BUDAPEST UNIVERSITY OF TECHNOLOGY AND ECONOMICS
DEPT. OF NETWORKED SYTEMS AND SERVICES

AUTOMATED SECURITY VERIFICATION OF NETWORKING PROTOCOLS AND QUERY AUDITING ALGORITHMS FOR WIRELESS SENSOR NETWORKS

Ta Vinh Thong

Thesis Booklet

Supervised by

Dr. Levente Buttyán
Dept. of Networked Systems and Services
Budapest University of Technology and Economics

Budapest, Hungary
2013

1 Introduction

Wireless Sensor Networks (WSNs) are given higher priority recently, thanks to their increasingly important role and widespread applications in everyday life. WSNs consist of spatially distributed sensors (called sensor nodes) to monitor physical or environmental conditions, such as temperature, sound, pressure, etc., at different locations. Each sensor node typically has a radio transceiver with an internal antenna or connection to an external antenna, a microcontroller, an electronic circuit for interfacing with the sensors and an energy source, usually a battery. WSNs consist of a large number of resource constrained sensor nodes and a few more powerful base stations. The sensors collect various types of data from the environment and send those data to the base stations using multi-hop wireless communications. For this reason, in the literature, the base stations are also called sink nodes. Communications in WSNs usually take place between the sensor nodes and the base stations, and it is important to distinguish the direction of those communications. In case of upstream communication, the sender is a sensor node, and the receiver is a base station, while in case of downstream communication, these roles are reversed. The goal of the sender is to reliably transmit to the receiver a full message that may consist of multiple fragments.

Up to date, numerous networking protocols and solutions have been proposed to ensure the reliable operation of WSNs applications in a hostile environment. However, despite the fact that WSNs are often envisioned to operate in hostile environments, some of the protocols and solutions do not address security issues at all, and as a consequence they ensure reliability only in a benign environment where no intentional attack takes place. Recognizing this problem, in recent years many research focused on proposing security protocols based on cryptographic methods. Unfortunately, designing security protocols is a very difficult and error-prone task, as confirmed by the fact that critical security holes can be found in many widely used protocols, including protocols secured by cryptographic operations, and believed to be secure by the protocol designers. The security vulnerabilities inherent in the designed protocols are often hard to spot, because of the huge number of behavioral scenarios defined in the protocols. In many cases, protocol and system designers only perform manual and informal analysis on their proposed protocols. The main problem is that informal analysis of protocols is error-prone, and security holes can be overlooked, hence, it is not considered to be a reliable approach. Addressing this problem, my research focuses on formal analysis and automated security verification of protocols for wireless sensor networks. Formal analysis is based on strong mathematical background, and uses formal languages that have expressive syntax and semantics, and give us a possibility to automate the security verification.

In my dissertation, I propose formal and automated verification methods for analyzing the security of protocols. I focus on the protocols and algorithms designed for wireless sensor networks (WSNs), which are related to the following three topics: (1) formal and automated security analysis of routing protocols for wireless ad-hoc sensor networks; (2) formal and automated verification of transport protocols for wireless sensor networks; and (3) query auditing algorithms for protecting sensitive information in statistical databases. In the following, I provide a brief overview of the three research topics that are covered in my dissertation. In this section, I only discuss the main problems that serve as motivation for my research in each topic. The research objectives, the major challenges, as well as the methodology are provided in the following three sections.

Topic 1: My first topic is related to a special application of wireless sensor networks. Namely, I focus on such applications in which the sensor nodes are deployed in devices which permanently change their locations, such as vehicular networks. This kind of network are also known as wireless ad-hoc sensor networks. Wireless ad-hoc sensor networks are not based on pre-defined topology, thus, in order to allow one party to communicate with another party, route discovery is accomplished. Once a route between two parties has been found, they start to exchange data on this route such that each party in the route forward the packet it received to the target. The route discovery procedure

is defined by routing protocols. Numerous attacks against routing protocols have been published, in which attacker(s) can achieve that the honest parties attempt to exchange data through a route that does not exist in reality, without being aware of it. This type of attacks, which I called as *route forging attack*, is critical because it can lead to futile energy consumption and can degrade the efficiency of the network.

Topic 2: The second topic is concerned with the security verification of transport protocols designed for wireless sensor networks. In some applications of WSNs, for instance, in case of multimedia sensor networks [6], the sensors capture and transmit high-rate data with some QoS requirements. Such applications require the use of a transport protocol that ensures reliable delivery and congestion control. Transport protocols used in wired networks (e.g., the well-known TCP) are not applicable in WSNs, because they perform poorly in a wireless environment and they are not optimized for energy consumption. Therefore, a number of transport protocols specifically designed for WSNs have been proposed in the literature (see e.g., [39] for a survey). The main design criteria that those transport protocols try to meet are reliability and energy efficiency. Unfortunately, existing transport protocols for WSNs do not include sufficient security mechanisms or totally ignore the security issue. Hence, many attacks have been found against existing WSN transport protocols. In general, we can talk about attacks against reliability and energy depleting attacks. An attack against reliability is considered to be successful if the loss of a packet (or packet fragment) remains undetected. In case of energy depleting attacks, the goal of the attacker is to force the sensor nodes to perform energy intensive operations, in order to deplete their batteries.

Topic 3: My third research topic focuses on the application of WSNs in hospital environment, where body mounted wireless sensor networks are used to collect medical data (e.g., ECG signals, blood pressure measurements, temperature samples, etc.) from a patient, and a personal device (e.g., a smart phone) is used to collect those data. The measured records are stored in a database on the personal device, and in the most cases they are sensitive information that only authorized person (e.g., attending physician) can access. In many cases, some kind of statistical information about the stored data is allowed to be accessed for external parties (e.g., hospital personnel, personal coach services, and health insurance companies, researchers). The statistical data is not sensitive for the patient, and one important requirement is that from the set of statistical data, the sensitive information cannot be inferred. For instance, the queries about the average of sensitive data are allowed to be provided, however, from these averages individual sensitive measurement data samples should not be deducible. To achieve this, the so called query auditors are deployed in the personnel devices.

Query auditing (QA) is the problem that has been studied intensively in the context of disclosure control in statistical databases. The goal of an off-line query auditing algorithm is to decide whether private information was disclosed by the responses of the database to a certain set of aggregate queries. Off-line query auditors work on queries received and responses provided in the past, therefore, they can only detect a privacy breach, but cannot prevent it. On-line query auditing algorithms, on the other hand, decide whether responding to a new incoming query would result in the disclosure of some private information, given the responses that have already been provided to past queries, and if responding to the new query would breach privacy, then the database can deny the response. Thus, on-line query auditing algorithms can prevent the unintended disclosure of private information. Various disclosure models are considered, namely, full disclosure and partial disclosure models. In the full disclosure case, the privacy of some data x breaches when x has been uniquely determined, while in the latter case x has been inferred to fall in a set consisting only small number of the possible values, or follows a skewed distribution.

2 Research Objectives

As discussed in Section 1, critical security holes can be found in many widely used routing and WSN transport protocols, including such protocols that are secured by cryptographic operations, and believed to be secure by the protocol designers. The security vulnerabilities inherent in the designed protocols are often hard to spot, because of the huge number of behavioral scenarios defined in the protocols. In many cases, protocol and system designers only perform manual and informal analysis on their proposed protocols. The main problem is that informal analysis of protocols is error-prone, and security holes can be overlooked, hence, it is not considered to be a reliable approach. One promising approach is to use formal methods, which have been widely-used in software engineering. The main advantage of formal analysis is that it helps increasing the confidence in a protocol by providing an analysis framework that is more systematic, and hence, less error-prone than the informal analysis. Moreover, formal analysis usually based on formal languages that have expressive syntax and semantics, which give us a possibility to automate the security verification.

Although in the literature there are several formal languages, as well as automated model-checking tools for verifying different properties of systems and protocols, e.g., [17], [30], [33], [8], [35], [28]. These methods are not designed specifically for analyzing routing protocols, hence, their specification languages lack several syntax and semantics elements required for routing protocols (e.g., broadcast sending). Therefore, they cannot be used to analyze routing protocols, or only in a very circumstantial way, applying model abstraction. In recent years, researchers focused on proposing specific methods for ad-hoc networks, e.g., [18], [19], [36], [7], [28], [12], [3], [38], [34]. However, the methods proposed in these related works have numerous drawbacks, for instance, they are not automated or they are based on less expressive formal languages that do not enable us to reason about route forging attacks in wireless ad-hoc networks.

