
This is the author’s final accepted version.

There may be differences between this version and the published version. You are advised to consult the publisher’s version if you wish to cite from it.

http://eprints.gla.ac.uk/222219/

Deposited on: 17 August 2020

Enlighten – Research publications by members of the University of Glasgow
http://eprints.gla.ac.uk
An Efficient Scheme to Secure Data Provenance in Home Area Networks

Zhaohui Tang  
School of Sciences  
University of Southern Queensland, Australia  
Zhaohui.Tang@usq.edu.au

Sye Loong Keoh  
School of Computing Science  
University of Glasgow, UK  
SyeLoong.Keoh@glasgow.ac.uk

Abstract—The 5G technology has raised new security issues in the new communication paradigm as there are more devices in the network. Internet-of-Things devices, including smart home devices, will be allowed to connect, communicate, and share data faster than ever. In the context of a Home Area Network (HAN) where smart home appliances are monitored and controlled in real-time, the explosive growth of networked devices poses an elevated security threat to data provenance. This paper focuses on data provenance issues in a HAN within a smart metering infrastructure. We propose a comprehensive yet efficient solution to guarantee that the reported energy usage is collected from the real appliance as claimed, at the designated location, and that it reflects the real consumption by the appliance.

I. INTRODUCTION

The world is transitioning towards a smart grid enabled ecosystem in which renewable energy can be integrated with power distribution and generation efficiently to address growing energy needs. As compared with the traditional grid system, smart grid enables the customers to have a better control of their electricity usage, and provides more efficient management of their daily consumption by reporting real-time energy consumption and supporting fine-grained customer monitoring and control.

While smart grid has been extensively adopted in the world, the security of smart grid [1] has drawn a big attention to researchers from both academic and industry due to massive challenges posed by cyber security threats [12].

A. Home Area Network

A home area network (HAN) is a subsystem within a smart grid system which extends smart grid capabilities into the home using different networking protocols. As shown in Fig. 1, HAN is typically found in consumer premise and it connects home (smart) appliances such as thermostats, refrigerators and other electrical devices to a smart meter. A HAN typically relies on a wireless link such as WiFi, ZigBee to communicate with the smart meter via a gateway. The gateway also acts as a bridge between the smart meter and the smart appliances in consumer’s home. The gateway collects energy usage data, network status from the utility for display to the consumer on the gateway itself. Moreover, the gateway also forwards demand/response and energy-pricing signals to the smart appliances for their information.

However, this also makes HAN one of the most vulnerable systems in the smart grid because the wireless communication is inherently insecure. It is possible for an attacker to intercept and monitor network traffic to gain sensitive information across the wireless communication.

B. Threat Modeling and Security Goals

Some real-life provenance attacks have been reported where a malicious user intruded into a HAN and deceived the smart grid system with fraudulent energy usage report. In October 2014, smart meters in Spain were hacked to cut power bills [4]. Exploiting the leaked encryption keys and unique identifier associated with the smart meter, the attackers played the role as home appliances and sent the utility energy consumption reports which were discovered to be under-reported. In another occasion, two German researchers demonstrated security vulnerabilities in a German energy company’s smart metering service, where attackers were able to rewrite electricity bills to a negative energy consumption rate of -106610 kWh [5]. In fact there are many attacks that can be launched to commit energy fraud in a home area network setting due to the potential vulnerabilities at smart plugs (i.e., a device that keeps track of the electricity usage of a power outlet):

- Firstly, a smart plug’s device identity could be spoofed by an adversary impersonating other smart plugs to send out fraudulent energy reading.
- Secondly, a compromised smart plug may alter the energy reading before sending it to the smart meter, which
breaches the data source authenticity as the measurement itself has already been tampered with at the smart plug prior to data transmission.

- Thirdly, a Man-in-the-Middle attacker could intercept and manipulate the data routed from smart plugs to the smart meter which implies data integrity breach.
- Finally, the location of a smart plug could be spoofed, in an extreme case, to a smart plug from another household, for example, to a neighbour household who has access to the victim household’s home Wifi network.

The aforementioned four types of attacks can all result in fraudulent energy usage reports.

