A Practical Smart Metering System Supporting Privacy Preserving Billing and Load Monitoring

Hsiao-Ying Lin[1], Wen-Guey Tzeng[2], Shiuan-Tzuo Shen[2], and Bao-Shuh P. Lin[1]

[1] Intelligent Information and Communications Research Center,
National Chiao Tung University, Taiwan
hsiaoying.lin@gmail.com, bplin@mail.nctu.edu.tw
[2] Department of Computer Science, National Chiao Tung University, Taiwan
{vink,wgtzeng}@cs.nctu.edu.tw

Abstract. Fine-grained meter readings enable applications in an advanced metering infrastructure. However, those meter readings threaten personal privacy by implying a sketch of daily activities of households. The privacy issue has been addressed in smart metering systems by either a trusted third party assumption or cryptographic primitives. We address the privacy issue by using a trusted platform module and lightweight cryptographic primitives. Our smart metering system simultaneously supports the billing and load monitoring applications in a privacy preserving manner. It allows an electricity service provider obtain sums of meter readings over a time period and a monitoring center obtain sums of meter readings from meters in an area at some recent time unit while keeping individual meter reading private. Moreover, we formally prove that our system is privacy preserving. Our system provides a simple yet very practical solution to a privacy preserving smart metering system.

Keywords: Trusted platform module, smart metering, privacy preserving technique, secure aggregation, pseudorandom number generator.

1 Introduction

The emergence of smart grids has established a trend towards building our next generation of power grid systems. As shown in Fig. 1, new features include two-way power flows and mutual communications between electricity entities. Smart grids integrate intelligence and automation into the conventional power grid system to increase energy efficiency and improve system reliability and quality. We can build advanced applications upon smart grids, such as load monitoring, automatic billing, dynamic pricing, and power generation planning.

One essential technology of smart grids is fine-grained meter reading within a very short period of time per household. However, meter readings of a household reveal detailed information about daily activities of the household and used appliances during a specific time period [7,14,11]. Fine-grained meter readings cause serious privacy issues. Actually, the granularity of meter readings often exceeds the need of some underlying applications. Current smart meters record

F. Bao, P. Samarati, and J. Zhou (Eds.): ACNS 2012, LNCS 7341, pp. 544–560, 2012.

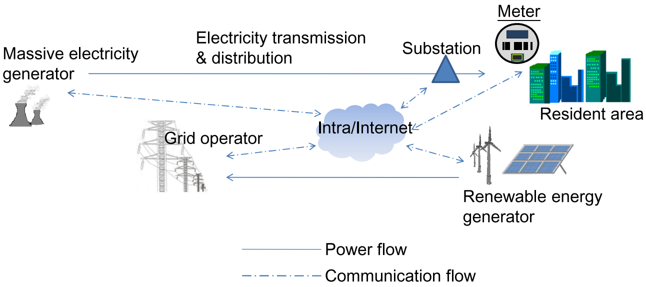

Meter
Massive electricity generator
Electricity transmission & distribution
Substation
Resident area
Intra/Internet
Grid operator
Renewable energy generator
Power flow
Communication flow

Fig. 1. Power is massively generated by a power station and transmitted by a grid operator from the generator to end-consumers. Local renewable energy can also be transmitted to other entities.

electricity usage every 5 to 60 minutes [8]. The next generation of smart meters will upgrade a time unit to seconds. In billing applications, the electricity service provider (ESP) only needs the amount of power consumption per hour to compute a bill. For example, in Ontario, Canada, the time-of-use price service during winters only needs the consumption data over two hours in an on-peak time period, six hours in a mid-peak time period, and 12 hours in an off-peak time period [1]. In load monitoring applications, the load monitoring center (LMC) collects the amount of electricity usage over a local area in order to monitor current activities of the power grid. LMC requires consumption data in much finer time granularity than ESP does. Nevertheless, LMC only needs the total power consumption over the area at recent time units.

To address the privacy issues against service providers, an approach of secure aggregation is proposed. By secure aggregation techniques, a service provider can only get an aggregated result of meter readings while individual meter reading remains private. For the billing application, previous works use public key homomorphic encryption schemes, commitment schemes, or a trusted third party to securely aggregate meter readings of a meter. For the load monitoring application, previous works use public key homomorphic encryption schemes, secret sharing techniques, or distributed random noise generation to securely aggregate meter readings of meters in an area.

On the other hand, many manufactures of smart meters use a hardware component to address various cyber-security issues. For example, Atmel provides electricity meters with a hardware security component for cryptographic authentication. Embedding a trusted platform module chip (TPM) into a smart meter is a general practice for securing metering services [12,13,9]. We shall assume that a TPM is embedded into a smart meter for providing securing functions.