Similarly, to the best of my knowledge, there is no any previous work which was focused on designing formal methods for analyzing WSN transport protocols. These protocols typically consist of complex behavioral elements, such as launching and resetting timers, probabilistic behavior, and performing cryptographic operations. Unfortunately, most related analyzing methods and tools (e.g., [30], [33], [8], [35], [28], [24]) are not well-suited for this purpose.

Addressing these problems, in the first two topics, my research focuses on proposing novel *formal analysis* and *automated security verification methods* designed for either proving the security of wireless ad-hoc network routing protocols and WSN transport protocols, respectively, or detecting security holes in them.

Query auditing is a problem that has been studied intensively in the context of disclosure control in statistical databases [5]. To the best of my knowledge, in all existing works on query auditing, the private information whose disclosure one wants to detect or prevent consists of the sensitive fields of individual records in the database (e.g., the salary of a given employee). The reason may be that statistical databases are mainly used for computing statistics over certain attributes of human users (e.g., the average salary of women employees), and in such applications, each database record corresponds to an individual person [5]. In contrast to these works, I define a novel setting for query auditing, where I want to detect or prevent the disclosure of aggregate values in the database (e.g., the maximum salary that occurs in the database), and I my goal is to propose efficient off-line and on-line query auditing algorithms in this new setting. More specifically, I study the problem of detecting or preventing the disclosure of the maximum (minimum) value in the database, when the querier is allowed to issue average queries to the database.

3 Challenges

In this section, I discuss the major challenges that one must face in each research topic. In case of the formal and automated verification of wireless ad-hoc network routing protocols, on the one hand, for specifying and reasoning about routing protocols, specific modeling elements such as broadcast communication, neighborhood, and communication range should be supported. Moreover, new theorems and bisimilarity definitions are required in order to enable us modeling the attacker model specific to wireless ad-hoc networks, and to analyze route forging attacks. On the other hand, during the automatic verification of routing protocols, a large number of network topologies and a strong attacker model need to be considered. This induces a huge number of states to be examined, which today's computer cannot always handle. My goal is to propose the (first) method that can handle arbitrary network topology and strong attacker model, which previous methods cannot provided.

The formal and automated security verification of WSN transport protocols is difficult because they typically consist of complex behavioral characteristics, such as real-time, probabilistic, and cryptographic operations. Moreover, the WSN transport protocols such as the SDTP protocol [11], includes these three behavioral characteristics at the same time. For these reasons, the analysis of this class protocols is very difficult. To the best of my knowledge, until now, there is no formal language which allows us to specify these three behavioral characteristics at the same time.

In most related works (e.g., [13], [31], [14], [23]) that address the query auditing problems, the values of individual attributes in the database are assumed to be unbounded real numbers. In contrast, I consider the query auditing problem in which the individual and the sensitive information take their values from some bounded interval $[\alpha, \beta]$, $\beta > \alpha$, of real numbers. The rationale behind this assumption is that in a hospital environment, the medical data (e.g., ECG signals, blood pressure measurements, temperature samples, etc.) is collected from a patient, which usually are lower-bounded and upper-bounded by some values. This assumption introduces some new problems, because the auditors which protect the privacy of sensitive data in case the attributes having unbounded domain, may not work in case of bounded attributes. For instance, if we want to protect the maximum value of a dataset, then the privacy is breached in case the attacker can deduce that some value in the dataset is equal to the upper-bound. This assumption provides the attacker more possibilities to deduce the sensitive information than in case of unbounded domain.

4 Methodology

For the first topic, I propose a variant of process algebra called the *sr*-calculus which, unlike previous works, provides expressive syntax and semantics for analyzing at the same time (i.) cryptographic primitives and operations, (ii.) the nature of broadcast communication, and (iii.) the specification of node's neighborhood in wireless medium, which are required for verifying secure routing protocols. The *sr*-calculus can be seen as the combination of the three calculi, the applied π-calculus [17], the omega calculus [36] and CMAN [18], with some modifications and extensions. I propose a fully automated verification method, called *sr*-verif, to verify the security of source routing protocols against route forging attacks. *sr*-verif is based on a backward deduction proof technique, combining with the well-known logic based resolution.

In the second topic, I propose a probabilistic timed calculus, called $crypt^{prob}_{time}$, for cryptographic protocols. The basic concept of $crypt^{prob}_{time}$ is inspired by the previous works [17], [20], [16] proposing solutions separately for each of the three discussed points. In particular, $crypt^{prob}_{time}$ is derived from the applied π-calculus [17], which defines an expressive syntax and semantics supporting cryptographic primitives to analyze security protocols; a probabilistic extension of the applied π-calculus [20]; and

a process calculus for timed automata proposed in [16]. I provide an approach for the automatic security verification WSN transport protocols with the PAT process analysis toolkit [37], which is a powerful general-purpose model checking framework. To the best of my knowledge, currently PAT is the most well-suited framework for this purpose due to its expressive syntax and semantics.

In the third research topic, in order to propose offline and online query auditors in the full disclosure model, I apply the well-known linear equation and linear optimization problems. In order to construct a simulatable query auditors for the case of probabilistic disclosure model, I apply the efficient random sampling approach [26], as well as the definition of the Chernoff bound and the Union bound, known in statistical theory.

5 New Results

In this section, I disscuss the main results that I proposed in my Ph.D thesis. Specifically, I provide the main and the sub-theses of my Ph.D dissertation, along with a brief overview of my proposed methods.

5.1 Formal and automated security verification of wireless ad-hoc routing protocols

My first thesis group contains the results for the first research topic, namely, the proposed formal and automated security verification methods for wireless ad-hoc routing protocols. I focus on verifying the security of on-demand source routing protocols in which the information about the route is included in request and reply messages in form of an ID list, against route forging attacks. I assume internal attacker nodes, meaning that they are compromised nodes, which can perform computations like honest nodes, and posses information that honest nodes can have according to the protocol. But unlike the honest nodes, attacker nodes can either decide to follow the protocol or not. In the latter case attacker nodes can modify messages, and when it intercepts a request it can remain idle and does nothing, or it can forward messages unchanged. Attacker nodes can cooperate with each other, and they can run parallel sessions at the same time.

Thesisgroup 1. *For verifying source routing protocols, I proposed a new variant of algebra based formal language, called sr-calculus, and a new automated verification method, called sr-verif, which are specifically designed for on-demand source routing protocols [Th05, Th06, Th07, Th08]. The main advantage of the proposed methods is that they support the modeling of cryptographic primitives, as well as the specification of broadcast communications, hence, they are suitable for reasoning about secured routing protocols. To the best of my knowledge, these are the first methods that are optimized for analyzing secured source routing protocols. I applied my methods to analyze well-known routing protocols such as the SRP [32], the Ariadne [22], and the endairA [3] protocols.*

The problem identified in [12] is the main motivation for proposing a formal and automated verification method for analyzing the correctness of secure routing protocols. In this paper, the authors showed that many routing protocols, which are believed to be secure by the protocol designers, turned out to be vulnerable. The authors indentified the necessity of formal analysis methods for routing protocols. Based on the formal framework, called simulation paradigm, they showed tricky attacks against the SRP and the Ariadne "secure" routing protocols. This paper is one of the first attempt to apply formal method for security analysis of routing protocols. The main drawback of this solution, however, is that it cannot be automated, because it does not based on a systematic

deduction approach, and no formal language is used for protocol specifications. More precise and systematic approaches are required.

