a tag and a reader can allow an adversary to recognise a tag or a set of tags even if the information exchanged can not be understood. All efforts to prevent traceability in the higher layers may be rendered useless if no care is taken at the physical layer.

5.1 Threats Due to Diversity of Standards

The parameters of radio transmission (frequency, modulation, timings, etc) follow given standards. Thus all tags using the same standard should send very similar signals. Signals from tags using different standards are easy to distinguish. A problem arises when we consider sets of tags rather than a single tag. In a few years, we may all be walking around with many tags in our belongings. If several standards are in use, each person may have a set of tags with a

[3] Note that in the particular case of the ICode1 tag, where the CRC-8 is applied on a 8-bit word, we can actually recover 8 bits of the identifier by sending only one singulation request! Therefore, by sending 4 requests with respectively $h = 0$, $h = 8$, $h = 16$, and $h = 24$, the adversary will be able to recover the 32 bits of the tag's singulation identifier.

characteristic mix of standards. This mix of standards may allow a person to be traced. This method may be especially good at tracing certain types of persons, like military forces or security personnel.

To reduce the threats of traceability due to characteristic groups of tags it is thus of paramount importance to reduce the diversity of the standards used in the market. Note that even if it is possible to agree on a single standard to use when RFID tags become popular, there will be times when a standard for a new generation of tags will be introduced. During the period of transition it will be possible to trace people due to characteristic mixes of old and new tags.

5.2 Threats Due to Radio Fingerprinting

Radio fingerprinting is a technique that has been used in mobile telephony to recognise cloned phones. By recording characteristic properties of the transmitted signals it is possible to tell a cloned cell-phone from the original one. Small differences in the transient behaviour at the very beginning of a transmission allows for the identification of transceivers even if they are of the same brand and model [20]. In the case of RFID tags, there will be too many tags in circulation to make it possible to distinguish a single tag from all other tags of the same model. Nevertheless, there will be several manufacturers in the market and their tags will have different radio fingerprints. It will thus be possible to trace a person by a characteristic mix of tags from different manufacturers.

Preventing traceability through radio fingerprinting seems quite difficult. There is no benefit for the manufacturers to produce tags that use exactly the same technology, producing the same radio fingerprint. Much more likely, manufacturers will experiment with different technologies in order to produce tags that have either better performance, price or size.

6 Conclusion

As we have shown in this paper, until now privacy issues in RFID systems have only been considered in classical cryptographic models with little concern for the practical effects on traceability when the theory is put into practice. We have shown that, contrary to the three basic concepts of cryptography, i.e., confidentiality, authentication, and integrity, traceability has to be considered with regard to the communication architecture. Thus, to create a fully privacy-friendly RFID system, privacy has to be ensured at each of the three layers of the communication model. We have described the threats that affect each of these layers and we have given some practical examples in order to illustrate our theories. We have included recommendations or solutions for each of these layers, although we have found that ensuring both privacy and scalability at the application layer seems difficult without sacrificing the low cost constraint.

Acknowledgments

The work presented in this paper was supported (in part) by the National Competence Center in Research on Mobile Information and Communication Systems (NCCR-MICS), a center supported by the Swiss National Science Foundation under grant number 5005-67322.

References

1. Auto-ID Labs. `http://www.autoidlabs.org`.
2. G. Avoine. Privacy issues in RFID banknote protection schemes. *Smart Card Research and Advanced Applications – CARDIS*, pp. 33–48, Kluwer, 2004.
3. G. Avoine and Ph. Oechslin. A scalable and provably secure hash based RFID protocol. *International Workshop on Pervasive Computing and Communications Security – PerSec 2005*, pp. 110–114, IEEE, 2005.
4. Electronic Product Code Global Inc. `http://www.epcglobalinc.org`.
5. EPC. Draft protocol specification for a 900 MHz class 0 radio frequency identification tag. `http://www.epcglobalinc.org`, February 2003.
6. M. Feldhofer, S. Dominikus, and J. Wolkerstorfer. Strong authentication for RFID systems using the AES algorithm. *Workshop on Cryptographic Hardware and Embedded Systems – CHES 2004*, LNCS 3156, pp. 357–370, Springer, 2004.
7. D. Henrici and P. Müller. Hash-based enhancement of location privacy for radio-frequency identification devices using varying identifiers. *Workshop on Pervasive Computing and Communications Security – PerSec 2004*, pp. 149–153, IEEE, 2004.
8. International Organization for Standardization. `http://www.iso.org`.
9. ISO/IEC 18000. Automatic identification – radio frequency identification for item management – communications and interfaces. `http://www.iso.org`.
10. ISO/IEC 18000-1. Information technology AIDC techniques – RFID for item management – air interface, part 1 – generic parameters for air interface communication for globally accepted frequencies. `http://www.iso.org`.
11. ISO/IEC 18000-3. Information technology AIDC techniques – RFID for item management – air interface, part 3 – parameters for air interface communications at 13.56 MHz. `http://www.iso.org`.
12. ISO/IEC 7498-1:1994. Information technology – open systems interconnection – basic reference model: The basic model. `http://www.iso,org`, November 1994.
13. A. Juels. "yoking-proofs" for RFID tags. *Workshop on Pervasive Computing and Communications Security – PerSec 2004*, pp. 138–143, IEEE, 2004.
14. A. Juels, R. Rivest, and M. Szydlo. The blocker tag: Selective blocking of RFID tags for consumer privacy. *Conference on Computer and Communications Security – ACM CCS*, pp. 103–111, ACM, 2003.
15. D. Molnar and D. Wagner. Privacy and security in library RFID: Issues, practices, and architectures. *Conference on Computer and Communications Security – ACM CCS*, pp. 210–219, ACM, 2004.
16. M. Ohkubo, K. Suzuki, and S. Kinoshita. Cryptographic approach to "privacy-friendly" tags. *RFID Privacy Workshop*, MIT, MA, USA, November 2003.
17. Philips. I-Code1 Label ICs protocol air interface, May 2002.
18. J. Saito, J.-C. Ryou, and K. Sakurai. Enhancing privacy of universal re-encryption scheme for RFID tags. *Embedded and Ubiquitous Computing – EUC 2004*, LNCS 3207, pp. 879–890, Springer, 2004.

19. S. Sarma, S. Weis, and D. Engels. RFID systems and security and privacy implications. *Cryptographic Hardware and Embedded Systems – CHES 2002*, LNCS 2523, pp. 454–469, Springer, 2002.
20. J. Toonstra and W. Kinsner. Transient analysis and genetic algorithms for classification. *IEEE WESCANEX 95. Communications, Power, and Computing*, volume 2, pp. 432–437, IEEE, 1995.
21. S. Weis. Security and privacy in radio-frequency identification devices (master thesis), May 2003.