C. Security Goals

Based on the four types of attacks described above, one can obviously see an urgent need to protect data provenance in a smart grid home area network. Henceforth, the primary aim of this work is to develop an efficient security solution ensuring that the reported energy usage is collected from the specific appliance as claimed, and reflects the real consumption of the energy. Specifically, we aim to achieve the following security goals in a HAN (we assume each smart appliance is associated with one smart plug for drawing the appliance’s electricity reading):

- Source Identity Authenticity which ensures that the energy consumption data is originated from the smart plug as claimed (i.e., not impersonated).
- Source Data Authenticity which means that the energy consumption measured by the smart plug on each appliance reflects the real consumption (i.e., no under-report or over-report).
- Data Integrity which guarantees that the data transmitted from the smart plug to the smart meter is not tampered with (i.e., no alteration when routed to the smart meter).
- Location Authenticity - ensures that the energy data is collected from the location of the power outlet as claimed (i.e., no location spoofing).

Out of the four security goals, three are directly linked to data provenance: source identity authenticity, source data authenticity, location authenticity. The fourth security goal, i.e., data integrity, is attained as a by-product of our proposed solution.

D. Our Contributions

While security goals are achieved, the efficiency of the solution is worth discussing due to the resource-constrained IoT devices within our context of HAN system. We therefore aim to propose a scheme ensuring the aforementioned security goals are achieved with high efficiency.

Our contribution includes a comprehensive yet efficient solution for a HAN system to accomplish all the security goals listed in Section I-C.

E. Organization

This paper is organized as follows. In Section II we introduce smart grid and work related to HAN security. Section II also includes preliminary knowledge that is needed for understanding our solution which is detailed in Section III. Security analysis of our proposed solution is presented in Section IV. We conclude the paper with future work and discussion in Section V.

II. RELATED WORK AND PRELIMINARY

A. Related Work

Aman et al. [6] proposed to use Physical Unclonable Function (PUF) to protect the identity of an IoT device and the wireless link’s Received Signal Strength Indicator (RSSI) values for preserving the device’s location authenticity. However, this research did not have any implementation or analysis to demonstrate the practicality of the approach, moreover, it did not investigate the source data authenticity issue.

Jiang et al. [7] introduced a classification-based machine learning approach to classify the load profiles of customer’s abnormal behaviour for detecting energy-theft suspects. This was complemented by a state-based detection method that monitors and correlates the various events derived from the smart grid network and thus vastly improves the detection rate. Yet, none of these detection schemes addressed fundamental issues such as hardware tampering where the energy consumption data could be manipulated at its source. i.e., smart plugs.

Another line of research focused on hardware tempering detection by generating and verifying tamper-evident proofs such as time-stamps, digital signatures of energy consumption data drawn by the IoT devices. Nevertheless, another critical provenance factor, namely, location authenticity, was not discussed in this series of literatures.

A recent work [15] has addressed all the security issues as defined above, however the scheme was digital signature based and thus required lengthy keys as well as intensive computations at the IoT devices. A more efficient scheme is needed for practicality consideration.

To summarize, there is a lack of practical scheme that achieves all the security goals defined in Section I-C. Motivated by this, our work aims to design a new security scheme that is more efficient and thus more practical than the existing work [15].

B. Preliminary

This section prepares some fundamental knowledge which will be used in later sections.

1) Monotone Span Program and Access Structure [3] [8] [9]:

Definition 1: (Monotone Span Program (MSP)) A Monotone Span Program (MSP), denoted by \( M \), is a quadruple \((F, M, \epsilon, \phi)\) where \( F \) is a finite field, \( M \in \mathbb{F}^{m \times d} \) is a full-rank matrix for some \( m \) and \( d \leq m \), \( \epsilon \in \mathbb{F}^d \) is an arbitrary non-zero vector called the target vector, and \( \phi : [m] \to F \) is a surjective labelling map of rows to parties. The size of \( M \) is defined to be \( m \), the number of rows of the matrix \( M \).