5.1.1 My proposed formal analysis method for secured source routing protocols

Thesis 1.1. *I proposed a new variant of process algebra, called the sr-calculus, for reasoning about the security properties and route forging attacks against source routing protocols [Th05, Th06, Th08]. The sr-calculus is based on the applied π-calculus [17], the omega-calculus [36] and CMAN [18]. The novelty of the sr-calculus is that it supports modelling elements for (i) the attacker's accumulated knowledge base, (ii) cryptographic primitives and operations, and (iii) broadcast communications. In order to formally prove or refute the security of source routing protocols, I proposed a new definition of labeled bisimilarity for wireless ad-hoc networks. Finally, I showed that the syntax and semantics of my proposed sr-calculus is well-defined.*

The advantage of my proposed *sr*-calculus is that its expressiveness allows for modelling broadcast communication, neighborhood, and transmission range like CMAN [18] and the ω-calculus [36], and cryptographic primitives like the applied π-calculus and the spi-calculus [1, 17], however, compared to them it includes the definition of *active substitution with range* that is novel and enables us to model attacker knowledge and attacks in the context of wireless ad-hoc networks. In addition, *sr*-calculus is equipped with new theorems and bisimilarity definitions that allow us to model the attackers specific to the context of wireless ad-hoc networks, and to analyze route forging attacks. The detailed description of the *sr*-calculus can be found in Section 2.5 of my dissertation. Sections 2.5.1 and 2.5.2 of my dissertation discuss the type system and the formal syntax of the *sr*-calculus, while the operational semantics of the *sr*-calculus can be found in Section 2.5.3 of the dissertation.

Thesis 1.2. *Using the proposed sr-calculus and labeled bisimilarity, I proved that the SRP protocol is vulnerable to route forging attacks in case of one compromised node [Th05, Th06, Th08].*

I proposed a new bisimilarity definition for the *sr*-calculus, namely, the labeled bisimilarity in context of wireless ad-hoc networks, with which one can formally prove the security properties of source routing protocols. The labeled bisimilarity tells if two wireless ad-hoc networks are equivalent, meaning that they cannot be distinguished by an observer which can eavesdrop on communications. Below I provide an intuitive but informal definition of the labeled bisimilarity. The more formal description of the labeled bisimilarity, which requires several auxiliary notations and definitions, can be found in Section 2.5.3 of my dissertation.

Definition 1. *Labeled bisimilary for a given network topology ($\approx_l^N$) is the largest such symmetric relation ($\Re$) between two networks with the same node IDs and topologies N, and the following three properties are fulfilled:*

1. *The two sets of the output messages in the two networks cannot be distinguished from each other. We say that the two networks are statically equivalent;*

2. *One network can simulate any internal reductions (internal computations) performed within the another network, such that they remain statically equivalent after performing these corresponding reductions, and vice versa.*

3. *One network can simulate any labeled transitions (message outputs/inputs) performed within the another network, such that they remain statically equivalent after performing these corresponding transitions, and vice versa.*

In order to verify the security of a given source routing protocol, *routeprot*, based on the labeled bisimilarity, I define two *sr*-calculus specifications for *routeprot*, namely, the real specification and the ideal specification. The real specification of *routeprot* is denoted by $E^{real}_{routeprot}$, which follows exactly the (informal) definition of the *routeprot* routing protocol. The ideal specification of *routeprot* is denoted by $E^{ideal}_{routeprot}$, which is defined in the same way as $E^{real}_{routeprot}$, except for the specification of the source node. The only difference between $E^{real}_{routeprot}$ and $E^{ideal}_{routeprot}$ is that in $E^{ideal}_{routeprot}$, the source node is able to check the validity of the returned route. I showed that $E^{real}_{routeprot}$ and $E^{ideal}_{routeprot}$ are not labeled bisimilar, hence SRP is vulnerable. The detailed analysis of SRP can be found in Section 2.6 of my dissertation.

Thesis 1.3. *I proposed the Backward Deduction proof technique for Source Routing protocols (BDSR), by combining the mathematical background of the sr-calculus and the backward deduction approach. Based on the proposed BDSR and labeled bisimilarity definition, I proved that the Ariadne protocol is vulnerable to route forging attacks in case of one compromised node. In addition, I showed that the endairA protocol is secure against the route forging attacks when one compromised node is assumed, and it is vulnerable when we allow several cooperative attacker nodes [Th06, Th08].*

I develope a systematic proof technique, called BDSR, that enables us to reason about the security of routing protocols in an efficient way. This proof technique is based on backward deduction, namely, we start with the assumption that the source has accepted an invalid route, and based on the definition of the protocol reason backward step-by-step to find out how this could have happened. In case we get a contradiction it means that the starting assumption must not be valid, and the protocol is secure. The novelty of the BDSR algorithm is that it combines the proposed labeled bisimilarity definition with the backward deduction method, which enables us to perform exhaustive analysis and to prove the security of protocols. The main advantage of BDSR is that it enables us to reason about the more complex routing protocols such as Ariadne and endairA. More details about the BDSR algorithm can be found in Section 2.7 of my dissertation.

5.1.2 My proposed automated verification method for secured source routing protocols

Thesis 1.4. *I proposed a fully automated verification method, called sr-verif, to verify the security of source routing protocols against route forging attacks [Th07, Th08]. sr-verif is based on the proposed backward deduction proof technique (BDSR), combining with the well-known logic based resolution. I proved that sr-verif is correct, that is, whenever an attack scenario is returned it is really an attack. I proved that the proposed sr-verif never gets into an infinite deduction loop and terminates within a finite number of steps, and also provided its worst-case complexity [Th07, Th08].*

The detailed discussion of the proposed *sr*-verif can be found in Section 2.8 of my dissertation. The automated deduction algorithm is discussed in Table 1 and Table 2 of Section 2.8.9 of my dissertation. The novelty of *sr*-verif compared to the related model-checking tools e.g., used in [7, 36] is that the operation of routing protocols can be given in the simplified version of the *sr*-calculus, which supports the modeling of cryptograghic operations and broadcast communication in a straightforward way. Compared to previous approaches that attempted to formalize the verification process of secure ad-hoc network routing protocols [12, 2, 3, 4, 18] *sr*-verif is fully automated. In addition, the main advantage of *sr*-verif compared with the related solutions (e.g., [7]) is that it does not assume any specific topology when performing the verification, instead it considers arbitrary topology. Finally, in contrast to [38, 34] my emphasis is deliberately on verifying security properties instead of loop-freedom properties.

My proposed *sr*-verif was inspired by the concept of the ProVerif automatic verification tool [9], however, as opposed to [9] it is designed for verifying routing protocols and includes numerous novelties such as broadcast communications, neighborhood, and considering an attacker model specific to wireless ad hoc networks. Namely, in *sr*-verif, the operation of routing protocols are specified in the *syntax of processes* of the simplified version of the *sr-calculus*. This is then translated to the well-known Horn-clauses, using translation rules. This set of clauses is called *protocol rules*. In addition, the topology and the initial knowledge of the attacker node are specified by a set of facts, while the computation ability of the attacker node is specified by the set of Horn-clauses. The clauses that specify the attacker computation ability are called *attacker rules*. The deductive algorithm is based on the resolution steps accomplished over these clauses and facts in a backward search manner.

In Section 2.8.10 of my dissertation (Lemmas 1 and 2), I proved that my proposed *sr*-verif terminates, and the deduction algorithm is infinite loop-free. In Section 2.8.11 of my dissertation (Lemma 3), I showed that my proposed sr-verif is correct, namely, whenever an attack is returned then it is a valid attack.

Thesis 1.5. *Using sr-verif I showed that the SRP and the Ariadne routing protocols are insecure in case of one attacker node, as well as the endairA protocol is vulnerable in case of several cooperative nodes [Th07, Th08].*

The systematic verification of SRP, Ariadne and endairA can be found in sections 2.8.13, 2.7.3, and 2.7.4 of my dissertation.

5.2 Formal and automated security verification of WSN transport protocols

In the second thesis group, I discuss the main thesis and the sub-theses related to the second research topic.