It is a challenge to design a smart metering system that simultaneously supports multiple privacy preserving applications without using a trusted third party and public key cryptographic primitives. We focus on the billing and load monitoring applications and consider the privacy requirements for them. Our main contribution is to propose a practical privacy preserving smart metering system that supports billing and load monitoring applications with TPM technologies. Our system uses a pseudorandom number generator and hash functions supported by TPM technologies. Features of our smart metering system are as follows:

- ESP can only query a meter for a sum of meter readings over a time period. Each meter reading remains private against ESP.
- LMC can only query a sum of meter readings from meters in an area at a time unit. Each meter reading remains private against LMC.
- Meter readings are securely stored in a semi-trusted storage system.
- Meters can freely join or leave our smart metering system without overhead.

Moreover, we formally define a privacy model with respect to time-series meter readings to capture privacy requirements and prove that our smart metering system meets the requirements.

2 Related Work

We briefly introduce existing privacy preserving protocols of smart metering systems and TPM technologies.

Privacy preserving metering protocols. Anonymous technology is suggested by NIST to anonymizing traces of meter readings [2]. For the billing application, Petrlic proposed a solution by using pseudonym of households against ESP where the grid operator to be a fully trusted intermediate translator [15]. Jawurek et al. constructed a privacy preserving billing protocol by integrating a homomorphic commitment scheme, zero knowledge proofs and a tamper-evident meter [9]. Meter readings are committed and aggregated by using the homomorphic commitment scheme. Only the final bill is opened to ESP and the correctness of the computation is verified by using zero knowledge proofs. Rial and Danezis took a similar approach [16], where they replaced the tamper-evident meter by TPM.

For the load monitoring application, Garcia and Jacobs proposed a solution by using a trusted aggregator in a substation and an additively homomorphic encryption scheme [6]. Each meter encrypts meter readings by using LMC's public key. The aggregator aggregates encrypted meter readings and only sends the aggregated result to LMC. Shi et al. proposed a privacy model for aggregation of time-series data (such as meter readings) while individual datum remains private [17]. In their system, the number of meters is fixed after the system is setup. The system must to be reset when meters join or leave. Later, Shi et al. proposed a new solution by using the subset cover technique to tolerate leaving meters [5]. Kursawe et al. proposed a privacy friendly aggregation method [10]. An aggregator and meters secretly share 0 for multiple times in parallel such that no share of a meter is revealed. Ács and Castelluccia [3] proposed a solution by using random noise and secret sharing. Meters independently generate random noise and pairs of meters secretly share 0. Meter readings are masked by random noise and encrypted by secret shares. The sum of masked and encrypted meter readings gives a noisy sum of meter readings.

Bohi et al. proposed a privacy model and two approaches for the billing and load monitoring applications, respectively [4]. First, they used a trusted third party to compute the bill for the billing application. Second, they introduced random noises on meter readings, where the distribution of the noise has a known mean and variance. LMC gets only an approximate sum of meter readings while individual reading is private.

Our work is distinguished as it simultaneously addresses both applications but only requires a simple and lightweight use of TPM for generating pseudorandom numbers without a trusted third party and without mutual communications among meters.

TPM technologies. TPM is a microcontroller that offers facilities for secure generation of cryptographic keys, the ability to limit the use of keys, non-volatile storage and a hardware pseudorandom number generator. It enables platform attestation and cryptographic primitives, such as RSA and SHA-1. The TPM specification is defined by the trusted computing group and the latest version is TPM 1.2 revision 116[1].

A TPM chip itself is a solid component through platform attestation. It employs platform configuration registers to record configurations of platform and software, and prevents unauthorized modifications on these configurations. By verifying configurations, TPM assures that the platform is initialized from a secure and correct condition.

3 System Model

We describe our time notation, smart metering system, and the billing and load monitoring applications. We also brief privacy requirements. Detailed descriptions of privacy requirements are provided in Section 5.

3.1 Time Notations

Time is divided into basic *time units* $t_1, t_2, \cdots$. Let l be a fixed positive integer, where $l \geq 2$. We set l to be the minimum number of time units where ESP gets the sum of meter readings. Based on the parameter l, we define *time periods* and the *current time window*. A time period T consists of al continuous time units for any positive integer a. The current time window W is the latest continuous l time units t_{z-l+1},t_{z-l+2}, $\cdots$, t_z, where t_z is the current time unit.