The target vector \( \epsilon \) can be an arbitrary non-zero vector, in this work we take a typical value of \( \epsilon \), that is, \( \epsilon = 1 = (1, 0, \cdots, 0) \). For a matrix \( M \in \mathbb{F}^{m \times d} \), we use \( M^T \) to denote its transpose.
Access structures are used in the study of security systems where multiple parties need to work together to obtain a resource. Groups of parties that are granted access are called qualified. The set of all such qualified sets is called the access structure of the system.

Definition 2: An Monotone Span Program (MSP) is said to compute an access structure \( \tau \) if it holds that \( Q \in \tau \) if and only if there \( \exists \lambda \in \mathbb{F}^m \) such that \( M_Q^T \cdot \lambda = \epsilon \). In other words, a set \( Q \) of participants can reconstruct the target vector \( \epsilon \) if and only if \( Q \) is qualified.

2) Message Authentication Codes: MACs [10], [11], one of the most utilized symmetric-key cryptosystem, can be used by the receiver to check that the message was indeed from the intended sender and not tampered throughout the transmission. A conventional MAC consists of three steps as below:

- **Key Generation:** For a given security parameter, a MAC key \( k \) is generated for the sender and receiver. Typically \( k \) is randomly chosen and of appropriate length.

- **MAC Generation:** Before sending out a message \( m \), the sender generates a tag \( t \) based on the key \( k \), message \( m \) and a chosen algorithm.

- **MAC Verification:** Upon receiving a message \( m \) and its tag \( t \), based on the MAC verification algorithm (derived from the aforementioned MAC generation algorithm), a receiver/verifier will return a value of True or False. Where True means that the message \( m \) is proven to be not tampered during the transmission from the sender to the receiver/verifier.

3) Multi-participant MAC Scheme: We propose a multi-participant MAC scheme, enabling multiple participants to jointly generate a MAC that can be verified by a receiver. This is similar to Threshold Cryptography [14] scheme, but using MAC which is symmetric-key based and less computational intensive, as compared to digital signature. A secure MAC, HMAC [19] is adopted by our proposed MAC scheme.

- **MAC-Key-Share-Generation** \((M, i, k)\): Given a randomly chosen key \( k \in \mathbb{F} \), a monotone span program \( M \) and a user’s ID \( i \), the receiver generates and sends securely a MAC key share \( k_i \) to Participant \( i \) by the formula \( k_i = M_i^T \cdot \epsilon \cdot * k \).

- **Partial-MAC-Generation** \((m, k)\): A message or MAC tag, \( m \) and \( k \), MAC=HMAC(m, k) [19].

- **Full-MAC-Generation** \((Q, t)\): For a set \( Q \) of participants, where \( Q = \{P_{i_1}, ..., P_{i_{|Q|}}\} \) and their corresponding partial MAC tags \( t_Q = \{t_{P_{i_1}}, ..., t_{P_{i_{|Q|}}}\} \). The full MAC \( t \) can be generated as: \( t = t_{P_{i_1}} \oplus t_{P_{i_2}} \oplus ... \oplus t_{P_{i_{|Q|}}} \).

- **Full-MAC-Verification** \((m, t, Q, k, M)\): The verifier should have access to the monotone span program \( M \) and key \( k \) that are used in the stage of MAC-Key-Share-Generation. Given a message \( m \) and a MAC tag \( t \), the participants’ identity \( Q = \{P_{i_1}, ..., P_{i_{|Q|}}\} \), the verifier checks the integrity of a message-MAC pair by checking whether the following equation holds:

\[ t = HMAC(m, M_i^T \cdot \epsilon \cdot * k) \oplus ... \oplus HMAC(m, M_{i_{|Q|}}^T \cdot \epsilon \cdot * k). \]

If the equation holds, then a True value is returned and the message-MAC pair is accepted, otherwise a False value is returned with the message-MAC pair being rejected.

### III. Proposed Solution

This paper proposes to adopt an integrated device and implement a 2-phase communication protocol, in order to fulfill all the security goals as defined in Section I-C. We briefly describe our approach in Section III-A. The integrated device is unveiled in Section III-B, while the 2-Phase communication protocol is detailed in Section III-C (Commissioning Phase) and Section III-D (Operation Phase).

