Title	PRGA: Privary-preserving Recording & Gateway-Assisted Authentication of Power Usage Information for Smart Grid
Author(s)	Chim, TW; Yiu, SM; Li, VOK; Hui, LCK; Zhong, J
Citation	IEEE Transactions on Dependable and Secure Computing, 2015, v. 12 n. 1, p. 85-97
Issued Date	2015
URL	http://hdl.handle.net/10722/218736
Rights	IEEE Transactions on Dependable and Secure Computing. Copyright © IEEE.

PRGA: Privacy-Preserving Recording & Gateway-Assisted Authentication of Power Usage Information for Smart Grid

Tat Wing Chim, *Fellow, IEEE*, Siu-Ming Yiu, Victor O.K. Li,
Lucas C.K. Hui, *Senior Member, IEEE*, and Jin Zhong

Abstract—Smart grid network facilitates reliable and efficient power generation and transmission. The power system can adjust the amount of electricity generated based on power usage information submitted by end users. Sender authentication and user privacy preservation are two important security issues on this information flow. In this paper, we propose a scheme such that even the control center (power operator) does not know which user makes the requests of using more power or agreements of using less power until the power is actually used. At the end of each billing period (i.e., after electricity usage), the end user can prove to the power operator that it has really requested to use more power or agreed to use less power earlier. To reduce the total traffic volume in the communications network, our scheme allows gateway smart meters to help aggregate power usage information, and the power generators to determine the total amount of power that needs to be generated at different times. To reduce the impact of attacking traffic, our scheme allows gateway smart meters to help filter messages before they reach the control center. Through analysis and experiments, we show that our scheme is both effective and efficient.

Index Terms—Smart grid network, authentication, privacy preserving, commitment, homomorphic encryption, bloom filter

1 Introduction

The smart grid network is considered as the next generation power supply network which facilitates reliable and effective transmission of electricity from power generators to end users. In this network, the amount of electricity generated can be adjusted based on the real-time demand of end users in two ways. First, for big users such as factories, on top of the basic level of power supply, they can express additional power request one day, one week, one month or even one year ahead. The power operator then imposes additional charge to them. Second, for all end users, the power operator publishes a discount table one day, one week, one month or even one year ahead such that if an end user can reduce certain level of power usage in the forthcoming period, it can get a discount in its electricity bill. Thus the end user expresses its power reduction plan (especially for the usage of electric appliances which consume lots of power) as a response. Both ways are very common in United States and European countries. They not only ensure that user demands are satisfied but also avoids excess electricity generation. The latter can in turn help increase the profit of the power operators and protect the environment. In the future, a smart grid network may also facilitate big end users to sell requested but unused electricity to other end users in the market.

The additional power request and power reduction plan can be considered as contracts between end users and the power operator. Then how can the power operator ensure that they are actually enforced? This is challenging because the communications network is independent of the power transmission network and due to the physical properties of power transmission, an end user can pull any amount of power from the grid anyway. One common approach is to make use of the smart meters installed at end users' houses or factories. At the end of each billing period, the power operator compares an end user's actual power usage (being recorded by its smart meter) with the contracted amount to see whether they match. If not, the power operator imposes a fine or just cancels the agreed discount for the end user. For the case of additional power request, this discourages an end user from requesting too much or using more power than requested. For the case of power reduction, this means an end user must reduce the power usage as contracted in order to get a discount.

Two recent works [1] and [2] proposed a comprehensive and hierarchical structure for smart grid communications. There are home area networks (HANs) at the user end, building area networks (BANs) at the building feeder and neighborhood area networks (NANs) among substations. A NAN is formed by a large number of BANs while a BAN is formed by a large number of HANs. Note that BANs may not exist in some real-world advanced metering infrastructure (AMI) deployments nowadays but we emphasize that our proposed scheme still works even without the aggregation at BANs. A simplified architecture of the communications

- *T.W. Chim, S.M. Yiu, L.C.K. Hui, and J. Zhong are with the Department of Computer Science, Hong Kong University, Room 519, 5/F, Haking Wong Building, Hong Kong. E-mail: {twchim, smyiu, hui}@cs.hku.hk, jzhong@eee.hku.hk.*
- *V.O.K. Li is with the Department of Electrical and Electronic Engineering, Hong Kong University, 613 Chow Yei Ching Bldg., Hong Kong. E-mail: vli@eee.hku.hk.*

Manuscript received 16 May 2013; revised 8 Dec. 2013; accepted 17 Mar. 2014. Date of publication 25 Mar. 2014; date of current version 16 Jan. 2015. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below. Digital Object Identifier no. 10.1109/TDSC.2014.2313861

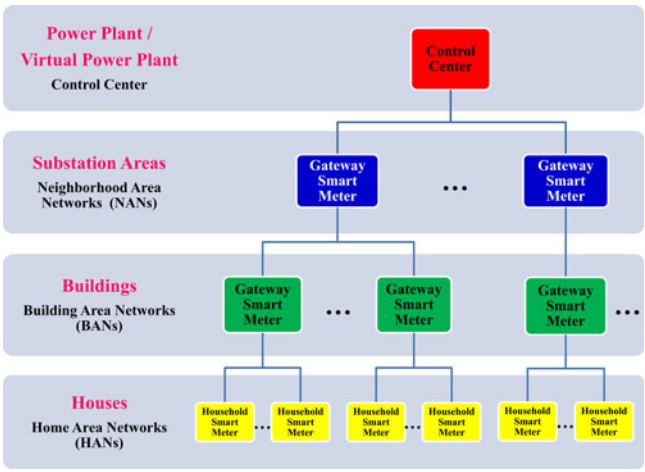

Fig. 1. Hierarchical architecture of smart grid.

network (ignoring the power generation and distribution network here) is shown in Fig. 1.

Basically, there is one control center, belonging to the power operator and located at the power plant, connected to multiple substation areas. Note that if there are distributed energy resources, this control center becomes the controlling unit of the so called virtual power plants [3] which integrate various distributed energy resources. Each substation area contains one NAN gateway smart meter connecting to buildings. Each building contains one BAN gateway smart meter connecting to houses. Each house in turn contains one household (or HAN gateway) smart meter connecting to all electric appliances in the house. Each BAN gateway smart meter will collect additional power request messages and power reduction plans from household smart meters, aggregate them and forward them to the NAN gateway smart meter. Similarly, each NAN gateway smart meter will collect messages from BAN gateway smart meters and forward them to the control center. Each BAN gateway also lets household smart meters download the up-to-date daily, weekly, monthly or yearly discount table so that an end user can make the power reduction plan in the forthcoming period accordingly. In terms of communications technologies, a household smart meter communicates with a BAN gateway smart meter (i.e., HAN-BAN connection) using Wi-Fi while a BAN gateway smart meter communicates with a NAN gateway smart meter (i.e., BAN-NAN connection) using WiMax. A NAN gateway smart meter at a substation in turn communicates with the control center using the supervisory control and data acquisition (SCADA) [4] system.

Unlike the kWh meters in the traditional power network, household smart meters can push information about an end user to gateway smart meters and then to the control center. There are many discussions about the functionalities of household smart meters. In general, we assume that a household smart meter can communicate with the electric appliances or electric machines using HAN (which usually adapts the Zigbee protocol), predict or project the overall electricity requirements of the household or the factory, and let end users input its daily, weekly, monthly or yearly power usage plans. It then forwards the information to the BAN gateway smart meter. Overall speaking, it can be regarded as an intelligent device with adequate amount of computational power. Readers may refer to [5] and [6] for more details.

In this paper, we focus on two major security issues of the communication between a gateway smart meter and the household smart meters in its region of responsibility. For the communications between the control center and NAN gateway smart meters, some security measures are already in place in the extended version of SCADA [7] and so we will skip this part due to space limitations. Information flow between individual household smart meters and gateway smart meters (and then between gateway smart meters and the control center) has a big impact on the reliability of power supply and is related to the charges for the end users. Security issues in a smart grid system (referred to as cyber security [8]) must not be overlooked. In particular, sender authentication and user privacy preservation are two major concerns [9].

The additional power request messages and power reduction plans sent by end user smart meters are used by the control center (the power operator) to determine how much electricity to generate in the forthcoming period and how to balance the load of different power generators. If the messages are not authenticated to confirm that they are from valid registered users, attackers can easily launch an illegal packet flooding attack[1] that can seriously affect the availability of the service. Unlike traditional kWh meters which only record the cumulative amount of electricity used, household smart meters transmit additional power requests or power reduction plans to substations via the communications network. As a result, future electricity usage patterns of end users can be easily traced and leaked. Such privacy leakage can then be used to reveal the daily habits of the end user such as when he/she is not at home or when a factory is in operation. This privacy preservation issue has been raised in both [9] and [11]. A proper privacy-preserving mechanism has to be adopted.