3.2 Smart Metering System

Our smart metering system consists of meters, a storage system, ESP and LMC, as shown in Fig. 2. We assume that meters are purchased by households and deployed by the grid operator. Households trust the grid operator that it honestly deploys meters. We assume that meters are trusted, that is, meters honestly follow defined steps. We also assume that ESP and LMC are honest-but-curious, that is, they follow defined steps but try to dig out individual meter readings

[1] The specification is available as international standard ISO/IEC 11889.

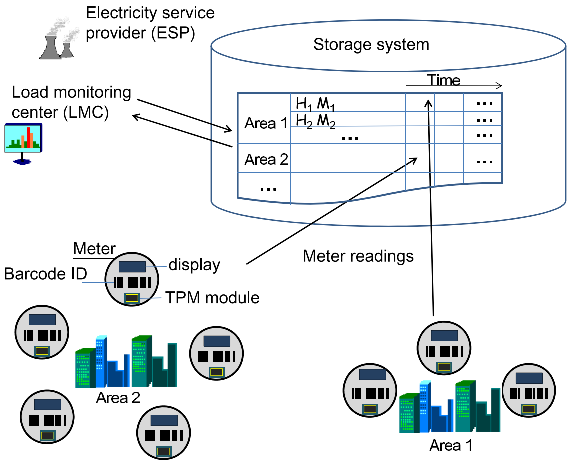


Fig. 2. Our system model consists of meters, a storage system, ESP, and LMC

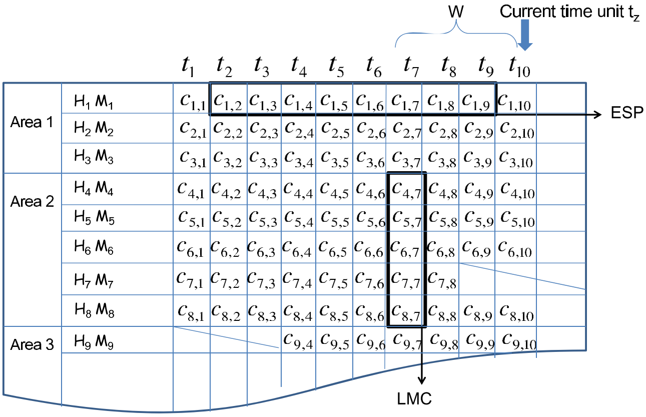
from what they obtain from communications. Moreover, we assume that ESP and LMC do not collude.

A household $\mathcal{H}_i$ has a meter $\mathcal{M}_i$ that records power consumption $d_{i,j}$ of $\mathcal{H}_i$ at time unit t_j. Households may physically move in or out an area. A meter $\mathcal{M}_i$ has a serial number $\mathcal{SN}_i$ assigned by the meter manufacture. A meter $\mathcal{M}_i$ *encrypts* a meter reading $d_{i,j}$ as $c_{i,j}$ and stores $c_{i,j}$ into the storage system. The storage system stores the encrypted meter reading according to the meter and the time unit. We assume that ESP and LMC can freely access the storage system after being authenticated by the storage system.

Meter readings are conceptually arranged in a matrix in the storage system, where a row represents meter readings of a household over time and a column represents meter readings of households in an area at a time unit. An example is shown in Fig. 3. From the time unit t_1, Areas 1 and 2 have 3 and 5 households, respectively. Each household $\mathcal{H}_i$ has a meter $\mathcal{M}_i$ for $1 \leq i \leq 9$. At t_4, new household $\mathcal{H}_9$ moves in Area 3 and then a row of $\mathcal{M}_9$ is added in the matrix. When household $\mathcal{H}_7$ moves out Area 2 at t_9, the row of $\mathcal{H}_7$ in Area 2 is deleted from the matrix.

3.3 Supporting Billing Applications

ESP is allowed to query the meter for decryption information of a sum of meter readings over a time period T. ESP sums up encrypted meter readings over T. By the decryption information, ESP decrypts the encrypted sum to obtain the power consumption of the household over T. In the example in Fig. 3, l is set to 4. ESP queries the meter $\mathcal{M}_1$ for decryption information of the time period $T = (t_2, t_3, t_4, t_5, t_6, t_7, t_8, t_9)$ and decrypts the encrypted sum $c = \sum_{j=2}^{9} c_{1,j}$ to obtain the sum $\xi = \sum_{j=2}^{9} d_{1,j}$ of the meter readings between t_2 and t_9.