#### A. Our Approach

The main advantage of our approach is that most of the existing scheme adopts an asymmetric key-based digital signature [15] to provide authenticity, integrity and data provenance. This paper improves efficiency by adopting a symmetric key based Message Authentication Codes (MACs) approach.

For ensuring the identity authenticity at smart plug, a tag (i.e., a MAC tag) must be generated by the smart plug with its own private credential (i.e., MAC key share which will be defined later).

The most challenging goal to achieve is source data authenticity. Similar to a previous work [15], we exploits a novel magnetic sensor [13] as a redundant measuring tool to simultaneously but independently collect a smart appliance’s energy. The energy reading drawn by the magnetic sensor (say, \( v_1 \)), will be used to cross check with the reading logged by the smart meter (say, \( v_2 \)). We assume at least one out of the two devices (smart plug, magnetic sensor) is trusted, and argue that source data authenticity at smart plug is attained if \( v_1 = v_2 \), otherwise breached.

The data authenticity and integrity at smart plug are essential, which are ensured by a MAC tag that can only be generated by the smart plug with its private MAC key share. On the other hand, since we rely on the magnetic sensor’s cross-checking, it is also important to ensure the data sent by magnetic sensor is authentic; a MAC tag and MAC key share are also needed at the magnetic sensor.

In this paper we assume the receiver (i.e., smart meter) is trusted while one of the two senders (smart plug, magnetic sensor) could be compromised. For gathering and verifying the smart plug’s location information, we propose to integrate a third sender, i.e., a Bluetooth device which is assumed not to collude with the smart plug or magnetic sensor. With all these security assumptions, we propose to let the smart meter choose a random key as his MAC key, based on which, distinct MAC key shares are produced and distributed to the three different senders: the smart plug, magnetic sensor, and Bluetooth device.

We allow the the smart plug (or magnetic sensor) and Bluetooth device to reconstruct the MAC key by combining their MAC key shares. In other words, the qualified sets to obtain the
original MAC key consist of: \{smart plug, bluetooth device\}, \{magnetic sensor, bluetooth device\}.

**B. The Integrated Device: SSP**

A device **Secure Smart Plug (SSP)** was proposed in [15], which is essentially an integration of three individual components: Smart Plug, Bluetooth Device, Magnetic Sensor. The SSP is typically located in an indoor environment to connect home appliances and reports the energy consumption to \(SM\). Fig. 2 illustrates the composition of an SSP and a use case of applying SSP in a HAN environment to gather and report the electricity consumption of a smart home appliance (for example, a smart refrigerator) to a smart meter. Here is a brief description of the entities included in the SSP and HAN:

- \(SM\) simulates the real smart meter (server) in a HAN.
- Smart Plug (SP) is used to measure the energy consumption of a smart home appliance.
- Bluetooth Device (BT) is required to identify and determine the location information of the SSP as this paper concerns the location authenticity.
- Magnetic Sensor (MS) is used as an alternative medium to draw the energy usage of a smart home appliance.

**C. Commissioning of SSP**

In this phase, the SSP in our use case are first commissioned with necessary security parameters which are required for the Operation phase later. There are two parts of commissioning: \*Key Deployment and MAC Key Share Distribution, Location Registration*.

1) **Key Deployment and MAC Key Share Distribution:**

As reasoned in Section III-A, the system first provision a MAC key share for \(SP, MS\) and \(BT\). These key shares will be used for generating their respective (partial) MACs. In this case, we define the participant set \(P\) in our use case as \(P = \{SP, MS, BT\}\). Assume there exists an MSP \(M = (F, M, \epsilon, \phi)\) computing \(\tau\), where \(\tau\) is the access structure:

\[
\tau = \{\{SP, BT\}, \{SP, MS\}\}.
\]

With \(M\), \(SM\) executes **MAC Key Share Generation** algorithm in the commissioning phase implementing the algorithm \((M, i, k)\). \(SM\) is also responsible for distributing the resultant three MAC key shares to the three participants: \(k_{SP}\) to SP, \(k_{BT}\) to BT, and \(k_{MS}\) for MS.

2) **Location Registration:** Our scheme relies on Bluetooth Low Energy (BLE) beacons to be deployed in order to determine the location of the SSP.