To tackle both issues is not trivial and subject to the following challenges. First, most data communications in a smart grid network are time critical. Any delay may result in the consumer experiencing electricity interruption. According to [8], the power generator system only has a few seconds to receive data from substations in each period. Any security schemes added to the system should be efficient in terms of computational complexity. While the identity of a sender needs to be authenticated, we have to preserve the privacy of the end users, and we do not want the power operator and the substations to be able to know how an end user will use power in the forthcoming period. Thus, techniques to enable a substation to authenticate and aggregate messages without knowing the actual contents and to enable the power operator to know the accumulated power demand in the forthcoming period are required.

In this paper, we utilize the hierarchical structure of smart grid and propose a novel Privacy-preserving Recording and Gateway-assisted Authentication (PRGA) of power

1. We remark that for other distributed denial of service (DDoS) attacks in which the packets are seemingly from valid users, we need to employ other techniques to protect the system [10]. This is also an important topic that requires more investigation, but is not the focus of this paper.

usage information for smart grid. We focus on the subsystem that connects gateway smart meters to end users. Our scheme has the following security features:

1) Gateway smart meters, which are physically more secure from being attacked, are responsible for basic authentication and aggregation of messages sent by user smart meters on their way to the control center. This can help ensure the availability of the power system (by filtering illegal packets) and reduce the traffic loading (by aggregating messages) in the communications network. The techniques of hash-based message authentication code (HMAC) and homomorphic encryption are used.
2) The real identity of any user smart meter and the power usage plans sent by it are kept secret even from the control center before the power is actually used. That means except for the end user, no one in this world knows how he/she will use power in the forthcoming period. We make this possible by encryption.
3) At the end of each billing period, the power operator can compare an end user's actual power usage (being recorded by its smart meter) with the contracted amount to see whether they match. We make this possible using the concept of commitment.

To summarize, we are going to propose a scheme such that the control center can record an end user's additional power request or power reduction plan anonymously. That is, even the power operator does not know how an end user will use power in advance. At the end of each billing period (i.e., after electricity usage), the end user can prove to the power operator that it has really made the additional power request or power reduction plan, together with the amount concerned. To reduce the traffic loading in the communications network, our scheme also allows gateway smart meters to aggregate messages such that power generators can know the total amount of power that needs to be generated at different times. Note that the power operator as well as the gateway smart meters do not know and do not need to know the exact power usage pattern of each end user in advance. It is because in the electricity distribution mechanism, they only need to know the total amount of electricity required and the power level to be maintained at each point of the power grid at each moment.

The rest of the paper is organized as follows. Related work is reviewed in Section 2. The system model and the problem statement are described in Section 3. Basic cryptographic concepts are summarized in Section 4. Our schemes are presented in Section 5. Security analysis and performance of our scheme are presented in Sections 6 and 8, respectively, and we conclude in Section 9.

2 Related Work

The smart grid project was initiated by the European Union in 2003 [12]. At around the same time, the IntelliGrid project [13] was started by the Electric Power Research Institute of the USA, while the US DOE started the Grid 2030 project [14]. Under the Energy Independence and Security Act of 2007, the National Institute of Standards and Technology (NIST) is responsible for coordinating the development of a framework for information management to achieve interoperability of smart grid devices. In 2010, NIST released a report [8] which describes the potential components of a smart grid. Some security issues (which they define as cyber security) are also discussed.

A recent work [15] elaborates on the importance of a smart grid especially with the consideration of renewable energy resources. A control model known as risk-limiting dispatch is adopted. Some new requirements of the communication architecture and potential security problems are identified. As such, there is an urgent need to establish protocols and standards for the smart grid.

In terms of security, the major issues are discussed in details in [9] and [11]. For the generator-to-substation communication, some security measures are already in place in the extended version of SCADA [7]. For the substation-to-smart-meter communication, sender authentication and user privacy preservation are considered as two major concerns as discussed earlier. A key objective of sender authentication is to avoid the so-called false data injection (i.e., an attacker injects a large volume of data into the control center in order to disturb its normal operation and to perform some sort of denial of service attack). In the old days, the research community usually adopts state estimation (i.e., finding out abnormality in the current state from previous ones) to identity false data being injected into a network. However, it was recently proved that this approach can easily be compromised [16]. Even if such a state estimation approach can be implemented, it is not appropriate for a smart grid system. Requiring the control center to perform state estimation implies that the control center needs to handle a large volume of valid and invalid data. Obviously, this violates the strict data reception assumption of the control center (recall that the control center only has a few seconds in each period for receiving data). In two recent efforts [1] and [2], the authors proposed a simple authentication scheme. After initial authentication using conventional public key infrastructure (PKI), any two parties in HANs, BANs or NANs can establish a shared key using the Diffie-Hellman technique. They then use that shared key to create HMAC signature for all ongoing communications. However, privacy issues are not addressed at all. In a recent work [17], a set of privacy-preserving protocols is proposed for a user to combine smart meter readings with a certified tariff policy to generate an electricity bill, which is then transmitted to the service provider together with a zero-knowledge proof to ensure its correctness and to avoid information leakage. Although this work also adopts commitment schemes like ours, it only handles the submission of smart meter readings. It is difficult to directly apply it to handle the submission of power usage plans. Note that power usage plans are important as they may show important deviations of power usage of the users and thus they need to be well protected.

Smart meters, especially end-user smart meters, are more vulnerable to physical disturbance and compromise since they are located at end-users' homes or factories and the power operator has no way to keep an eye on them. In fact, some attacks on smart meters have already been identified [18], [19], [20]. To secure smart meters, researchers follow

two major directions. One is based on software protection and attestation schemes like [21]. The other is based on a trusted platform module (TPM) [22] which is attached to the smart meter for storing secret information such as keys. This TPM is assumed to be tamper-resistant such that information stored in it cannot be read or modified easily. In this paper, we also assume the existence of such a TPM platform. Nevertheless, even if such smart meter attacks take place, our proposed scheme also provides a backdoor solution. In particular, our scheme allows the control center to "isolate" a seemingly compromised smart meter by not allowing it to obtain the updated master secret. For details, please refer to Section 5.5.

3 System Model and Security Requirements

In this section, we discuss our assumptions and security requirements in details.

3.1 System Model and Assumptions

Recall that we consider a smart grid network as a hierarchical architecture. There are home area networks at the user end, building area networks at the building feeder and neighborhood area networks among substations. A NAN is formed by a large number of BANs while a BAN is formed by a large number of HANs (see Fig. 1). Basically, there is one control center, belonging to the power operator and located at the power plant, connected to multiple substation areas. Each substation area contains one NAN gateway smart meter connecting to buildings. Each building contains one BAN gateway smart meter connecting to houses. Each house in turn contains one household (or HAN gateway) smart meter connecting to all electric appliances in the house. It can be easily observed that these layers have different physical security level assumptions.

1) The control center (belonging to the power operator) is assumed to be secure and fully trusted.
2) Gateway smart meters are usually physically locked from outside access. For example, a BAN gateway smart meter is usually locked inside the control room of a building while a NAN gateway smart meter is usually locked inside the substation. These regions are relatively more difficult to be compromised by attackers. In this paper, we assume that they are secure.
3) End-user smart meters are more vulnerable to physical disturbance and compromise since they are located at end-users' homes or factories and the power operator has no way to keep an eye on them. They are made more secure by attaching tamper-resistant trusted platform modules to them. TPM is assumed to be tamper-resistant such that keys stored in them are difficult to be cracked or altered.

Without loss of generality, we assume that servers at the control center and gateway smart meters have higher computational power than the average home personal computers. Although this may not be true for the so called "industry computer" used in the field right now, this may change as computers are upgraded when smart grid network becomes more mature. Also we assume that any two parties can establish a secure channel using conventional public key infrastructure to minimize the impact of network-level attacks.