Thesisgroup 2. *For verifying WSN transport protocols, I proposed a probabilistic timed calculus for cryptographic protocols, called crypt$^{prob}_{time}$, and an automated verification method based on the well-known PAT process analysis toolkit [37]. Using crypt$^{prob}_{time}$ and PAT, I analyzed two previously proposed WSN transport protocols, the DTSN [27] and the SDTP [11] protocols, and showed that they are vulnerable. I proposed a new secured WSN transport protocols, SDTP^{+} [Th11], and proved that it is secure against the vulnerabilities that can be found in DTSN and SDTP. My related publications in this research topic are [Th11, Th13, Th12].*

Despite the fact that WSNs are often envisioned to operate in hostile environments, existing transport protocols for WSNs do not address security issues at all and, as a consequence, they ensure reliability and energy efficiency only in a benign environment where no intentional attack takes place [10]. Broadly speaking, attacks against WSN transport protocols can be attacks against reliability and energy depleting attacks. An attack against reliability is considered to be successful if the loss of a data packet (or packet fragment) remains undetected. In case of energy depleting attacks, the goal of the attacker is to force the sensor nodes to perform energy intensive operations, in order to deplete their batteries. Due to the complexity of WSN transport protocols, informal analyis of the designed protocols is error-prone, and subtle attack scenarios can be overlooked. Hence, formal analysis methods are required to be proposed for WSN transport protocols, which increase the reliability of the analysis.

5.2.1 My proposed formal analysis method for secured WSN transport protocols

Thesis 2.1. *I proposed a probabilistic timed calculus, called crypt$^{prob}_{time}$, for cryptographic protocols [Th13, Th12]. To the best of my knowledge, this is the first of its kind in the sense that it combines the following three features: (i.) it supports formal syntax and semantics for cryptographic primitives and operations; (ii.) it supports time constructs similar to the concept of timed automata that enables us to verify real time systems; (iii.) it also includes the syntax and semantics of probabilistic constructs for analyzing systems that perform probabilistic behavior. In addition, I proposed the novel definition of weak probabilistic timed bisimilarity for proving and refuting the security properties WSN transport protocols.*

The basic concept of *crypt*$^{prob}_{time}$ is inspired by the previous works [17], [20], [16] proposing solutions separately for each of the three discussed points. In particular, *crypt*$^{prob}_{time}$ is derived from the applied π-calculus [17], which defines an expressive syntax and semantics supporting cryptographic primitives to analyze security protocols; a probabilistic extension of the applied π-calculus [20]; and a process calculus for timed automata proposed in [16]. The design methodology of *crypt*$^{prob}_{time}$ is based on the terminology proposed in these works, it can be seen as the modification and extension of them, and contains some novelties.

Note that, although in my dissertation the proposed *crypt*$^{prob}_{time}$ calculus is used for analyzing WSN transport protocols, it is also suitable for reasoning about other systems that include cryptographic operations, as well as real-time and probabilistic behavior. Note that with *crypt*$^{prob}_{time}$, my purpose is to develop a formal proof method for probabilistic timed cryptographic protocols, and the question of how can an automated verification method based on *crypt*$^{prob}_{time}$ be designed is left for the future. In my dissertation, I used the well-known PAT process analysis toolkit [37] for automating the verification, instead of designing an automatic method based on *crypt*$^{prob}_{time}$.

I defined *crypt*, a variant of the applied π-calculus [17], as the base calculus of *crypt*$^{prob}_{time}$ which supports cryptographic primitives and operations. I defined *crypt* in a similar way as in the applied π-calculus, except that the recursive process invocation is used instead of process replication (because I want to follow the automata semantics). In addition, to analyze WSN transport protocols, I had to add some extra modeling elements. For supporting the comparison between integers, the set of extended processes in [17] is improved with the corresponding comparison processes. The timed extension of *crypt* is based on the timed calculus proposed in [25], [16], and it is also based on the semantics of the well-known timed automata. The probabilistic extension is inspired by the syntax and semantics of the probabilistic extension of the applied π-calculus proposed in [20], and the probabilistic automata in [16]. The main difference between my work and the related methods is that I focus on extending *crypt*, which is different from the calculus used in those works. In addition, I combine both timed and probabilistic elements at the same time. Finally, I also propose a new definition called *weak probabilistic timed bisimilarity* for proving the existence of the attacks against security protocols.

The concept of *crypt*$^{prob}_{time}$ is based on the concept of probabilistic timed automata, hence, the correctness of *crypt*$^{prob}_{time}$ comes from the correctness of the automata because the semantics of *crypt*$^{prob}_{time}$ is equivalent to the semantics of the probabilistic timed automata.

The detailed description of *crypt*$^{prob}_{time}$ can be found in Section 3.5 of my dissertation. Sections 3.5.1 and 3.5.2 of my dissertation discuss the formal syntax and the operational semantics of *crypt*$^{prob}_{time}$.

5.2.2 Security analysis of WSN transport protocols using *crypt*$^{prob}_{time}$

In this subsection, I give a brief overview about how to apply the proposed *crypt*$^{prob}_{time}$ for analyzing WSN transport protocols. This subsection is an excerpt of the Section 3.6 of the dissertation.

Definition 2. (***Weak prob-timed labeled bisimulation for crypt$^{prob}_{time}$ processes***)
We say that two states $s_1 = (A^1, v_1)$ and $s_2 = (A^2, v_2)$ are weak prob-timed labeled bisimilar, namely (s_1 $\Re^p_t$ s_2) iff

1. *an observer cannot distinguish the message outputs in A^1 and A^2 (statically equivalence);*
2. *If from s_1 we can reach the state s'_1 after a silent (internal) action after d time units, then s_2 can simulate this action via the corresponding silent action trace, leading to some s'_2, and s'_1 $\Re^p_t$ s'_2 holds again.*
3. *If from s_1 we can reach the state s'_1 after a non-silent labeled transition after d time units, then s_2 can simulate this action via the corresponding labelled transition trace, leading to some s'_2, and s'_1 $\Re^p_t$ s'_2 holds again,*

and vice versa.

In my provided formal proofs, I applied the proof technique that is usual in process algebras. Namely, I define an ideal version of the protocol run, in which I specify the ideal/secure operation of the real protocol. This ideal operation, for example, can be defined such that honest nodes always know what is the correct message they should receive/send, and always follow the protocol correctly, despite the presence of attackers. Then, I examine whether the real and the ideal versions, running in parallel with the same attacker(s), are weak prob-timed bisimilar.

Definition 3. *Let the processes Prot() and Prot^{ideal}() specify the real and ideal versions of some protocol Prot, respectively. We say that Prot is secure if Prot() and Prot^{ideal}() are weak probabilistic timed bisimilar: Prot() $\approx_{pt}$ Prot^{ideal}().*

The main difference between the ideal and the real systems is that in the ideal system, honest nodes are always informed about what kind of packets or messages they should receive from the honest sender node. This can be achieved by defining hidden or private channels between honest parties, on which the communication cannot be observed by attacker(s). The honest sender informs the honest receiver about the message it should receive, and the receiver ignores the incorrect messages.

Thesis 2.2. *Using crypt$^{prob}_{time}$ I specified the behavior of the previously proposed DTSN protocol. I proved that the DTSN protocol is vulnerable to both the modification/forging of data packets and control packets, when there is one compromised node in the path from the source to the destination [Th13, Th12].*

The security properties I want to check in case of the DTSN protocol is that how secure it is against the manipulation of control and data packets. In particular, can the manipulation of packets prevent DTSN from achieving its design goal. To prove or refute the bisimilarity relation, I define *Prot*(*params*) and *Prot^{ideal}*(*params*) such that in *Prot^{ideal}*(*params*) a hidden communication channel is defined between every honest node pair, which is used to inform about the correct message the addressee should receive. To prove that the DTSN protocol is vulnerable against packet modification or forging attacks, I show that there is a transition in the real protocol *Prot*(*params*) which cannot be simulated by any corresponding transition trace in the ideal protocol *Prot^{ideal}*(*params*). The transition that makes the difference between the real and ideal variants of the protocol is related to the attack scenario in which a honest intermediate node accepts the forged packet in *Prot*(*params*), but not in *Prot^{ideal}*(*params*). In other words, I proved the violation of Definition 3. The specification of the DTSN protocol in *crypt$^{prob}_{time}$* can be found in Section 3.6.1, while its security analysis is discussed in Section 3.7.1 of my dissertation.