The SSP indoor location can be accurately determined on the basis of three known reference points from the BLE beacons using Trilateration Scheme as shown in Fig. 3. More precisely, using RSSI filtering methodology [18], the SSP’s location information is identified based on its distances from the three BLE beacons whose positions are known and fixed in an indoor environment. This location information will be used in the future to verify whether the three-in-one SSP device has been relocated when its built-in components \(SP\) and \(MS\) are being used for reporting energy readings. The verification will be mainly conducted by the third built-in component, i.e. \(BT\) from the SSP.

**D. Operation**

During the operation phase, energy consumption reading is reported to \(SM\) periodically. With our proposed multi-participant MAC Scheme, the \(SM\) is able to verify the integrity and authenticity of the data it receives. SM is also ensured on the location authenticity of the SSP (and thus SP).

There are three steps involved in the Operation Phase. One can see that the following three algorithms from Section II-B3 are used in the Operation Phase: **Partial-MAC-Generation, Full-MAC-Generation, Full-MAC-Verification**.

1) **Step 1:** Extract Energy Reading, Generate Partial MACs, and Verify \(SP\)’s Location.

The three components of \(SP, MS\) and \(BT\) execute Step 1 concurrently with details shown below.

- **Step 1-SP:** There are three sequential activities implemented by SP:
  - \(SP\) extracts the home appliance’s energy reading \(r_{SP}\) and forms its message \(m_{SP} = (SP, r_{SP}, t_{SP})\) where \(SP\) represents its identity, and \(t_{SP}\) records the time when the energy reading \(r_{SP}\) is drawn.
- SP further generates its partial MAC, denoted as $PM_{SP}$, based on $m_{SP}$ and its MAC key share $k_{SP}$, by running:

$$PM_{SP} = \text{Partial-MAC-Generation}(m_{SP}, k_{SP}).$$  \hspace{1cm} (1)

- Finally, SP sends BT the following concatenated message:

$$(m_{SP}, PM_{SP}).$$

**Step 1-MS:** Similar to SP, MS in Step 1 is expected to complete the following three tasks:
- MS extracts the home appliance’s energy reading $r_{MS}$ and forms its message $m_{MS} = (MS, r_{MS}, t_{MS})$ where MS represents its identity, and $t_{MS}$ records the time when the energy reading $r_{MS}$ is drawn.
- MS further generates its partial MAC, denoted as $PM_{MS}$, based on $m_{MS}$ and its MAC key share $k_{MS}$, by running

$$PM_{MS} = \text{Partial-MAC-Generation}(m_{MS}, k_{MS}).$$  \hspace{1cm} (2)

- Finally, MS sends the following concatenated message to BT:

$$(m_{MS}, PM_{MS}).$$

**Step 1-BT:** When $SP$ is extracting the energy reading, $BT$ is responsible for verifying whether $SP$'s location remains the same as registered in Commissioning Phase. If it is the same, $BT$ will continue with **Step 2 Generate Full MACs** (as described next); otherwise it will end the protocol and notify the system that $SP$ has been relocated and hence not being able to authenticate its location.

2) **Step 2:** Generate Full MACs. Note that $BT$ has received messages from both $SP$ ($m_{SP}$ in Step 1-SP) and $MS$ ($m_{MS}$ in Step 1-MS), based on which $BT$ is able to generate partial MACs for both messages with $BT$’s privately-possessed MAC key share $k_{BT}$. $BT$ will further generate full MACs based on the existing partial MACs. Details can be seen below:

- For $m_{SP}$, $BT$ first generates its partial MAC $PM_{BT4SP}$ as:

$$PM_{BT4SP} = \text{Partial-MAC-Generation}(m_{SP}, k_{BT}).$$  \hspace{1cm} (3)

Subsequently, $BT$ generates full MAC for $m_{SP}$, denoted by $FM_{SP}$, using

$$FM_{SP} = PM_{SP} \oplus PM_{BT4SP},$$  \hspace{1cm} (4)

and sends the (message, full MAC) pair $(m_{SP}, FM_{SP})$ to Smart Meter (SM).