3.2 Security Requirements

We aim at designing an authentication scheme to validate messages sent by end user smart meters which are located at end users' homes or factories while at the same time, preserve the end-users' privacy (such as future daily electricity usage pattern). The security requirements are summarized as follows:

1) Power plan message authentication. The power plan message from any household smart meter has to be properly authenticated before they are forwarded to and processed by the control center. Also an attacker cannot impersonate any valid smart meter to send out fake power plan messages.
2) Privacy preservation of future power usage plan. Before actual power usage, no one including the control center and gateway smart meters can know how an end-user will use power in the forthcoming period although this information is listed in a power usage plan.
3) Non-repudiation of power usage plan. At the end of a billing period, an end user has to prove to the control center that its power usage plan submitted earlier agrees with the actual power usage. The end user cannot deny any power usage plan submitted or cannot argue that it has made a certain power usage plan which it has not actually made.
4) Traceability. Although no one knows how an end-user will use power in the forthcoming period, a power usage plan can be related to the corresponding end-user at the end of the billing period. This is necessary for the stability of power grid and for the accuracy of charging bill calculation. Also this can lead to some deterrent effects to attacking activities as the power company can impose punishments to the users concerned upon discovering attacking activities.

4 Preliminaries

In this section, we explain the concepts of public-key encryption and digital signature, hash-based message authentication code, homomorphic encryption, Bloom filter, and commitment, respectively.

4.1 Public-Key Encryption and Digital Signature

Public-key encryption is a function provided by the public key infrastructure and is also known as asymmetric encryption. A trusted party assigns each user a pair of public key and private key. The public key can be known by everyone while the private key is kept secret. To securely send a message, the sender encrypts the message using the receiver's public key. The receiver can then obtain the message by decrypting using the corresponding private key. RSA [23] is a well-known algorithm for public-key encryption. Throughout this paper, we denote the process of encrypting plaintext M with public key PK to obtain ciphertext C as

$C = ENC_{PK}(M)$. Similarly, we denote the process of decrypting ciphertext C with private key SK to obtain plaintext M as $M = DEC_{SK}(C)$. In this paper, we assume that any two parties can establish a secure channel using PKI to minimize the impact of network-level attacks.

4.2 Hash-Based Message Authentication Code

Hash-based message authentication code is a specific construction for computing a message authentication code (MAC) using a cryptographic hash function in combination with a secret key. Both data integrity and authenticity of a message can be achieved using such a technique. Well-known hash functions such as SHA-1 [24] and MD5 [25] can be extended to produce an HMAC. Due to the nature of hash functions, an HMAC value can be computed in a much shorter time than a traditional digital signature. Throughout this paper, we denote the HMAC value generated on message M using the secret key K as $HMAC_K(M)$.

4.3 Homomorphic Encryption

Homomorphic encryption is a special kind of encryption having the following property. If we want to perform an operation (e.g., to compute the sum) on the plain data but these data are now encrypted, one can obtain the resulting value by performing the computation on the encrypted values such that the one who carries out the computation cannot know the values of the data, but the receiver is able to decrypt and obtain the correct resulting answer. For example, the homomorphic encryption used in Paillier cryptosystem [26] has the following property. The encryption of the sum of two numbers is equivalent to the product of the encrypted values of the individual numbers. To be specific, let the public key used for encryption in this system be $PK = (m, g)$. Given two numbers x_1 and x_2. The encrypted values of x_1 and x_2 are $ENC_{PK}(x_1) = g^{x_1}{r_1}^m$ and $ENC_{PK}(x_2) = g^{x_2}{r_2}^m$, respectively, where r_1 and r_2 are random numbers. The product of the encrypted values of x_1 and x_2 is $g^{x_1}{r_1}^m \times g^{x_2}{r_2}^m = g^{x_1+x_2}(r_1r_2)^m$ which is equivalent to the encrypted value of $x_1 + x_2$ (i.e., $ENC_{PK}(x_1 + x_2)$). Therefore, Paillier cryptosystem carries the property of homomorphic encryption: $ENC_{PK}(x_1 + x_2) = ENC_{PK}(x_1) \times ENC_{PK}(x_2)$. Having this property, a third party holding encrypted numbers can help to compute their sum without the need of first decrypting them or even knowing the numbers in advance.

4.4 Bloom Filter

A Bloom filter provides an efficient representation of a set $A = a_1, a_2, \ldots, a_n$ of n elements to support membership queries. The idea is to allocate a vector v with m bits, initially all set to 0, and then choose k independent hash functions, h_1,h_2,$\ldots, h_k$, each with range $1, \ldots, m$. For each element $a \in A$, the bits at the positions $h_1(a)$,$h_2(a)$,$\ldots, h_k(a)$ in v are set to 1 (A particular bit might be set to 1 multiple times). To answer if a value b is in A, we check the bits at positions $h_1(b)$, $h_2(b)$,$\ldots, h_k(b)$. If any of them is 0, then b is definitely not in the set A. Otherwise we conjecture that b is in the set although there is a certain probability that we are wrong (called a false positive). After inserting n keys into the vector with m bits with k hash functions, the probability that a particular bit is still 0 is $(1 - \frac{1}{m})^{kn} \sim e^{-\frac{kn}{m}}$ assuming that on any input value, the hash functions pick each position with equal probability. Hence the probability of a false positive is $(1 - (1 - \frac{1}{m})^{kn})^k \sim (1 - e^{-\frac{kn}{m}})^k$. Let $f(k) = (1 - e^{-\frac{kn}{m}})^k$ and let $g(k) = lnf(k) = kln(1 - e^{-\frac{kn}{m}})$. By finding $\frac{dg}{dk}$ and making $\frac{dg}{dk} = 0$, it can be shown that to minimize the probability of having false positives, k should be set to $\frac{mln2}{n}$. On the other hand, as long as the functions used are one-way functions, one cannot retrieve any information being put into a bloom filter.

4.5 Commitment

A commitment scheme allows one party to commit to a value while keeping it secret from the other party. At a later time, the first party can reveal the committed value and prove to the other party that this revealed value is the same as the committed one. To achieve this purpose, two functions are defined—$Commit(.)$ and $CheckReveal(.)$ The first function takes a secret value and a commitment key as input and produces a commitment while the second function takes the commitment, the value to be revealed and a decommitment key as input and produces a positive or a negative answer. Let us take a look at an example. Assume that Party A wants to commit the value X during his conversation with Party B. Party A first generates a commitment and decommitment keys, denoted by CK_A and DK_A respectively. Party A then computes $C_A = Commit(X, CK_A)$ and sends it to Party B. At a later time, Party A sends C_A, X and DK_A to Party B which then invokes the function $CheckReveal(C_A, X, DK_A)$. RSA with random padding [23] is a common implementation for commitment functions. In this case, CK_A and DK_A form a public and private key pair and are kept by Party A. $C_A = Commit(X, CK_A)$ then becomes $C_A = ENC_{CK_A}(X)$. $CheckReveal(C_A, X, DK_A)$ thus involves the checking of $X = DEC_{DK_A}(C_A)$. Since by the property of RSA with random padding, it is computationally hard to find two messages that can be encrypted to the same ciphertext, we can conclude that Party A cannot modify X when revealing its value to Party B.

5 Our Solution

This section presents our privacy-preserving recording and gateway-assisted authentication of power usage information for smart grid in details. Throughout this paper, we denote the process of encrypting plaintext M with public key PK to obtain ciphertext C, the process of decrypting ciphertext C with private key SK to obtain plaintext M and the HMAC value generated on message M using the secret key K as $C = ENC_{PK}(M)$, $M = DEC_{SK}(C)$ and $HMAC_K(M)$, respectively. On the other hand, to save space, we use a variable without any subscript to represent a collection of the same variable with subscripts. For example, we use Xs to represent a collection of X_i.

Our scheme contains four phases and we will discuss them one by one:

1) Preparation phase. In this phase, the control center sets up system parameters.