		t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}
Area 1	H_1 M_1	$c_{1,1}$	$c_{1,2}$	$c_{1,3}$	$c_{1,4}$	$c_{1,5}$	$c_{1,6}$	$c_{1,7}$	$c_{1,8}$	$c_{1,9}$	$c_{1,10}$
	H_2 M_2	$c_{2,1}$	$c_{2,2}$	$c_{2,3}$	$c_{2,4}$	$c_{2,5}$	$c_{2,6}$	$c_{2,7}$	$c_{2,8}$	$c_{2,9}$	$c_{2,10}$
	H_3 M_3	$c_{3,1}$	$c_{3,2}$	$c_{3,3}$	$c_{3,4}$	$c_{3,5}$	$c_{3,6}$	$c_{3,7}$	$c_{3,8}$	$c_{3,9}$	$c_{3,10}$
Area 2	H_4 M_4	$c_{4,1}$	$c_{4,2}$	$c_{4,3}$	$c_{4,4}$	$c_{4,5}$	$c_{4,6}$	$c_{4,7}$	$c_{4,8}$	$c_{4,9}$	$c_{4,10}$
	H_5 M_5	$c_{5,1}$	$c_{5,2}$	$c_{5,3}$	$c_{5,4}$	$c_{5,5}$	$c_{5,6}$	$c_{5,7}$	$c_{5,8}$	$c_{5,9}$	$c_{5,10}$
	H_6 M_6	$c_{6,1}$	$c_{6,2}$	$c_{6,3}$	$c_{6,4}$	$c_{6,5}$	$c_{6,6}$	$c_{6,7}$	$c_{6,8}$	$c_{6,9}$	$c_{6,10}$
	H_7 M_7	$c_{7,1}$	$c_{7,2}$	$c_{7,3}$	$c_{7,4}$	$c_{7,5}$	$c_{7,6}$	$c_{7,7}$	$c_{7,8}$		
	H_8 M_8	$c_{8,1}$	$c_{8,2}$	$c_{8,3}$	$c_{8,4}$	$c_{8,5}$	$c_{8,6}$	$c_{8,7}$	$c_{8,8}$	$c_{8,9}$	$c_{8,10}$
Area 3	H_9 M_9				$c_{9,4}$	$c_{9,5}$	$c_{9,6}$	$c_{9,7}$	$c_{9,8}$	$c_{9,9}$	$c_{9,10}$

Fig. 3. Meter readings are conceptually arranged in a matrix

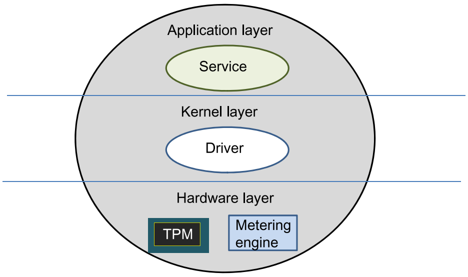


Fig. 4. We model a meter in three layers

The correctness for ESP is that ESP obtains the correct sum ξ of a household over T. The privacy requirement for ESP is that it cannot get individual meter readings of a household.

3.4 Supporting Load Monitoring Applications

LMC is allowed to query meters in an area for *approximate* decryption information of a time unit t_j in the current time window W. LMC sums up encrypted meter readings in the area at t_j to get the encrypted sum. By the approximate decryption information, LMC decrypts and gets an approximate overall power consumption of households in the area at t_j. In the same example, the current time unit is t_{10} and the current time window is (t_7, t_8, t_9, t_{10}). LMC can query meters in Area 2 for decryption information at t_7 and decrypts the encrypted sum $c = \sum_{i=4}^{8} c_{i,7}$ to obtain an approximate sum for $\xi = \sum_{i=4}^{8} d_{i,7}$.

The correctness for LMC is that LMC obtains a good approximate sum $\tilde{\xi}$ for ξ. We formulate the approximation by the error ratio $\omega = |\tilde{\xi} - \xi|/\xi$, a threshold value ϵ and a confidence probability δ as $\Pr[\omega \leq \epsilon] > 1-\delta$. With sufficiently small ϵ and δ, LMC obtains a good approximate sum $\tilde{\xi}$ for ξ with a higher probability. The privacy requirement for LMC is that it cannot get exact individual meter readings of a household.

3.5 Meter Model

As shown in Fig. 4, we have a three-layer model for a meter. The hardware layer consists of hardware components, such as a TPM chip, a metering engine, a processing and communication engine. The kernel layer consists of drivers of hardware components. We assume that a driver is in charge of meter readings of the metering system. The application layer is built upon the kernel layer to provide services, such as a web interface for observing current meter readings.