- For $m_{MS}$, $BT$ generates its partial MAC $PM_{BT4MS}$ as:

$$PM_{BT4MS} = \text{Partial-MAC-Generation}(m_{MS}, k_{BT}).$$  \hspace{1cm} (5)

and then generates the full MAC for $m_{MS}$, denoted by $FM_{MS}$, via

$$FM_{MS} = PM_{MS} \oplus PM_{BT4MS},$$  \hspace{1cm} (6)

and sends the (message, full MAC) pair $(m_{MS}, FM_{MS})$ to Smart Meter (SM).

3) **Step 3:** Verify Data Integrity & Source Data Authenticity.

Two security goals are achieved at $SM$ in this step: (1) Data integrity of messages $m_{SP}, m_{MS}$ which are originated from $SP$ and $MS$ respectively (in Step 1). (2) Source data authenticity of the energy consumption value sent by $SP$.

- **Step 3-Verify Full MAC:** Upon receiving the two pairs of (message, full MAC) from $BT$ (which implies that $SP$’s location has been verified to be authentic), $SM$ calculates the two messages’ respective full MACs:
  - One messages is originated from $SP$ but forwarded by $BT$, appended with partial MACs based on $SP$ and $BT$’s MAC key shares.
  - The other message is originated from $MS$ but forwarded by $BT$, appended with partial MACs based on $MS$ and $BT$’s MAC key shares.

For calculating the full MACs, $SM$ needs the key shares which it generated and distributed in Commissioning Phase (Section III-C1) for the participants: $SP$, $BT$ and $MS$:

- For $m_{SP}$, $SM$ implements:

$$\text{Full-MAC-Verification}(m_{SP}, FM_{SP}, \{SP, BT\}, k, M)$$  \hspace{1cm} (7)

- For $m_{MS}$, $SM$ implements:

$$\text{Full-MAC-Verification}(m_{MS}, FM_{MS}, \{MS, BT\}, k, M)$$  \hspace{1cm} (8)

$SM$ ends the protocol if (7) or (8) returns a False value. Otherwise, $SM$ continues to check whether $r_{SP} = r_{MS}$ (recall that $r_{SP}$ is one part of $m_{SP}$ while $r_{MS}$ is one part of $m_{MS}$). If $r_{SP} = r_{MS}$, then $SM$ is ensured that energy reading from $SP$ is authentic; otherwise manipulated.

IV. SECURITY ANALYSIS

The security attainment of our proposed MAC scheme can be seen from the theorem below.

**Theorem 1:** The proposed MAC scheme is a secure MAC scheme for a HAN system.

**Proof 1:** It suffices to prove that the proposed MAC scheme is secure for our defined access structure $\hat{A}$. That is, under the proposed MAC scheme, a set $Q$ of participants can generate a valid (message, full MAC) pair if and only if $Q$ is qualified. More precisely, we prove that only the set of participants $\{SP, BT\}$ is able to generate a valid full MAC for the message $m_{SP}$ while only $\{MS, BT\}$ is able to generate a valid full MAC for the message $m_{MS}$. As follows we show the proof for the message $m_{SP}$, which can be easily extended for the message $m_{MS}$.
Recall that the full MAC for $m_{SP}$ is calculated by Equation (4). Due to the XOR operation, Equation (4) can be obtained only if both Equation (1) and Equation (3) are successfully obtained. Since both SP and BT’s MAC key shares are secrets, this further implies that the chosen HMAC is insecure, which contradicts with the fact the HMAC is chosen to be a secure MAC when we propose the multi-participant MAC scheme.

V. DISCUSSION AND FUTURE WORK

In this paper we study all the data provenance issues within a Home Area Network for smart metering infrastructure, including potential attacks to impersonate the smart plug, manipulate a smart appliance’s energy consumption data, relocate the smart plug, etc.

An existing work has addressed these problems by proposing a digital signature scheme. However, it is known that signature based scheme is asymmetric key based and thus inefficient due to lengthy keys and extensive computations required. We therefore in this paper take an alternative approach, that is, symmetric key based which is more efficient because of shorter keys required.

Part of our future work is to implement the proposed MAC scheme, particularly, to design a monotone span program that is of smallest size yet computing the access structure defined for our use case.

REFERENCES