TABLE 1
Notations Used in this Paper

Symbol	Meaning
PK_{CC}	Public key of control center
SK_{CC}	Private key of control center
HSM_i	Household smart meter i
$HSMID_i$	Identity of HSM_i
PK_{HSM_i}	Public key of HSM_i
SK_{HSM_i}	Private key of HSM_i
BSM_i	BAN gateway smart meter i
$BSMID_i$	Identity of BSM_i
PK_{BSM_i}	Public key of BSM_i
SK_{BSM_i}	Private key of BSM_i
NSM_i	NAN gateway smart meter i
$NSMID_i$	Identity of NSM_i
PK_{NSM_i}	Public key of NSM_i
SK_{NSM_i}	Private key of NSM_i
s	System master secret
n	Number of pre-defined sub-periods
U_i	Power usage plan by HSM_i
u_{ij}	Amount of additional power requested or power reduction agreed by HSM_i in j^{th} sub-period
E_i	Encrypted power usage plan by HSM_i
T	Current timestamp
CK_i	Commitment key of HSM_i
DK_i	De-commitment key of HSM_i
$Commit(M, CK_i)$	Commitment on M
$CheckReveal(C, M, DK_i)$	Reveal M with commitment C
H_i	Hash of $HSMID_i$, T and U_i by HSM_i
C_i	Commitment of H_i with CK_i by HSM_i
AE_j	Aggregated Es by BSM_j
HBF_j	Bloom filter for storing Hs by BSM_j
CBF_j	Bloom filter for storing Cs by BSM_j
AAE_k	Aggregated AEs by NSM_k
$ENC_x(M)$	Encryption of plaintext M using key x
$SIG_x(M)$	Signature on message M using key x
$HMAC_x(M)$	HMAC on message M using key x

2) Power plan submission phase. In this phase, a smart meter submits power plans to request for additional power or to express its intention to reduce power usage. The control center cannot relate the power plans to the actual users in this phase.
3) Power plan processing phase. In this phase, power plans submitted by smart meters are processed by their gateway smart meters and by the control center.
4) Reconciliation phase. In this phase, a smart meter needs to prove to the control center that its actual power usage is consistent with the power plans submitted earlier.
5) System master secret updating phase. In this phase, the control center updates the system master secret in all non-compromised smart meters.

We first summarize the notations that will be used in this paper in Table 1 to enhance readability. Then we will describe each of these four phases one by one.

5.1 Preparation Phase

As discussed earlier, each household smart meter (located at an end user's home) or gateway smart meter (located at different locations of the power transmission network) is assumed to be tamper-resistant such that private keys and a system master secret can be stored on them securely without the worry of being tampered. Each tamper-resistant device has a clock which runs on its own battery. These clocks are assumed to be roughly synchronized.

The control center performs the following:

1) Generates its public and private keys by following the property of Paillier cryptosystem [26] or any other Homomorphic encryption cryptosystem. We denote PK_{CC} as its public key which is assumed to be preloaded into all tamper-resistant devices. Such a preloading process has been widely adopted in studies that involve the use of tamper-resistant devices. [27] is an example. SK_{CC} is its private key (corresponding to PK_{CC}) and is kept private.
2) Generates for each household smart meter HSM_i an identity $HSMID_i$, a pair of conventional public and private keys by following the property of any PKI. Let PK_{HSM_i} and SK_{HSM_i} be its public and private keys respectively. $< HSMID_i, PK_{HSM_i} >$ is stored into the control center's database while SK_{HSM_i} is preloaded into the smart meter. These keys are used for the purpose of initial transmission or updating of system master secret (details will be discussed in the last point).
3) Generates for each BAN gateway smart meter BSM_i an identity $BSMID_i$, a pair of conventional public and private keys by following the property of any PKI. Let PK_{BSM_i} and SK_{BSM_i} be its public and private keys respectively. $< BSMID_i, PK_{BSM_i} >$ is stored into the control center's database while SK_{BSM_i} is preloaded into the smart meter. These keys are used for the purpose of initial transmission or updating of system master secret (details will be discussed in the last point).
4) Generates for each NAN gateway smart meter NSM_i an identity $NSMID_i$, a pair of conventional public and private keys by following the property of any PKI. Let PK_{NSM_i} and SK_{NSM_i} be its public and private keys respectively. $< NSMID_i, PK_{NSM_i} >$ is stored into the control center's database while SK_{NSM_i} is preloaded into the smart meter. These keys are used for the purpose of initial transmission or updating of system master secret (details will be discussed in the next point).
5) Generates a system master secret s and securely transmits it to each household smart meter HSM_i, each BAN gateway smart meter BSM_i and each NAN gateway smart meter NSM_i by encrypting using the corresponding PK_{HSM_i}, PK_{BSM_i} and PK_{NSM_i}, respectively. At a later time and if there is a neccessity (e.g., a smart meter is proved to be compromised), the control center can generate a new system master secret and securely transmit it to each non-compromised household or BAN gateway smart meter in the same way.

5.2 Power Plan Submission Phase

An end user can request for additional power or to express the intention to reduce power at any time (e.g., one-day ahead, one-week ahead or one-month ahead). At this moment, the user smart meter HSM_i performs the following steps:

1) Prepares an array of entries: $U_i = [u_{i0}, u_{i1}, \ldots, u_{i(n-1)}]$. Here we assume that there are altogether n predefined sub-periods in the forthcoming power

provisioning period. For example, if we are using one-day ahead power plan submission scheme, n can be set to 24 and the basic unit becomes hour). u_{ix} represents the amount of additional power required (if it carries a positive value) or power reduction agreed (if it carries a negative value) by the household smart meter HSM_i in the xth sub-period. u_{ix} can take up a zero value if the end user concerned does not require additional power or does not want any power reduction in the xth sub-period.

2) Encrypts each entry in the array using the control center's public key PK_{CC} by incorporating randomly generated numbers (as specified in Paillier cryptosystem [26] or any other Homomorphic encryption cryptosystem standard) to form: $E_i = [e_{i0}, e_{i1}, \ldots, e_{i(n-1)}]$ where $e_{ix} = ENC_{PK_{CC}}(u_{ix})$. In this way, no gateway smart meter can know the value of any u_{ix}.
3) Generates a pair of commitment and decommitment keys and saves them locally. Without loss of generality, let CK_i be the commitment key and let DK_i be the corresponding de-commitment key.
4) Computes the hash of the array U_i together with its identity $HSMID_i$ and the current time stamp T as: $H_i = h(HSMID_i, T, U_i)$.
5) Commits H_i to form: $C_i = Commit(H_i, CK_i)$.
6) Computes the HMAC signature with the system master secret s as the key on E_i, H_i and C_i to form $HMAC_s(E_i \| H_i \| C_i)$ where $\|$ stands for simple concatenation.
7) Sends $ENC_{PK_{BSM_j}}(E_i, H_i, C_i, HMAC_s(E_i \| H_i \| C_i))$ to its upper level BAN gateway smart meter BSM_j.
8) Stores CK_i, DK_i, T, U_i and C_i locally.

In this paper, we only focus on the security requirement that a user's power usage pattern is kept private from anyone before he/she uses the power. In fact, a user's power usage pattern may still be inferred after the control center has collected a large volume of power usage information over time. As a future work, we will investigate a possible solution to this problem. As a preliminary idea, request messages from all users in a region will be aggregated so that the utility can only get the long-term usage trend of a region, but not of an individual end user.

5.3 Power Plan Processing Phase

The BAN gateway smart meter BSM_j does not forward the received power plans from its lower level household smart meters to its upper level NAN gateway smart meter immediately. Instead, it only performs such a forwarding at regular intervals. For example, if one-day ahead scheme is adopted, the BAN gateway forwards the plans to the NAN gateway smart meter every mid-night. Without loss of generality, a BAN gateway smart meter should receive more than one power plans during such an interval. Upon the time of forwarding, BSM_j performs the following steps:

1) For each $ENC_{PK_{BSM_j}}(E_i, H_i, C_i, HMAC_s(E_i \| H_i \| C_i))$ received, BSM_j decrypts the block using its private key SK_{BSM_j} and re-computes the HMAC signature $HMAC_s(E_i \| H_i \| C_i)$ based on the received E_i, H_i and C_i to see whether it is the same as the one attached. This ensures that the power usage plan is sent by a valid user smart meter and also it is not modified by anyone. If the computed value is not the same as the received one, it simply drops the power plan message or requests the household smart meter concerned to re-submit its power usage plan.
2) Aggregates the received power usage plans by computing the product of each array entry in Es to form: $AE_j = [ae_{j0}, ae_{j1}, \ldots, ae_{j(n-1)}]$ where $ae_{jx} = e_{0x} \times e_{1x} \times \cdots \times e_{(m-1)x}$. Here we assume that there are altogether m power plans received and to be aggregated.
3) Prepares two bloom filters HBF_j and CBF_j (we will discuss about how to set their size in Section 8.1. BSM_j then adds H_0, $H_1, \ldots, H_{m-1}$ into HBF_j and adds $C_0, C_1, \ldots, C_{m-1}$ into CBF_j.
4) Computes the HMAC signature with the system master secret s as the key on AE_j, HBF_j and CBF_j to form $HMAC_s(AE_j \| HBF_j \| CBF_j)$ where $\|$ stands for simple concatenation.
5) Forwards $ENC_{PK_{NSM_k}}(BSMID_j, AE_j, HBF_j, CBF_j, HMAC_s(BSMID_j \| AE_j \| HBF_j \| CBF_j))$ to its upper level NAN gateway smart meter NSM_k.