Thesis 2.3. *Using $crypt^{prob}_{time}$ I specified the behavior of the SDTP protocol, which is the security extension of DTSN. I proved that the SDTP protocol is secure against the modification/forging of data packets and control packets in case of one compromised node. However, I proved that SDTP is vulnerable against the forging of data and control packets in case of two cooperative attacker nodes [Th13, Th12].*

Similar to the case of DTSN, to analyze the security of SDTP I define a real and an ideal version of the protocol. Thereafter, I prove that SDTP is secure against the packet manipulation attack in case of one attacker node, but it is vulnerable when two cooperative attacker nodes are considered. The specification and the security analysis of the SDTP protocol in $crypt^{prob}_{time}$ can be found in sections 3.6.2 and 3.7.2 of my dissertation.

Thesis 2.4. *I proposed a new secured transport protocol for WSNs, called SDTP$^+$ [Th11], in order to patch the security holes that can be found in DTSN and SDTP. I proved that SDTP$^+$ is secure against the modification/forging of data packets and control packets in case of either one or two compromised nodes [Th11, Th12].*

I proposed SDTP$^+$ [Th11], a new secured WSNs transport protocols in order to patch the security weaknesses can be found in DTSN and SDTP. SDTP$^+$ aims at enhancing the authentication and integrity protection of control packets, and is based on an efficient application of asymmetric key crypto and authentication values, which are new compared to SDTP. The security mechanisms proposed in SDTP$^+$ are based on the application of hash-chains [15] and Merkle-trees [29], which have been broadly used in designing security protocols. Hash chains have been used in many secure routing protocols (e.g., Ariadne [22]). Similarly, Merkle trees have been used in securing WSN protocols [21]. My main contribution is the application of hash-chains and Merkle-trees in a new context, namely, for securing WSN transport protocol. The complete description of the SDTP$^+$ protocol can be found in Section 3.3 of my dissertation.

As for the analysis of SDTP$^+$, I defined the real and the ideal versions of SDTP$^+$. Based on the definition of weak prob-time bisimilarity, I showed that hash-chains and Merkle-trees eliminate the security vulnerabilities in DTSN and SDTP, as well as it can be efficiently applied in the context of WSN transport protocols.

5.2.3 Automated verification of WSN transport protocols using the PAT toolkit

This subsection is an excerpt of the Section 3.8 of the dissertation. I provide an approach for the automatic security verification of the DTSN and SDTP protocols with the PAT process analysis toolkit [37], which is a powerful general-purpose model checking framework. To the best of my knowledge PAT has not been used for this purpose before, however, in my dissertation I show that the expressiveness of PAT makes it well-suited for checking some interesting security properties defined for this class of protocols.

Thesis 2.5. *Using PAT I showed that the DTSN protocol is vulnerable to the control packet modification attack in presence of one attacker node [Th13, Th12].*

The detailed discussion of verifying DTSN can be found in Section 3.8.4 of my dissertation. The first main design goal of DTSN is to provide reliable delivery of packets. Let us consider the topology $S - I - A - D$, where S, I, A, D are IDs of the source, the intermediate node, the attacker, and the destination, respectively. The assertion, denoted by *violategoal1*, for verifying the security of DTSN regarding this first main goal is the following:

```
PAT code:
 #define violategoal (OutBufL == 0 && BufI == 0 && numNACK > 0);
```

where the (global) variables *OutBufL* and *BufI* are the number of the occupied cache entries at the source and intermediate nodes, respectively. The variable *numNACK* represents the number of the packets that are requiring to be retransmitted, namely, the number of the gaps in the data packet stream received by the destination. The goal *violategoal* represents the state in which the cache of S and I are emptied, but at the same time D has not received all of the packets sent by the source yet. Thereafter, I run the PAT process analysis toolkit for *violategoal*, and got an attack scenario. More assertions for more attack scenarios and corresponding topologies are discussed in my dissertation.

Thesis 2.6. *Based on PAT, I showed that SDTP is secure against the control packet modification attack in case of one attacker node, as well as it is vulnerable to the data and control packet forging attack in presence of cooperative attackers [Th13, Th12].*

The detailed discussion of verifying SDTP can be found in Section 3.8.5 of my dissertation. The SDTP protocol is the first security extension of the DTSN protocol with cryptographic primitives. As I already mentioned, the current form of the PAT toolkit does not support a convenient way for modeling crypto primitives. Hence, for instance the MAC (message authentication code) computed over the message *msg* with the *Kmac* is modelled by the pair *msg.Kmac*. In PAT *msg.Kmac* represents a composite message with two parts, which by default can be accessed by the attacker. Hence, to model the MAC with *msg.Kmac* we have to model the behavior of the attacker node such that when it obtains *msg.Kmac* it cannot use the key *Kmac* and cannot change the part *msg* to some other *msg'*. I model the rest crypto primitives in the same way, which are detailed in my dissertation.

First, I examined whether the security solution of SDTP can eliminate the main vulnerability of DTSN. I analyzed the security of SDTP against the successful attack scenarios in case of DTSN. Namely, for the topologies $S - I - A - D$ and $S - A - I - D$, besides *violategoal*. As result PAT returns *Not Valid*, which means that the security extension protects SDTP from the packet modification/forging attack causing the honest nodes empty their buffers while there are packets to be retransmitted.

However, I showed that SDTP is still vulnerable when two cooperative attacker nodes are assumed. In particular the attacker nodes can achieve that all the intermediate nodes delete more packets in their buffer than required, and in the worst case their buffers can be totally emptied. This basically, brings SDTP back to the end-to-end retransmission model, which is contrary to the design objective of SDTP. Let the assertion *violategoal2* represent the state where the intermediate nodes delete more packets than what the destination requested. Then, I analysed SDTP for the topology $S - A1 - I - A2 - D$ and *violategoal2*, and got Valid along with an attack scenario. The main vulnerability of SDTP that enables this attack is that the intermediate nodes do not perform any verification of the message origin.

5.3 Query auditing for protecting sensitive information in statistical databases

I address a new auditing problem by considering an *aggregation* value of a dataset to be sensitive and concentrating on protecting the privacy of aggregation values. In particular, I consider the problem of detecting or preventing the disclosure of the maximum (minimum) value, denoted by MAX (MIN), in the database, when the querier is allowed to provide average queries to the database.

Specifically, the query auditing problems that I am considering are defined as follows: Given t queries $q_1, \ldots, q_t$ over the stored data set $X = \{x_1, \ldots, x_n\}$. Each query q_i is of the form (Q_i, AVG), where $Q_i \subseteq \{1, 2, \ldots, n\}$, and the value of each x_i is assumed to be a real number that lies in a finite interval $[\alpha, \beta]$, where $\beta > \alpha$.

- Given all the t corresponding answers $a_1, \ldots, a_t$ for the corresponding queries $q_1, \ldots, q_t$, the task of the offline auditor is to detect if the value of MAX is fully disclosed.
- Given the first $t-1$ answers $a_1, \ldots, a_{t-1}$ for the corresponding queries $q_1, \ldots, q_{t-1}$, when a new q_t is posed, the task of the online auditor is to make a decision whether to answer or deny the query so that the privacy of MAX is preserved.

I denote the class of auditors that accept average queries and protect the privacy of the maximum value (as defined above), by $\text{Auditor}_{avg}^{max}$.

In contrast to the previous works, I assume that the domain of sensitive values is bounded, which leads to some new problems. As for the attacker model, I assume that there is only one attacker at a time, hence, I do not deal with the collusion attackers case. Moreover, I consider only one session at a time, not interleaving sessions. Moreover, within a session the attacker repeatedly poses average queries and its goal is to deduce somehow the maximum (minimum) values. The attacker can use any algorithm to compute the secrets based on the queries and answers.

In chapter 4 of my dissertation, I only discuss the auditors for the maximum value, however, I note that the case of protecting the minimum value (beside average queries) can be constructed in the same way, which is briefly discussed in [Th10].