The power consumption is often measured in kWh. Since we consider a finer time granularity, the unit of measurement is changed to Wh so that integer representation is enough. Moreover, we assume that meter readings (in integers) at a time unit are much less than a defined number p. From the statistics of U.S. Energy Information Administration, in 2009, the average power consumption per household per month is 908 kWh. That is $105Wh$ per 5 minutes. Thus, we set p to be of length 64 bits.

4 Privacy Preserving Smart Metering System

We describe our smart metering system and two types of queries supporting billing and load monitoring applications.

4.1 Metering System Construction

We assume that a meter is deployed or reset by the grid operator when a household moves in an area. At the beginning, the metering system consists of a storage system, ESP and LMC. Later, meters join in. Choose a large number p, where $p \geq 2\sqrt{p}$. Let the initial time unit be t_1. Let the pseudorandom number generator be g, where $g : \{0,1\}^\tau \times \{0,1\}^\lambda \rightarrow \mathcal{Z}_p$. Let h and h' be cryptographic hash functions, where $h : \{0,1\}^* \rightarrow \{0,1\}^\lambda$ and $h' : \{0,1\}^* \rightarrow \{0,1\}^\tau$. A meter $\mathcal{M}_i$ runs as follows.

Meter initialization.

1. $\mathcal{M}_i$ takes a user input as a seed s_i. The TPM of $\mathcal{M}_i$ generates a master key k_i by using the seed s_i, the serial number $\mathcal{SN}_i$, and the hash function h', where $k_i = h'(s_i||\mathcal{SN}_i)$ and $||$ is the operator of concatenation. The master key k_i is then securely stored in non-volatile storage of the TPM of $\mathcal{M}_i$.
2. The driver of $\mathcal{M}_i$ creates and initializes l first-in first-out memory slots as 0.
3. The TPM of $\mathcal{M}_i$ generates l pseudorandom numbers $r_{i,1}, r_{i,2}, \cdots, r_{i,l-1}$ and $R_{i,1}$, where

$$\begin{aligned} &r_{i,j} = g(k_i, t_j), 1 \leq j \leq l-1, \text{ and} \\ &R_{i,1} = g(k_i, h(t_1||t_2||\cdots||t_l)) \end{aligned}$$

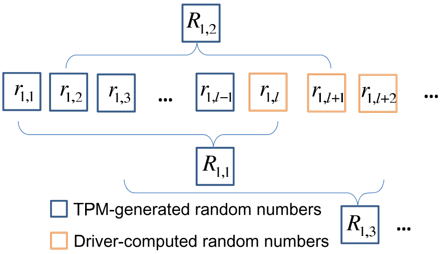


Fig. 5. The TPM generates random numbers $r_{1,1}, r_{1,2}, \cdots, r_{1,l-1}$ and $R_{1,1}$ in the initialization part and a random number $R_{1,i+1}$ at time unit t_i. Random numbers $r_{1,i}$ for $i \geq l$ are computed on the fly.

Then, the TPM of $\mathcal{M}_i$ passes all pseudorandom numbers to the driver of $\mathcal{M}_i$.

4. The driver of $\mathcal{M}_i$ computes $r_{i,l}$ as follows:

$$r_{i,l} = \left(R_{i,1} - \sum_{j=1}^{l-1} r_{1,j} \right) \mod p$$

Then, the driver stores l pseudorandom numbers $r_{i,1}, r_{i,2}, \cdots, r_{i,l}$ in memory slots.

Storage of meter readings at t_j, $j \geq 1$.

1. $\mathcal{M}_i$ measures the consumption $d_{i,j}$ and encrypts it as $c_{i,j}$, where

$$c_{i,j} = (d_{i,j} + r_{i,j}) \mod p$$

$c_{i,j}$ is sent and stored to the storage system.

2. The TPM of $\mathcal{M}_i$ generates a random number $R_{i,j+1}$ and passes it to the driver of $\mathcal{M}_i$, where

$$R_{i,j+1} = g(k_i, h(t_{j+1}||t_{j+2}||\cdots||t_{j+l})).$$

The driver of $\mathcal{M}_i$ computes

$$r_{i,j+l} = \left(R_{i,j+1} - \sum_{\alpha=j+1}^{j+l-1} r_{i,\alpha} \right) \mod p.$$

The driver then replaces $r_{i,j}$ with $r_{i,j+l}$ in the memory slot.