To facilitate any user smart meter to check whether its information is being aggregated by its upper level BAN gateway smart meter BSM_j, we require that the latest version of AE_j, HBF_j and CBF_j to be posted publicly and can be downloaded by any user smart meter for checking at any time. In this way, any user smart meter can keep track of any change on the three aggregated values before and after its power usage plan submission.

Upon receiving from multiple BANs, the upper level NAN gateway smart meter NSM_k performs the following steps:

1) For each $ENC_{PK_{NSM_k}}(BSMID_j, AE_j, HBF_j, CBF_j, HMAC_s(BSMID_j \| AE_j \| HBF_j \| CBF_j))$ received, NSM_k decrypts the block using its private key SK_{NSM_k} and re-computes the HMAC signature $HMAC_s(BSMID_j \| AE_j \| HBF_j \| CBF_j)$ based on the received $BSMID_j$, AE_j, HBF_j and CBF_j to see whether it is the same as the one attached. This ensures that the message is sent by a valid BAN gateway smart meter and also it is not modified by anyone. If the computed value is not the same as the received one, it simply drops the power plan message or requests the BAN gateway smart meter concerned to re-transmit the message.
2) Aggregates the received AEs in the same way as what the BAN gateway smart meter does. That is, it computes the product of each array entry in the AEs received to form an aggregated array AAE_k.
3) Computes the HMAC signature with the system master secret s as the key on its identity $NSMID_k$, AAE_k, $BSMID$s, HBFs and CBFs to form $HMAC_s(NSMID_k \| AAE_k \| BSMID_0 \| HBF_0 \| CBF_0 \| BSMID_1 \| HBF_1 \| CBF_1 \| \ldots \| BSMID_{p-1} \| HBF_{p-1} \| CBF_{p-1})$ where $\|$ stands for simple concatenation and p is the total number of HBFs and CBFs received.

4) Forwards $ENC_{PK_{CC}}(NSMID_k,\ AAE_k,\ BSMID_0,\ HBF_0, CBF_0, BSMID_1, HBF_1, CBF_1, \ldots, BSMID_{p-1}, HBF_{p-1}, CBF_{p-1}, HMAC_s(NSMID_k, AAE_k \| BSMID_0 \| HBF_0 \| CBF_0 \| BSMID_1 \| HBF_1 \| CBF_1 \| \ldots \| BSMID_{p-1} \| HBF_{p-1} \| CBF_{p-1}))$ to the control center.

Finally, the control center performs the following:

1) For each $ENC_{PK_{CC}}(NSMID_k, AAE_k, BSMID_0, HBF_0, CBF_0, BSMID_1, HBF_1, CBF_1, \ldots, BSMID_{p-1}, HBF_{p-1}, CBF_{p-1},\ HMAC_s(NSMID_k,\ AAE_k \| BSMID_0 \| HBF_0 \| CBF_0 \| BSMID_1 \| HBF_1 \| CBF_1 \| \ldots \| BSMID_{p-1} \| HBF_{p-1} \| CBF_{p-1}))$ received, the control center decrypts the block using its private key SK_{CC} and recomputes the HMAC signature $HMAC_s(NSMID_k, AAE_k \| BSMID_0 \| HBF_0 \| CBF_0 \| BSMID_1 \| HBF_1 \| CBF_1 \| \ldots \| BSMID_{p-1} \| HBF_{p-1} \| CBF_{p-1})$ based on the received $NSMID_k$, AAE_k, $BSMID$s, HBFs and CBFs to see whether it is the same as the one attached. This ensures that the message is sent by a valid NAN gateway smart meter and also it is not modified by anyone. If the computed value is not the same as the received one, it simply drops the power plan message or requests the NAN gateway smart meter concerned to re-transmit the message.
2) Aggregates the AAEs received in the same way as what the BAN and NAN gateway smart meters do. That is, it computes the product of each array entry in the AAEs received. It then decrypts each entry in the aggregation using its private key SK_{CC} to obtain the aggregated power demand information in each sub-period.
3) Stores $< BSMID_j, HBF_j, CBF_j >$ for each BSM_j into its own database for use in the reconciliation phase (details will be discussed in the next section).

Note that BANs may not exist in some real-world advanced metering infrastructure deployments nowadays. In this case, our scheme works in almost the same way except that there is no more aggregation at BAN gateway smart meters and the two bloom filters HBF_j and CBF_j are now prepared by the upper level NAN gateway smart meter NSM_k.

5.4 Reconciliation Phase

This phase is carried out at the end of each billing period. The control center requests each household smart meter to prove that it has submitted a certain power plan earlier. Such a proof is essential for two reasons. First, for end users who have used additional power without making request beforehand, penalties should be assessed. Second, for end users who have agreed to use less power and can actually use less power, discounts should be offered on their electricity bills.

Having received the request, an end user responds by sending its identity $SMID_i$, the time stamp used T, the original array U_i, the commitment C_i and the de-commitment key DK_i to the control center.

The control center then performs the following steps:

1) Computes $H_i = h(SMID_i, T, U_i)$ and ensures that both H_i and C_i are in the aggregated bloom filters HBF_j and CBF_j, respectively, submitted by its corresponding BAN gateway smart meter BSM_j. Note that mapping HSM_i to BSM_j is simple in a smart grid network since the geographical location of any end user is fixed.
2) Verifies the commitment information by invoking the function $CheckReveal(C_i, H_i, DK_i)$ to see whether it returns a positive value.
3) If yes, compares the agreed power plan and the actual power usage (to be physically measured by the smart meter like what the kWh meter does nowadays) of that end user to see whether they match. If not, penalties or additional charge will be imposed on the end user.

5.5 System Master Secret Updating Phase

Our scheme provides a mechanism for updating the system master secret s in case the control center believes that any smart meter has been compromised (e.g., by means of the software attestation protocol in [21]). Preliminary physical investigations may also be involved.

Assume that the control center wants to update the system master secret from s to s'. It finds the public keys of all smart meters except the one that has been compromised. For each of such smart meter SM_i, it composes the message $ENC_{PK_{SM_i}}(s')$ together with its signature $SIG_{SK_{CC}}(ENC_{PK_{SM_i}}(s'))$.

Upon receiving the key update message, SM_i's TPM first verifies the signature of the control center using the public key PK_{CC}. Then it decrypts the message using its own private key SK_{SM_i} to obtain s' and then replaces the old regional system key with the new one.

The time complexity of this key update procedure may be high but it can be carried out during non-peak hours (e.g., in the early morning). Also this phase can be applied to all household smart meters, BAN gateway smart meters and NAN gateway smart meters.

6 Security Analysis

We analyse our scheme with respect to the security requirements listed in Section 3.

1) Power plan message authentication. Before a smart meter transmits a request message to the control center, it has to include an HMAC signature on the encrypted message (i.e., E_i) using a system master secret s. This system master secret is only preloaded into all valid smart meters. Assuming that smart meters are tamper-resistant, this system master secret cannot be retrieved by an attacker easily. Hence an outside attacker (who is not a valid smart meter) does not know how to generate a valid HMAC signature. Thus our scheme is protected from outsider attacks.

   It is true that a compromised tamper-resistant device can still launch attacks (known as insider attacks) to the system since it possesses the system master secret s. However, as mentioned in Section 5.1, our scheme provides a way for the control center to update the system master secret using simple PKI. In addition, the new system key is sent in encrypted form to all non-compromised devices. The tamper-resistant device that is compromised

cannot get it. Also, the control center can set a bound for different types of users (e.g., domestic or commercial) to avoid a user requesting a huge amount of power as an insider attack.

2) Privacy preservation of future power usage plan. In the power plan submission phase (Section 5.2), an end user's power usage plan is encrypted using the control center's public key PK_{CC} before sending to the BAN gateway smart meter. The BAN gateway smart meter then aggregates the power usage plans of multiple end users using the technique of homomorphic encryption though it does not know how to decrypt them. At the control center side, it can decrypt using its private key SK_{CC} the aggregate amount of power usage in each sub-period. It cannot know the power usage plan of any individual end user.