Thesisgroup 3. *I proposed three $\text{Auditor}_{avg}^{max}$ query auditors of three different types. Namely, I proposed polynomial time off-line and on-line $\text{Auditor}_{avg}^{max}$ query auditors in the full disclosure model, as well as a simulatable on-line $\text{Auditor}_{avg}^{max}$ query auditor in the partial disclosure model [Th09, Th10].*

5.3.1 Offline and Online $\text{Auditor}_{avg}^{max}$ in the full disclosure model

Thesis 3.1. *I proposed a polynomial time offline $\text{Auditor}_{avg}^{max}$ query auditor for in case of full disclosure model. My proposed auditor is based on the application of the well-known linear optimization problem. I showed that the proposed offline auditor is sound, namely, if the auditor returns either that based on a series of t queries and the corresponding t answers, the value of MAX is fully disclosed or not, then this is really the case.*

The proposed offline auditor (more details can be found in Section 4.5 of my dissertation): I consider a method that takes into account the bounds of x_i's and the answers. For this purpose, I propose the application of the well-known linear optimization problem as follows: The t queries are represented by a matrix $\bar{A}$ of t rows and n columns. Each row $r_i = (a_{i,1}, \ldots, a_{i,n})$ of $\bar{A}$ represents the query set Q_i of the query q_i. The value of $a_{i,j}$, $1 \leq i, j \leq n$, is 1 wherever x_j is in the query set Q_i, and is a 0 otherwise. The corresponding answers are represented as a column vector $\bar{b} = (b_1, \ldots, b_t)^T$ in which b_i is the answer for q_i.

Since each attribute x_i takes a real value from a bounded interval $[\alpha, \beta]$ we obtain the following special linear equation system, also known as *feasible set*, which includes equations and inequalities:

$$\mathcal{L} = \begin{cases} \bar{A}\bar{x} = \bar{b}, \text{where } \bar{x} \text{ is the vector } (x_1, \ldots, x_n)^T. \\ \alpha \leq x_i \leq \beta, \forall x_i : x_i \in \{x_1, \ldots, x_n\} \end{cases}$$

Then, by appending each objective function $maximize(x_i)$ to $\mathcal{L}$, we get n linear programming problems P_i, for $i \in \{1, \ldots, n\}$. Let $x_i^{max} = maximize(x_i)$, then the maximum value of $x_1, \ldots, x_n$ is the maximum of the n maximized values, $x^{opt} = max\{x_1^{max}, \ldots, x_n^{max}\}$. Let us denote the whole linear programming problem above for determining the maximum value x^{opt} as $\mathcal{P}$. Note that x^{opt} returned by $\mathcal{P}$ is the exact maximum value if (i) $\mathcal{L}$ has a unique solution or (ii) $\mathcal{L}$ does not have a unique solution but there exist some x_i that can be derived to be equal to x^{opt}. Otherwise, x^{opt} is the best estimation of the exact maximum. Note that in our case $\mathcal{L}$ always has a solution, because one possible solution is actually the values stored in the database.

Based on this linear programming problem, our offline auditor will follow the next steps. Given t queries $q_1, \ldots, q_t$ over $X = \{x_1, \ldots, x_n\}$ and their corresponding answers $a_1, \ldots, a_t$, the value of MAX is fully disclosed in any of the following two cases:

- (F1) In case $\mathcal{L}$ has a unique solution, the value of MAX is equal to x^{opt}.
- (F2) In case $\mathcal{L}$ does not have a unique solution: If by following the solving procedure of $\mathcal{L}$ (e.g., basic row and column operations), there exist some x_i that can be uniquely determined such that $x_i = x^{opt}$, then the value of MAX is x_i. This is because x^{opt} is always at least as large as the value of MAX.

Otherwise, the attacker cannot uniquely deduce the value of MAX.

Thesis 3.2. *I proposed two variants of polynomial time online Auditor$_{avg}^{max}$ query auditors for in case of full disclosure model. My proposed auditors are based on the application of the well-known linear equation and linear optimization problem. I showed that the proposed online auditors ensure the privacy of the value of MAX in the full disclosure model [Th09, Th10].*

The proposed online auditor (can be found in Section 4.6 of my dissertation): Let us consider the first $t-1$ queries and answers over the data set similarly defined as in the offline case above. When a new q_t is posed, the task of the online auditor is to make a decision in *real-time* whether to answer or deny the query. More specifically, our goal is to propose an auditor that detects if answering with true a_t causes full disclosure of MAX. The proposed online auditor is based on the well-known linear optimization problem.

Algorithm 2/a: Online auditor Auditor$_{avg}^{max}$

Inputs: $q_1, \ldots, q_t$, $a_1, \ldots, a_t$, d_{tr}, α, β;

Let $\mathcal{L}_t^*$ be the feasible set formed by the t queries/answers

Let x_t^{opt} be the returned maximum by solving $\mathcal{P}$ with $\mathcal{L}_t^*$

if $|x_t^{opt} - MAX| > d_{tr}$ AND $(MAX - max_t) > d_{tr}$ **then** output a_t; **endif**

else if $|x_t^{opt} - MAX| \leq d_{tr}$ OR $(MAX - max_t) \leq d_{tr}$ **then** output DENY; **endif**

Algorithm 2/b: Online auditor Auditor$_{avg}^{max}$

Inputs: $q_1, \ldots, q_t$, $a_1, \ldots, a_t$, d_{tr}, α, β;

Let $\mathcal{L}_t$ be the feasible set formed by the t queries/answers

if with $\mathcal{L}_t$ the linear equation system has unique solution **then** output DENY; **return**; **endif**

else if there is a x_i that can be uniquely determined **then**

  if $(MAX - x_i) > d_{tr}$ AND $(MAX - max_t) > d_{tr}$ **then** output a_t; **return**; **endif**

  else if $(MAX - x_i) \leq d_{tr}$ OR $(MAX - max_t) \leq d_{tr}$ **then** output DENY; **return**; **endif**

endif else output a_t;

I showed in the Section 4.6 of my dissertation that the proposed online auditors ensure, in the full disclosure model, the privacy of the maximum value. $|x^{opt} - MAX|$ denotes the absolute distance between x^{opt} and MAX, while max_t is the maximum of the first t answers, and d_{tr} is the security treshold, defining that how close we allow the querier to the value of MAX.

5.3.2 Simulatable auditor$^{max}_{avg}$ in the partial disclosure model

Thesis 3.3. *I proposed an efficient simulatable Auditor$^{max}_{avg}$ query auditors for the case of probabilistic disclosure model. My proposed auditor is based on the application of the efficient random sampling approach [26], as well as the Chernoff bound and the Union bound, known in statistical theory. I showed that the proposed auditor is simulatable, and hence it provides the privacy of the value of MAX in the probabilistic disclosure model [Th09, Th10].*

In this subsection, I propose a simulatable auditor Auditor$^{max}_{avg}$ in the partial disclosure model. This subsection is the excerpt of the Section 4.7 of my dissertation. Consider an arbitrary data set $X = \{x_1, \ldots, x_n\}$, in which each x_i is chosen independently according to the same distribution $\mathcal{H}$ on $(-\infty, \infty)$. Let $\mathcal{D} = \mathcal{H}^n$ denote the joint distribution.

The predicate λ-Safe and AllSafe are a bit differ from the traditional definitions for individual values, since I am considering the maximum of n values instead of single values. Hence, the definitions are modified as follows:

Definition 4. *The sequence of queries and answers, $q_1, \ldots, q_t$, $a_1, \ldots, a_t$ is said to be λ- Safe with respect to an interval $I \subseteq [\alpha, \beta]$ if the following Boolean predicate evaluates to 1:*

$Safe_{\lambda,I}(q_1, \ldots, q_t, a_1, \ldots, a_t) =$

$$\begin{cases} 1 & \text{if } 1/(1+\lambda) \leq \frac{P_{\mathcal{G}^t_{post}}(MAX \in I | \wedge_{j=1}^{t}(avg(Q_j) = a_j))}{Pr_{\mathcal{G}_{max}}(MAX \in I)} \leq (1+\lambda) \\ 0 & \text{otherwise} \end{cases}$$

where $\mathcal{G}^t_{post}$ is the distribution of the posteriori probability, and $\mathcal{G}_{max}$ is the distribution of MAX. The definition of AllSafe is then given over all ω-significant intervals J of $[\alpha, \beta]$. Here the notion of ω-significant interval is defined over the maximum value instead of individual values: An interval J is ω-significant if $P_{\mathcal{G}_{max}}(\text{MAX} \in J) \geq \frac{1}{\omega}$. I only care about the probability changes with respect to the so called *significant intervals*.