An example of $\mathcal{M}_1$ is shown in Fig. 5. The TPM of $\mathcal{M}_1$ generates l random numbers in the initialization part and generates a random number at each time unit on the fly. For any time period T with l continuous time units, a random number generated by the TPM of $\mathcal{M}_1$ helps decrypt the sum of meter readings over T. The snapshots at t_l of $\mathcal{M}_1$ is shown in Fig. 6. At any time unit, the driver of $\mathcal{M}_1$ maintains the random numbers used for time units in W.

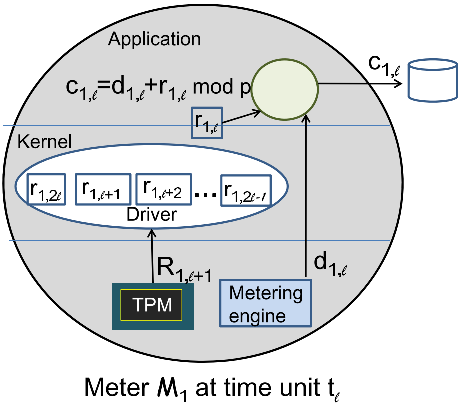


Fig. 6. Snapshot of the meter $\mathcal{M}_1$ at t_l

4.2 Supporting Billing Application

To compute a bill of a household $\mathcal{H}_i$, ESP queries the meter $\mathcal{M}_i$ of $\mathcal{H}_i$ for a time period T of al continuous time units, where $T = (t_\beta, t_{\beta+1}, \cdots, t_{\beta+al-1})$. The application layer of $\mathcal{M}_i$ divides the time period into a sub-periods and asks the TPM to regenerate the corresponding random numbers $R_{i,\beta}$, $R_{i,\beta+l}$, $R_{i,\beta+2l}$, $\cdots$, $R_{i,\beta+(a-1)l}$. The application layer of $\mathcal{M}_i$ computes the sum B of random numbers and sends B to ESP, where

$$B = \left(\sum_{j=0}^{a-1} R_{i,\beta+jl}\right) \mod p$$

ESP gets encrypted meter readings ($c_{i,\beta}$, $c_{i,\beta+1}, \cdots$, $c_{i,\beta+al-1}$) over T from the storage system and computes

$$\sum_{j=\beta}^{\beta+al-1} d_{i,j} = \left(\sum_{j=\beta}^{\beta+al-1} c_{i,j} - B\right) \mod p \qquad (1)$$

ESP correctness is that ESP can obtain the consumption of a household $\mathcal{H}_i$ over T. Note that for $0 \leq j \leq a-1$,

$$R_{i,\beta+jl} = (\sum_{k=0}^{l-1} r_{i,\beta+jl+k}) \mod p.$$

Thus,

$$\left(\sum_{j=\beta}^{\beta+al-1} c_{i,j} - B\right) \mod p = \left(\sum_{j=\beta}^{\beta+al-1} (d_{i,j} + r_{i,j}) - B\right) \mod p$$
$$= \left(\sum_{j=\beta}^{\beta+al-1} d_{i,j}\right) \mod p \qquad (2)$$

By Equation (2), Equation (1) holds when

$$\sum_{j=\beta}^{\beta+al-1} d_{i,j} = (\sum_{j=\beta}^{\beta+al-1} d_{i,j}) \mod p$$

That is, the sum of meter readings must be less than p. By choosing a large p, ESP correctness is guaranteed. In practice, it is sufficient to set p to be of 64-bits when each meter reading is of 32-bits and al is up to 2^{32}.

4.3 Supporting Load Monitoring Application

To monitor the power consumption in an area, LMC queries meters in an area at t_j in current time widow W. A meter $\mathcal{M}_i$ in the area should reply. The driver of $\mathcal{M}_i$ chooses a random number n according to a normal distribution $N(0, \sigma^2)$ with the mean 0 and the variance σ^2 and computes the noise $n_{i,j}$ as the floor of the chosen random number n, i.e. $n_{i,j} = \lfloor n \rfloor$. The variance σ shall be defined later. The driver of $\mathcal{M}_i$ then passes a noised random number $\tilde{r}_{i,j}$ to LMC, where

$$\tilde{r}_{i,j} = (r_{i,j} + n_{i,j} - \lceil \sqrt{p} \rceil) \mod p$$

Recall that the stored meter reading $c_{i,j} = (d_{i,j} + r_{i,j}) \mod p$. By $\tilde{r}_{i,j}$, LMC computes a noised meter reading $\tilde{d}_{i,j}$ of the meter $\mathcal{M}_i$ as follows:

$$\begin{aligned}\tilde{d}_{i,j} &= (c_{i,j} - \tilde{r}_{i,j} \mod p) - \lceil \sqrt{p} \rceil \\ &= (d_{i,j} - n_{i,j} + \lceil \sqrt{p} \rceil \mod p) - \lceil \sqrt{p} \rceil\end{aligned}$$