   Besides the aggregated power usage, the control center also receives bloom filters containing H_i and C_i submitted by each household smart meter SM_i. However, by the property of bloom filter, one cannot retrieve any information being put into a bloom filter. Therefore, the control center obtains no information about any end user's power usage (even H_i and C_i values) from the bloom filters. The privacy of any end user's future power usage is thus preserved.

3) Non-repudiation of power usage plan. At the end of a billing period (during the Reconciliation Phase in Section 5.4), an end user has to send its identity $SMID_i$, the time stamp used T, the original array U_i, the commitment C_i and the de-commitment key DK_i to the control center. By the property of RSA, it is hard to find two messages that can be encrypted to the same ciphertext. Thus an end user cannot commit another power usage plan and generate the same commitment. Also the end user involved is the only party who holds the decommitment key DK_i which corresponds to the commitment key CK_i used to generate the commitment. Therefore, as long as an end user needs to submit a proof, he/she cannot deny any power usage plan submitted or cannot argue that it has made a certain power usage plan which it has not actually made.

   It is true that an end user can deny submitting any proof (i.e., deny opening its previously submitted commitment). However, this is of his/her own disadvantage. For additional power usage request, the power operator simply treats him/her as using additional power without making request. Penalties result. For agreement of power saving, the power operator simply treats him/her of not submitting any agreement of power saving. No discount is given to his/her electricity bill as a result.

4) Traceability. At the end of a billing period (during the reconciliation phase in Section 5.4), an end user has to send its identity $SMID_i$ to the control center. Thus the control center can issue correct charging bills to each end user.

7 Analysis of Time Complexity

In this section, we briefly analyze the time complexity of our scheme. Note that we ignore the time complexity involved in preparation phase and system master secret updating phase since they can be done offline and are only done once occasionally (e.g., when the control center wants to update the system master secret). It is not critical to the efficiency of our scheme.

Let T_{henc} denote the time required to perform one homomorphic encryption, T_{hdec} the time required to perform one homomorphic decryption, T_{enc} the time required to perform one conventional asymmetric encryption, T_{dec} the time required to perform one conventional asymmetric decryption, T_{sig} the time required to perform one digital signature, T_{hash} the time required to perform one hash computation, T_{hmac} the time required to perform one HMAC computation, $T_{commkeygen}$ the time to generate a commitment and decommitment key pair, T_{comm} the time to compute commitment, $T_{vercomm}$ the time to verify a commitment, T_{mul} the time to compute the product of two real numbers, T_{pbf} the time to add a number into a bloom filter and T_{cbf} the time to check whether a number is in a bloom filter. Note that conventional asymmetric encryption can actually be done in a hybrid manner. That is, a message is first symmetrically encrypted using a session key, which is then asymmetrically encrypted using the receiver's public key. Similarly conventional asymmetric decryption can also be done in a hybrid manner. That is, the receiver first performs asymmetrically decryption using his/her private key to obtain the session key, which is then used for symmetric decryption to obtain the message. Based on similar ideas, digital signature can also be done in a hybrid manner. That is, a message is first hashed into a fixed output, which is then digitally signed using the sender's private key.

We implemented all these functions onto an old-fashioned computer with processor speed 750 MHz and RAM size 4 GB. This hardware configuration is roughly equivalent to that of FriendlyARM [28], which is a possible microcontroller platform for implementing smart meters. We adopted Paillier cryptosystem with 512 bits of modulus and at least $1 - 2^{-64}$ certainty of primes generation for homomorphic encryption and decryption, RSA with 1,024 bits key for asymmetric encryption, decryption, digitial signature, commitment generation and verification, AES with 128 bits key for symmetric encryption and decryption, and MD5 algorithm for hash and HMAC computation. We summarize the average time required (over 10 experiments) for each of the above operations in Table 2.

According to Section 5.2, the user smart meter HSM_i takes time $n \times T_{henc}$ to produce E_i, $T_{commkeygen}$ to generate a commitment and decommitment key pair, T_{hash} to compute H_i, T_{comm} to obtain C_i and, and T_{hmac} to compute the final HMAC signature. As such, the total time required is $38.2n + 577.6 + 0.4 + 812.4 + 0.4 = 38.2n + 1390.8$ msec. For example, if $n = 24$, the total time required becomes 2307.6 msec which is about 2.3 sec.

According to Section 5.3, the BAN gateway smart meter BSM_j takes time T_{hmac} to re-compute the HMAC signature, $m \times n \times T_{mul}$ to aggregate power usage plans, $2m \times T_{pbf}$ to add hash and commitment values into bloom filters, and T_{hmac} to compute HMAC signature. As such, the total time required is $0.4 + 0.2mn + 0.2m + 0.4 = 0.2mn + 0.2m + 0.8$ msec. For example, if $n = 24$ and $m = 100$, the total time required becomes 500.8 msec which is about 0.5 sec.

TABLE 2
Average Time Required for Each Function

Operations	Average Time Required
T_{henc}	38.2 msec
T_{hdec}	97.1 msec
T_{enc}	812.4 msec
T_{dec}	9.0 msec
T_{sig}	9.0 msec
T_{hash}	0.4 msec
T_{hmac}	0.4 msec
$T_{commkeygen}$	577.6 msec
T_{comm}	812.4 msec
$T_{vercomm}$	9.0 msec
T_{mul}	0.2 msec
T_{pbf}	0.1 msec
T_{cbf}	0.1 msec

Similarly, the NAN gateway smart meter NSM_k takes time T_{hmac} to re-compute the HMAC signature, $m \times n \times T_{mul}$ to aggregate power usage plans (assuming messages from m BAN gateway smart meters are received), and T_{hmac} to compute HMAC signature. As such, the total time required is $0.4 + 0.2mn + 0.4 = 0.2mn + 0.8$ msec. For example, if $n = 24$ and $m = 100$, the total time required becomes 480.8 msec which is about 0.5 sec. The control center takes time T_{hmac} to re-compute the HMAC signature, $m \times n \times T_{mul}$ to aggregate power usage plans (assuming messages from m NAN gateway smart meters are received), and $n \times T_{hdec}$ to obtain the aggregated power demand information in each sub-period. As such, the total time required is $0.4 + 0.2mn + 97.1n = 0.2mn + 97.1n + 0.4$ msec. For example, if $n = 24$ and $m = 100$, the total time required becomes 2810.8 msec which is about 2.8 sec.

According to Section 5.4, for each user, the control center takes time T_{hash} to compute H_i, $2 \times T_{cbf}$ to check whether H_i and C_i are in the aggregated bloom filters HBF_j and CBF_j, and $T_{vercomm}$ to verify the commitment made by a user. As such, the total time required is $0.4 + 0.2 + 9.0 = 9.6$ msec.

8 Performance Analysis

8.1 Discussion on Bloom Filter Approach

This section analyses our proposed bloom filter approach for the BAN and NAN gateway smart meters to aggregate H and C values from household smart meters. We show that the probability of having false positives is very small if we set the parameters for the bloom filters appropriately.

The probabily of having a false positive in our bloom filter apporach is equal to the probability that all k bits are set in the bloom filter. Thus such a probability is $Pr(PF) = (1 - (1 - \frac{1}{m})^{kn})^k \sim (1 - e^{-\frac{kn}{m}})^k$. This probability can be minimized when $k = \frac{m \ln 2}{n}$. Hence we set the number of hash functions to $\frac{m \ln 2}{n}$ in our scheme and $Pr(PF) \sim (0.6185)^{\frac{m}{n}}$. We represent this function graphically in Fig. 2. It can be shown that when $\frac{m}{n} = 5$, $Pr(PF)$ is about 0.09. When $\frac{m}{n} = 10$, $Pr(PF)$ drops to 0.008 only. Therefore, we suggest users to set the size of bloom filters to 10 times the number of households for which the information will be put into the bloom filters. For example, if a building contains 200 households, the size of HBFs and CBFs should be set to 2,000 bits (i.e., 250 bytes).

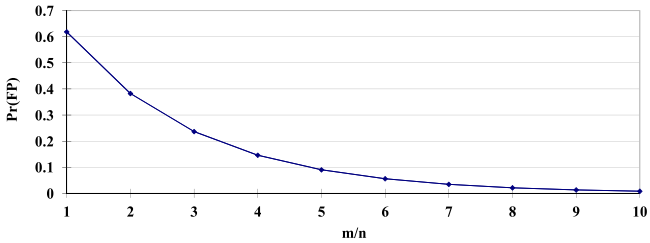

Fig. 2. Pr(PF) with different values of m/n.