Definition 5. $AllSafe_{\lambda,\omega}(q_1, \ldots, q_t, a_1, \ldots, a_t) =$

$$\begin{cases} 1 & \text{if } Safe_{\lambda,J}(q_1, \ldots, q_t, a_1, \ldots, a_t) = 1, \forall\ J \\ 0 & \text{otherwise} \end{cases}$$

For the probabilistic disclosure model, in the following I provide the definition of randomized auditor.

Definition 6. *A randomized auditor is a randomized function of queries $q_1, \ldots, q_t$, the data set X, and the probability distribution D that either gives an exact answer to the query q_t or denies the answer.*

Next I introduce the notion of (λ, ω, T)-privacy game and (λ, δ, ω, T)-private auditor. The (λ, ω, T)-privacy game between an attacker and an auditor, where in each round t (for up to T rounds):

1. The attacker (adaptively) poses a query $q_t = (Q_t, f_t)$.
2. The auditor decides whether to allow qt or not. The auditor replies with $a_t = f_t(Q_t)$ if q_t is allowed, and denies otherwise.
3. The attacker wins if $\text{AllSafe}_{\lambda,\omega}(q_1, \ldots, q_t, a_1, \ldots, a_t) = 0$.

Definition 7. *I say that an auditor is (λ, δ, ω, T)-private if for any attacker A*

$$P\{A \text{ wins the } (\lambda, \omega, T)\text{-privacy game}\} \leq \delta.$$

The probability is taken over the randomness in the distribution $\mathcal{D}$ and the coin tosses of the auditor and the attacker.

My proposed probabilistic $\text{Auditor}_{avg}^{max}$ auditor is implemented by the Algorithm 3 and Algorithm 4 in Section 4.7 of my dissertation. I showed that my proposed probabilistic $\text{Auditor}_{avg}^{max}$ auditor satisfies the Definition 7 above.

6 Conclusion

In this Ph.D Thesis, I focus on security problems in different application fields of wireless sensor networks. I proposed formal and automated verification methods for analyzing the security of protocols designed for WSNs, as well as query auditing algorithms for protecting sensitive information in statistical databases. My dissertation is composed of three theses groups, which are related to three different research topics.

The first theses group contains the following main contributions: I proposed a variant of process algebra called the *sr*-calculus, which provides expressive syntax and semantics for analyzing at the same time (i.) cryptographic primitives and operations, (ii.) the nature of broadcast communication, and (iii.) the specification of node's neighborhood in wireless medium, which are required for verifying secure routing protocols. I proposed a systematic and exhaustive proof technique for analyzing routing protocols with the *sr*-calculus.

In addition, I proposed a fully automatic verification method, called *sr*-verif, for secured ad-hoc network routing protocols, which is based on logic and a backward reachability approach. My method

has a clear syntax and semantics for modelling secure routing protocols, and handles arbitrary network topologies. Using my verification methods, I proved that the well-known routing protocols (DSR, SRP, Ariadne, endairA) are vulnerable to route forging attacks.

My main contributions in the second theses group are the following: I proposed a probabilistic timed calculus for cryptographic protocols, called $crypt^{prob}_{time}$, and demonstrated how to use it for proving security or vulnerability of protocols. To the best of my knowledge, this is the first such process calculus that supports an expressive syntax and semantics, real-time, probabilistic, and cryptographic issues at the same time. Hence, it can be used to verify systems that involve these three properties. For demonstration purposes, I applied $crypt^{prob}_{time}$ to prove that both of the two previously proposed protocols, DTSN and SDTP, are vulnerable to the EAR flag setting attack, and the tricky sandwich attack. Taking into account the security holes in DTSN and SDTP, I proposed a new secured WSN transport protocol, called SDTP$^+$, and proved that the discussed attacks against DTSN and SDTP do not work in SDTP$^+$.

In addition, I proposed an automatic verification method, based on the PAT process analysis toolkit for this class of protocols, and used it to verify the security of the DTSN and SDTP protocols. To the best of my knowledge, PAT has not been used to verify WSN transport protocols before, however, I showed that it is well-suited for this purpose.

Finally, my main theses in the third theses group is composed of the following results: I defined a novel setting for query auditing, where instead of detecting or preventing the disclosure of individual sensitive values, I want to detect or prevent the disclosure of aggregate values in the database. As a specific instance of this setting, in the dissertation, I studied the problem of detecting or preventing the disclosure of the maximum value in the database, when the querier is allowed to issue average queries to the database. I proposed efficient off-line and on-line query auditors for this problem in the full disclosure model, and an efficient simulatable on-line query auditor in the partial disclosure model.

References

[1] M. Abadi and A. Gordon. A calculus for cryptographic protocols: the Spi calculus. Technical Report SRC RR 149, Digital Equipment Corporation, Systems Research Center, January 1998.

[2] G. Acs, L. Buttyan, and I. Vajda. Provable security of on-demand distance vector routing in wireless ad hoc networks. In *In In Proceedings of the Second European Workshop on Security and Privacy in Ad Hoc and Sensor Networks (ESAS 2005*, pages 113–127, 2005.

[3] G. Acs, L. Buttyan, and I. Vajda. Provably secure on-demand source routing in mobile ad hoc networks. In *IEEE Transactions on Mobile Computing*, volume 5, 2006.

[4] G. Acs, L. Buttyan, and I. Vajda. The security proof of a link-state routing protocol for wireless sensor networks. In *IEEE Workshop on Wireless and Sensor Networks Security*, 2007.

[5] Charu C. Aggarwal and Philip S. Yu, editors. *Privacy-Preserving Data Mining - Models and Algorithms*, volume 34 of *Advances in Database Systems*. Springer, 2008.

[6] I. F. Akyildiz, T. Melodia, and K. R. Chowdhury. A survey on wireless multimedia sensor networks. *Computer Networks*, 51(4):921–960, 2007.

[7] Todd R. Andel and Alec Yasinsac. Automated evaluation of secure route discovery in manet protocols. In *SPIN '08: Proceedings of the 15th international workshop on Model Checking Software*, pages 26–41, 2008.

[8] J. Bengtsson and F. Larsson. Uppaal a tool for automatic verification of real-time systems. *Technical Report,Uppsala University, (96/67)*, 1996.

[9] Bruno Blanchet. Automatic Proof of Strong Secrecy for Security Protocols. In *IEEE Symposium on Security and Privacy*, pages 86–100, Oakland, California, May 2004.

[10] L. Buttyan and L. Csik. Security analysis of reliable transport layer protocols for wireless sensor networks. In *Proceedings of the IEEE Workshop on Sensor Networks and Systems for Pervasive Computing (PerSeNS)*, pages 1–6, Mannheim, Germany, March 2010.

[11] L. Buttyan and A. M. Grilo. A Secure Distributed Transport Protocol for Wireless Sensor Networks. In *IEEE International Conference on Communications*, pages 1–6, Kyoto, Japan, June 2011.

[12] L. Buttyán and I. Vajda. Towards provable security for ad hoc routing protocols. In *SASN '04: Proceedings of the 2nd ACM workshop on Security of ad hoc and sensor networks*, pages 94–105, 2004.

[13] Francis Chin. Security problems on inference control for sum, max, and min queries. *J. ACM*, 33:451–464, May 1986.

[14] Francis Chin and Gultekin Ozsoyoglu. Auditing for secure statistical databases. In *Proceedings of the ACM '81 conference*, pages 53–59, New York, NY, USA, 1981.

[15] D. Coppersmith and M. Jakobsson. Almost optimal hash sequence traversal. In *Fourth Conference on Financial Cryptography*, pages 102–119, Southampton, Bermuda, March 2002.

[16] PedroR. D'Argenio and Ed Brinksma. A calculus for timed automata. In Bengt Jonsson and Joachim Parrow, editors, *Formal Techniques in Real-Time and Fault-Tolerant Systems*, volume 1135 of *Lecture Notes in Computer Science*, pages 110–129. Springer Berlin Heidelberg, 1996.

[17] C. Fournet and M. Abadi. Mobile values, new names, and secure communication. In *In Proceedings of the 28th ACM Symposium on Principles of Programming, POPL'01*, pages 104–115, 2001.