The number $\sqrt{p}$ is used to prevent an overflowing issue for correctness. Note that $\tilde{d}_{i,j}$ may be negative. To obtain $\tilde{d}_{i,j} = d_{i,j} - n_{i,j}$, we need

$$d_{i,j} + \lceil \sqrt{p} \rceil \geq n_{i,j} \geq d_{i,j} - p + \lceil \sqrt{p} \rceil$$

Since $p \geq 2\sqrt{p}$, we bound the probability by

$$\Pr[|n_{i,j}| \leq d_{i,j} + \lceil \sqrt{p} \rceil] \geq 1 - \frac{\sigma^2}{p}$$

Since p is very large and σ is sufficiently small, the error probability is negligible.

Let LMC obtain m noised meter readings $\tilde{d}_{i_\alpha,j}$ from m meters $\mathcal{M}_{i_\alpha}$, where $1 \leq \alpha \leq m$. LMC computes an approximate value $\tilde{\xi}$ for the overall consumption ξ at t_j in the area, where

$$\tilde{\xi} = \sum_{\alpha=1}^{m} \tilde{d}_{i_\alpha,j} \quad \text{and} \quad \xi = \sum_{\alpha=1}^{m} d_{i_\alpha,j}$$

Since some meters may fail to reply due to various reasons, LMC needs to set a maximal waiting time period T_{max}.

LMC correctness requires that LMC obtains an approximate value for the overall consumption. The error between the approximate sum $\tilde{\xi}$ and ξ depends on the number m and the variance σ^2. For m meter readings at t_j, let $x = \tilde{\xi} - \xi = \sum_{\alpha=1}^{m} n_{i_\alpha,j}$. We measure the error by using the error ratio $\omega = |x|/\xi$. Let $\hat{d}$ be the average value of meter readings per time unit. Thus, we assume $\xi = m\hat{d}$.

Since each noise is randomly chosen from a normal distribution $N(0, \sigma^2)$, the distribution of x is a normal distribution $N(0, m\sigma^2)$. By the Chebyshev inequality, we have

$$\Pr[\omega \leq \epsilon] = \Pr[|x|/\xi \leq \epsilon] = \Pr[|x - 0| \leq \xi\epsilon] \geq 1 - \frac{m\sigma^2}{(\xi\epsilon)^2} = 1 - \frac{\sigma^2}{m\hat{d}^2\epsilon^2} \qquad (3)$$

Let $\delta = \frac{\sigma^2}{m\hat{d}^2\epsilon^2}$. Equation (3) shows that when σ is sufficiently small and m is sufficiently large, LMC obtains a good approximate with high probability $1 - \delta$. We set $\hat{d}$ to be 105 (an average meter reading in Wh per 5 minutes) according to the statistics of U.S. Energy Information Administration. We fix $\delta = 1\%$ and present values of m and σ for achieving $\epsilon = 10\%$, $\epsilon = 7\%$, and $\epsilon = 5\%$ in Fig. 7. When $m = 600$, σ is about 25, 18 and 12, respectively. The parameter σ is a tradeoff between LMC correctness and LMC privacy requirement. Here we obtain that a better approximate needs a smaller σ. We will see that LMC privacy requirement needs a larger σ in the subsection 5.2.

5 Privacy Requirements and Analysis

We formally define ESP and LMC privacy requirements and show that our system meets the requirements. We also show that meter readings are securely stored.

5.1 ESP Privacy Requirement and Analysis

ESP privacy requirement is that ESP cannot get individual meter readings of a household, where ESP gets sums of meter readings in time periods and accesses encrypted meter readings. We capture the ESP privacy requirement in a security game $\mathcal{G}$, where the power of ESP is enlarged to adaptively decide meter readings for non-challenge time periods. Even having the ability of adaptively setting meter readings and observing resulting encrypted meter readings, ESP