8.2 Experimental Results

This section explains our experimental results. In our experiments, we consider a virtual city like Hong Kong for which by 2011, there were 2,367,000 households and about 7,000 buildings [29]. We roughly fine-tune these numbers to reflect the situation in cities of different population densities. To be concise, we vary the number of households from 1 to 10 millions while fixing the number of buildings to 7,000. We then investigate the gain in terms of total traffic volume by our gateway-assisted aggregation approach. We also repeat this experiment by varying the number of sub-periods from 4 to 48. Next, we introduce some attacking traffic into the network and show that our gateway-assisted authentication approach is effective in minimizing its impact. Note that all previous efforts have different security assumptions from ours. For example, [1] and [2] only considers message authentication issue but not end user privacy preservation issue. It is difficult to directly compare with them. Therefore, in our experiments below, we just compare our scheme with and without aggregation and filtering at gateway smart meters.

In the first experiment, we fix the number of sub-periods to 24 and vary the number of households from 1 to 10 millions in steps of 1 million and investigate its impact on the total traffic volume with and without aggregation at gateway smart meters. The result is shown in Fig. 3. It is found that the total traffic volume increases linearly with the number of households. This is normal as more power usage plans are submitted when there are more households in the city. Aggregation at gateway smart meters yields significant decrease in the total traffic volume. No matter how the number of households varies, the total traffic volume with aggregation is about 66 percent lower than that without aggregation. This shows that aggregation at gateway smart meters is helpful in reducing the total traffic volume.

Next, we fix the number of households to 2 millions (i.e., roughly the statistics in Hong Kong) and vary the number of sub-periods from four to 48 in steps of 4 and investigate its impact on the total traffic volume with and without aggregation at gateway smart meters. Obviously, with more

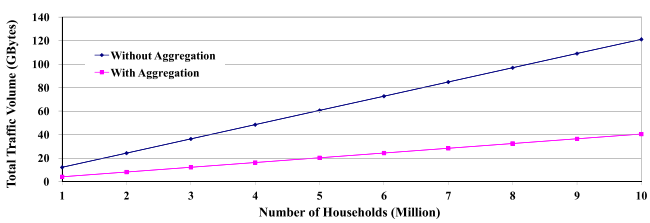

Fig. 3. Total traffic volume versus number of households.

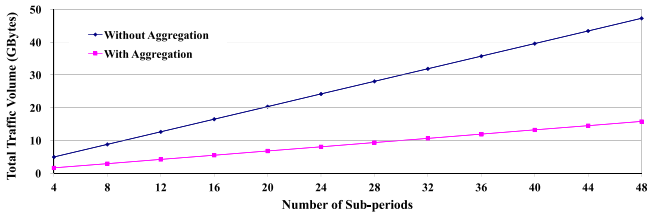

Fig. 4. Total traffic volume versus number of sub-periods.

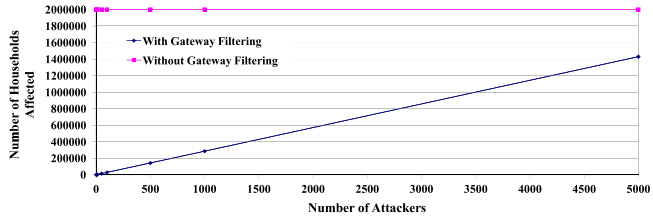

Fig. 5. Number of households affected versus number of attackers.

sub-periods defined, an end user's power usage plan can reflect his/her actual power usage more accurately. For example, if an end user requires additional power for just 1 hour but there are only 4 sub-periods defined in a day (i.e., 6 hours in each sub-period), a waste of 5 hours' additional power will be resulted. The result is shown in Fig. 4. It is found that the total traffic volume increases linearly with the number of sub-periods. This is normal as more power usage prediction information is submitted when there are more sub-periods defined. Aggregation at gateway smart meters yields significant decrease in the total traffic volume. No matter how the number of sub-periods varies, the total traffic volume with aggregation is about 66 percent lower than that without aggregation. This again shows that aggregation at gateway smart meters is helpful in reducing the total traffic volume.

Finally, we consider the case that there are some attackers in the smart grid network. We fix the number of households to 2 millions and the number of sub-periods to 24. We then introduce different numbers of attackers into the network and investigate how many households are affected with and without message filtering at gateway smart meters. We consider a household is affected if its message to the control center experiences any network congestion or even packet dropping. Due to the nature of our scheme, having different numbers of attackers inside the same building yields no difference in terms of performance. Therefore, we assume that the attackers are evenly distributed into different buildings in the city. For example, if there are 10 attackers, we assume that they are evenly distributed into 10 different buildings.

Without message filtering at gateway smart meters, all traffic from attackers will be forwarded to the control center. Thus the control center becomes a single point of failure and cannot handle (or can only handle at lower speed) power usage plans submitted by normal users. As a result, almost all households in the city are affected. With message filtering at BAN gateway smart meters, traffic from attackers are filtered at the building level. As a result, only households located in the same building as the attacker are affected. The control center and thus households located in buildings without attacker will not be affected at all. We summarize this result in Fig. 5. This shows that message filtering at gateway smart meters is helpful in reducing the impact of attacking traffic.

9 Conclusions

In this paper, we proposed a scheme for privacy-preserving recording and gateway-assisted authentication of power usage information for the smart grid network. By observing that gateway smart meters are usually physically harder to be compromised, they are utilized to help authenticate messages sent by household smart meters before these messages actually reach the control center. We assume that smart meters are tamper-resistant devices which are secure from data cracking or operation disturbance. A major feature of our scheme is that the privacy of any end user, especially their future power usage plan, can be preserved while at the same time the control center can generate a proper amount of electricity. That is, our scheme allows the control center to record an end user's additional power request or power reduction plan one day ahead, one week ahead, one month ahead or even one year ahead anonymously. This goal is achieved using the concept of cryptographic commitment. On the other hand, our scheme allows gateway smart meters to aggregate power usage plans from multiple household smart meters. This goal is achieved using the techniques of homomorphic encryption and bloom filter. Through experimental study, we show that aggregation at gateway smart meters can help reduce the total traffic volume by 66 percent. On the other hand, message filtering at BAN gateway smart meters can help to significantly reduce the impact of attacking traffic. In particular, only households in buildings with attackers are affected.

Note that the two security issues we address are the first step towards a secure smart grid system. Further investigation probably would come up with better solutions and other security issues (e.g., DDoS attacks) need to be addressed and integrated into the system. We will also study possible privacy leakage caused when the control center has built up long-term load profiles of end users. In this paper, we assume that TPM is tamper-resistant such that keys stored in them are difficult to be cracked or modified. We will have an in-depth investigation about how to achieve this assumption in the future. We are now in the process of collecting statistics on real power usage. We believe that if we have such statistics, we could perform a more realistic simulation in the future. Besides, we are considering other secure applications in smart grid networks.

Acknowledgments

This research was supported in part by the Collaborative Research Fund of the Research Grants Council of Hong Kong under Grant No. HKU10/CRF/10.

References

[1] M. M. Fouda, Z. M. Fadlullah, N. Kato, R. Lu, and X. S. Shen, "Towards a light-weight message authentication mechanism tailored for smart grid communications," in *Proc. First Int. Workshop Secur. Comput., Netw. Commun.*, Jun. 2011, pp. 1018–1023.

[2] M. M. Fouda, Z. M. Fadlullah, N. Kato, R. Lu, and X. S. Shen, "A Lightweight Message Authentication Scheme for Smart Grid Communications," *IEEE Trans. Smart Grid*, vol. 2, no. 4, pp. 675–685, Nov. 2011.