[18] Jens Chr. Godskesen. A calculus for mobile ad hoc networks. In *COORDINATION*, pages 132–150, 2007.

[19] Jens Chr. Godskesen. A calculus for mobile ad-hoc networks with static location binding. *Electron. Notes Theor. Comput. Sci.*, 242(1):161–183, 2009.

[20] Jean Goubault-Larrecq, Catuscia Palamidessi, and Angelo Troina. A probabilistic applied picalculus. In Zhong Shao, editor, *Programming Languages and Systems*, volume 4807 of *Lecture Notes in Computer Science*, pages 175–190. Springer Berlin Heidelberg, 2007.

[21] Y. C. Hu, D. B. Johnson, and A. Perrig. Secure efficient distance vector routing in mobile wireless ad hoc networks. *Ad Hoc Networks*, 1(1):175–192, July 2003.

[22] Y.-C. Hu, A. Perrig, and D. B. Johnson. Ariadne: a secure on-demand routing protocol for ad hoc networks. *Wirel. Netw.*, 11(1-2):21–38, 2005.

[23] Krishnaram Kenthapadi, Nina Mishra, and Kobbi Nissim. Simulatable auditing. In *In ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems, PODS*, pages 118–127, 2005.

[24] M. Kwiatkowska, G. Norman, and D. Parker. PRISM 4.0: Verification of probabilistic real-time systems. In G. Gopalakrishnan and S. Qadeer, editors, *Proc. 23rd International Conference on Computer Aided Verification (CAV'11)*, volume 6806 of *LNCS*, pages 585–591. Springer, 2011.

[25] Ruggero Lanotte, Andrea Maggiolo-Schettini, and Angelo Troina. Weak bisimulation for probabilistic timed automata. In *PROC. OF SEFM03, IEEE CS*, pages 34–43. Press, 2003.

[26] László Lovász and Santosh Vempala. The geometry of logconcave functions and sampling algorithms. *Random Struct. Algorithms*, 30:307–358, May 2007.

[27] B. Marchi, A. Grilo, and M. Nunes. DTSN - distributed transport for sensor networks. In *Proceedings of the IEEE Symposium on Computers and Communications*, pages 165–172, Aveiro, Portugal, July 2007.

[28] John D. Marshall, II, and Xin Yuan. An analysis of the secure routing protocol for mobile ad hoc network route discovery: Using intuitive reasoning and formal verification to identify flaws. Technical report, THE FLORIDA STATE UNIVERSITY, 2003.

[29] R. C. Merkle. Protocols for Public Key Cryptosystems. In *Symposium on Security and Privacy*, pages 122–134, California, USA, April 1980.

[30] Robin Milner, Joachim Parrow, and David Walker. A calculus of mobile processes, parts i and ii. *Inf. Comput.*, 100(1):1–77, September 1992.

[31] Shubha U. Nabar, Bhaskara Marthi, Krishnaram Kenthapadi, Nina Mishra, and Rajeev Motwani. Towards robustness in query auditing. In *International Conference on Very Large Data Bases (VLDB)*, pages 151–162, 2006.

[32] P. Papadimitratos and Z.J. Haas. Secure routing for mobile ad hoc networks. In *In Proceedings of the SCS Communication Networks and Distributed Systems Modeling and Simulation Conference*, pages 1–13, 2002.

[33] A. Perrig, R. Szewczyk, V. Wen, D. Culler, and J.D. Tygar. SPINS: security protocols for sensor networks. In *ACM MobiCom*, Rome, Italy, July 2001.

[34] Mayank Saksena, Oskar Wibling, and Bengt Jonsson. Graph grammar modeling and verification of ad hoc routing protocols. In *Proceedings of the Theory and practice of software, 14th international conference on Tools and algorithms for the construction and analysis of systems*, TACAS'08/ETAPS'08, pages 18–32, Berlin, Heidelberg, 2008. Springer-Verlag.

[35] Steve Schneider. *Concurrent and Real Time Systems: The CSP Approach*. John Wiley & Sons, Inc., New York, NY, USA, 1st edition, 1999.

[36] Anu Singh, C. R. Ramakrishnan, and Scott A. Smolka. A process calculus for mobile ad hoc networks. *Sci. Comput. Program.*, 75(6):440–469, 2010.

[37] Jun Sun, Yang Liu, and Jin Song Dong. Model checking csp revisited: Introducing a process analysis toolkit. In *In ISoLA 2008*, pages 307–322. Springer, 2008.

[38] O. Wibling, J. Parrow, and A. Pears. Automatized verification of ad hoc routing protocols. *Formal Techniques for Networked and Distributed Systems FORTE*, pages 343–358, 2004.

[39] J. Yicka, B. Mukherjeea, and D. Ghosal. Wireless sensor network survey. *Computer Networks*, 52(12):2292–2330, Aug. 2008.

Publications

[Th01] Levente Buttyán and Ta Vinh Thong. Biztonsgi API analízis a spi-kalkulussal. *Híradástechnika*, LXII(8):16–21, July 2007. (in Hungarian).

[Th02] Levente Buttyán and Ta Vinh Thong. Security API analysis with the spi-calculus. *Híradástechnika*, LXIII(1):43–49, April 2008.

[Th03] Frank Kargl, Panagiotis Papadimitratos, Levente Buttyán, Michael Mueter, Elmar Schoch, Bjorn Wiedersheim, Ta Vinh Thong, Gorgio Calandriello, Albert Held, Antinio Kung, and Jeanpierre Hubaux. Secure Vehicular Communication Systems: Implementation, Performance, and Research Challenges. *IEEE Communications Magazine*, 46(11):110–118, November 2008.

[Th04] Levente Buttyán, Gábor Pék, Ta Vinh Thong. Consistency verification of stateful firewalls is not harder than the stateless case. *Infocommunications Journal*, LXIV(2009/2-3):2–8, March 2009.

[Th05] Levente Buttyán and Ta Vinh Thong. Formal verification of secure ad-hoc network routing protocols using deductive model-checking. In *IFIP Wireless and Mobile Networking Conference (WMNC 2010)*, IFIP, pp 1–6, Budapest, 2010.

[Th06] Levente Buttyán and Ta Vinh Thong. Formal verification of secure ad-hoc network routing protocols using deductive model-checking. *Periodica Polytechnica Electrical Engineering Journal*, Vol. 1245, pp 1–20, 2011.

[Th07] Ta Vinh Thong and Levente Buttyán. On automating the verification of secure ad-hoc network routing protocols. *Telecommunication Systems Journal, Springer*, ISSN 1572-9451, pp 1–28, August 2011.

[Th08] Ta Vinh Thong. Formal verification of secure ad-hoc network routing protocols using deductive model-checking. *Cryptology Eprint Archive, IACR*, 1–77, March 2012.

[Th09] Ta Vinh Thong and Levente Buttyán. Query Auditing for Protecting Max/Min Values of Sensitive Attributes in Statistical Databases. In *9th International Conference on Trust, Privacy, Security in Digital Business (Trustbus 2012)*, Springer LNCS, Volume 7449, pp 192–206, July 2012, Wien.

[Th10] Ta Vinh Thong and Levente Buttyán. Query Auditing for Protecting Max/Min Values of Sensitive Attributes in Statistical Databases. *Technical Report, pp. 1–15, 2012, CrySyS Lab., BME*

[Th11] Amit Dvir, Levente Buttyán, Ta Vinh Thong. SDTP+: Securing a Distributed Transport Protocol for WSNs using Merkle Trees and Hash Chains. In *IEEE International Confenrence on Communications (ICC 2013), Communication and Information Systems Security Symposium*, IEEE, pp 1–6, June 2013, Budapest.

[Th12] Ta Vinh Thong and Amit Dvir. On Formal and Automated Security Verification of WSN Transport Protocols. *Cryptology Eprint Archive, IACR*, 1–81, January 2013.

[Th13] Ta Vinh Thong and Levente Buttyán and Amit Dvir. On Formal and Automated Security Verification of WSN Transport Protocols. *International Journal of Distributed Sensor Networks, Hindawi*, 1–28. (submitted on May 3, 2013).