[3] C. Romero. (2013). LightsOn: Virtual power plants: Making distributed energy resources actionable in smart grid commercial operations, *Elect. Energy Online.com* [Online]. Available: http://www.electricenergyonline.com/?page=show_article&article=367.
[4] ARC Advisory Group. SCADA systems for smart grid [Online]. Available: http://www.arcweb.com/Research/Studies/Pages/SCADA-Power.aspx.
[5] Department of Energy and Climate Change, "Smart metering implementation programme prospectus," Ofgem/Ofgem E-Serve, Jul. 2010.
[6] Smart Metering Team, "Smart metering implementation programme: Implementation strategy," Ofgem/Ofgem E-Serve, Jul. 2010.
[7] Juniper Networks Inc., "Architecture for secure SCADA and distributed control system networks," 2009.
[8] Office of the National Coordinator for Smart Grid Interoperability, *NIST special publication 1108: NIST framework and roadmap for smart grid interoperability standards, release* 1.0, Nat. Inst. Standards Tech, Jan. 2010.
[9] H. Khurana, M. Hadley, N. Lu, and D. A. Frincke, "Smart-grid security issues," *IEEE Security Privacy*, vol. 8, no. 1, pp. 81–85, Jan./Feb. 2010.
[10] Y. S. Choi, J. T. Oh, J. S. Jang, and J. C. Ryou, "Integrated DDoS attack defense infrastructure for effective attack prevention," in *Proc. IEEE 2nd Int. Conf. Inform. Tech. Convergence Services*, Aug. 2010, pp. 1–6.
[11] The Smart Grid Interoperability Panel Cyber Security Working Group, "Second draft NISTIR 7628 smart grid cyber security strategy and requirements," Nat. Inst. Standards Tech, Feb. 2010.
[12] SmartGrids, European smartgrids technology platform: Vision and strategy for Europe's electricity networks of the future, *Eur. Comm., Directorate-Gen. Res., Sustainable Energy Syst., EUR 22040*, 2006.
[13] Electric Power Research Institute, (2001). Intelligrid [Online]. Available: http://intelligrid.epri.com/
[14] US Department of Energy, Grid 2030: A national vision for electricity's second 100 years, 2003.
[15] V. O. K. Li, F. F. Wu, and J. Zhong, "Communication requirements for risk-limiting dispatch in smart grid," in *Proc. IEEE Workshop Smart Grid Commun.*, May 2010, pp. 1–5.
[16] Y. Liu, P. Ning, and M. K. Reiter, "False data injection attacks against state estimation in electric power grids," in *Proc. 16th ACM Conf. Comput. Commun. Security*, Nov. 2009, pp. 21–32.
[17] A. Rial and G. Danezis, "Privacy-preserving smart metering," in *Proc. 10th Annu. ACM Workshop Privacy Electron. Soc.*, Dec. 2011, pp. 49–60.
[18] F. M. Tabrizi and K. Pattabiraman, "A model for security analysis of smart meters," *Proc. IEEE/IFIP 42nd Int. Conf. Dependable Syst. Netw. Workshops (DSN-W)*, pp. 25–28, Jun. 2012.
[19] E. Naone. (2009). Meters for the smart grid [Online]. Available: http://www.technologyreview.com/hack/414820/meters-for-the-smart-grid/?a=f
[20] T. Goodspeed. (2010). Smartgird Skunkworks. [Online]. Available: http://travisgoodspeed.blogspot.hk/2010/03/smartgrid-skunkworks.html
[21] K. Song, D. Seo, H. Park, H. Lee, and A. Perrig, "OMAP: One-way memory attestation protocol for smart meters," in *Proc. 9th IEEE Int. Symp. Parallel Distrib. Process. Appl. Workshops*, 2011, pp. 111–118.
[22] Trusted Computing Group. (2013). Trusted platform module (TPM) specifications [Online]. Available: http://www.trustedcomputinggroup.org/resources/tpm_main_specification
[23] B. Kaliski and J. Staddon, *RSA Cryptography Specifications Version 2.0*, IETF RFC2437, 1998.
[24] D. Eastlake and P. Jones, *US Secure Hash Algorithm 1 (SHA1)*, IETF RFC3174, 2001.
[25] R. Rivest, *The MD5 Message-Digest Algorithm*, IETF RFC1321, 1992.
[26] P. Paillier, "Public-key cryptosystems based on composite degree residuosity classes," in *Proc. 17th Int. Conf. Theory Appl. Cryptographic Tech.*, May 1999, pp. 223–238.
[27] C. Zhang, R. Lu, X. Lin, P. H. Ho, and X. Shen, "An efficient identity-based batch verification scheme for vehicular sensor networks," in *Proc. IEEE Conf. Comput. Commun.*, Apr. 2008, pp. 816–824.
[28] FriendlyARM. FriendlyARM [Online]. Available: http://www.friendlyarm.net/
[29] Census and HKSAR Statistics Department. (2012). Hong Kong Statistics. [Online]. Available: http://www.censtatd.gov.hk/hkstat/sub/bbs.jsp

Tat Wing Chim received the BEng, MPhil and PhD degrees in information engineering, electrical and electronic engineering, and computer science, respectively, from the University of Hong Kong, in 2002, 2004 and 2011, respectively. He is currently a post-doctoral fellow in the Department of Computer Science and the Department of Electrical and Electronic Engineering, University of Hong Kong and being funded by Prof. Victor O.K. Li from the Department of Electrical and Electronic Engineering, University of Hong Kong. His research interests include information security and network routing.

Siu-Ming Yiu received the PhD degree in computer science from the Department of Computer Science, The University of Hong Kong, in 1996. He is currently an associate professor in the same department. His research interests include information security, cryptography, and bioinformatics.

Victor O. K. Li received the SB, SM, EE, and ScD degrees in electrical engineering and computer science from the Massachusetts Institute of Technology, Cambridge, MA, in 1977, 1979, 1980, and 1981, respectively. He joined the University of Southern California (USC), Los Angeles, CA, in February 1981, and became professor of Electrical Engineering and the director of the USC Communication Sciences Institute. Since September 1997 he has been with the University of Hong Kong, Hong Kong, China, where he is the chair professor of Information Engineering, the associate dean of Engineering, and the head of the Department of Electrical and Electronic Engineering. He was also the managing director of Versitech Ltd. (http://www.versitech.com.hk/), the technology transfer and commercial arm of the University, and on various corporate boards. His research is in information technology, especially its application to clean energy and environment. He is the co-director of the Area of Excellence in Information Technology funded by the Hong Kong government. Sought by government, industry, and academic organizations, he has lectured and consulted extensively around the world. He chaired the Computer Communications Technical Committee of the IEEE Communications Society 1987-1989, and the Los Angeles Chapter of the IEEE Information Theory Group 1983-1985. He co-founded the International Conference on Computer Communications and Networks (IC3N), and chaired its Steering Committee 1992-1997. He also chaired various international workshops and conferences, including, most recently, IEEE INFOCOM 2004 and IEEE HPSR 2005. He was an editor of *IEEE Network, IEEE JSAC Wireless Communications Series*, *Telecommunication Systems*, *ACM/Springer Wireless Networks*, and *IEEE Communications Surveys and Tutorials*. He also guest edited special issues of *IEEE Journal on Selected Areas in Communications*, *Computer Networks* and *ISDN Systems*, and *KICS/IEEE Journal of Communications and Networking*. He has been appointed to the Hong Kong Information Infrastructure Advisory Committee by the Chief Executive of the Hong Kong Special Administrative Region (HKSAR). He was a part-time member of the Central Policy Unit of the Hong Kong Government. He was also on the Innovation and Technology Fund (Electronics) Vetting Committee, the Small Entrepreneur Research Assistance Programme Committee, the Engineering Panel of the Research Grants Council, and the Task Force for the Hong Kong Academic and Research Network (HARNET) Development Fund of the University Grants Committee. He was a distinguished lecturer at the University of California at San Diego, at the National Science Council of Taiwan, and at the California Polytechnic Institute. He has also delivered keynote speeches at many international conferences. He has received numerous awards, including the PRC Ministry of Education Changjiang chair professorship at Tsinghua University, Beijing, the UK Royal Academy of Engineering Senior Visiting Fellowship in Communications, the Outstanding Researcher Award of the University of Hong Kong, the Croucher Foundation Senior Research Fellowship, and the Order of the Bronze Bauhinia Star, Government of HKSAR, China. He was elected an IEEE fellow in 1992. He is also a fellow of the HKIE and the IAE.

Lucas C.K. Hui received the BSc and MPhil degrees in computer science from the University of Hong Kong, and the MSc and PhD degrees in computer science from the University of California, Davis. He is the founder and the honorary director of the Center for Information Security & Cryptography, and concurrently an associate professor in the Department of Computer Science, The University of Hong Kong. His research interests include information security, computer crime, cryptographic systems, and electronic commerce security. He is a member of the HKIE and a senior member of the IEEE.

Jin Zhong received the BSc degree from Tsinghua University in 1995, MSc degree from the Electric Power Research Institute, China, in 1998 and the PhD degree from the Chalmers University of Technology, Sweden, in 2003. From June 2002 to Nov. 2002, she was a visiting scholar at Washington State University. She is currently an associate professor in the Department of Electrical and Electronic Engineering.