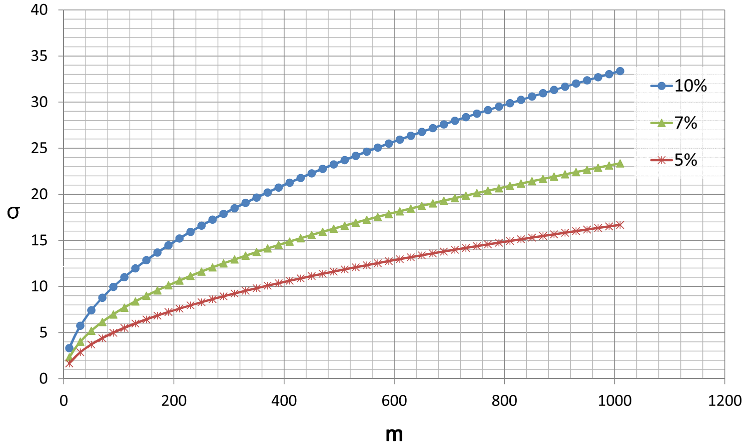


Fig. 7. Values of m and σ for achieving the error ratio ω less than 10%, 7%, and 5%, respectively, where $\hat{d} = 105$ and $\delta = 1\%$

still cannot distinguish meter readings from two possible sets of meter readings in the game $\mathcal{G}$. The security game $\mathcal{G}$ is described in the following.

The challenger $\mathcal{C}$ represents the metering system of a household $\mathcal{H}_i$ and the adversary $\mathcal{A}$ represents the honest-but-curious ESP. A query phase proceeds at beginning. $\mathcal{A}$ adaptively chooses meter readings $d_{i,j}$ at time units t_j from $j = 1$ and $\mathcal{C}$ returns the encrypted meter readings $c_{i,j}$ back to $\mathcal{A}$. $\mathcal{A}$ then decides to enter the challenge phase at t_{j_1}. This phase simulates that $\mathcal{A}$ reveals meter readings and their decryption information at time units earlier than t_{j_1}. In the challenge phase, $\mathcal{A}$ chooses a time period from the time unit t_{j_1} to a later time unit t_{j_2}, where $j_2 - j_1 = al$ for a positive integer a, and two challenge sets D_0, D_1 of meter readings for time units between t_{j_1} and t_{j_2}, where $D_v = \{d^v_{i,j_1}, d^v_{i,j_1+1}, \cdots, d^v_{i,j_2}\}$ for $v \in \{0, 1\}$ and $\sum_{s=j_1}^{j_2} d^0_{i,s} = \sum_{s=j_1}^{j_2} d^1_{i,s}$. $\mathcal{A}$ sends D_0 and D_1 to $\mathcal{C}$. $\mathcal{C}$ throws a random coin b and encrypts meter reading $d^b_{i,s}$ in D_b as $c_{i,s}$ for $s \in [j_1, j_2]$. Let $C = \{c_{i,j_1}, c_{i,j_1+1}, \cdots, c_{i,j_2}\}$. After getting encrypted meter readings C, $\mathcal{A}$ enters the second query phase. Again, $\mathcal{A}$ adaptively chooses a meter reading $d_{i,j'}$ for arbitrary time unit $t_{j'}$ where $j' > j_2$ and $\mathcal{C}$ adaptively returns the encrypted meter reading $c_{1,j'}$ back to $\mathcal{A}$. $\mathcal{A}$ then outputs a guessing b' for b.

If $b' = b$, $\mathcal{A}$ wins the game $\mathcal{G}$. That is, $\mathcal{A}$ successfully distinguishes which set D_b is encrypted. The advantage of an adversary is defined as $|\Pr[b' = b] - 1/2|$.

Definition 1. *A smart metering system satisfies ESP privacy requirement if for any probabilistic polynomial time algorithm $\mathcal{A}$ and a negligible function ε, $|\Pr_{\mathcal{A}}[b' = b] - 1/2| < \varepsilon$.*

A similar game is defined in [17], where the adversary needs to choose challenge sets D_0 and D_1 at the very beginning. It only captures a snapshot of meter readings at a time unit. In our security game, the queries from the adversary is adaptive. As a result, our security game models a stronger security requirement.

ESP Privacy Analysis. We first rephrase the description of the pseudorandom number generator in our smart metering system in Definition 2. Consider a set E_0 of a polynomial number $f(\tau)$ of elements randomly and uniformly chosen from $\mathcal{Z}_p$ and a set E_1 of the same number $f(\tau)$ of elements generated by g.

Definition 2. *Let $\tilde{b} \in \{0,1\}$ be a random coin. A function $g : \{0,1\}^\tau \times \{0,1\}^\lambda \to \mathcal{Z}_p$ is a pseudorandom number generator if given a set $E_{\tilde{b}}$ of elements, no probabilistic polynomial time algorithm guesses $\tilde{b}$ with an advantage more than ε', where ε' is a negligible function in τ.*

Theorem 1 states that our system satisfies ESP privacy requirement.

Theorem 1. *Let g be a pseudorandom number generator. Our smart metering system satisfies ESP privacy requirement, where $\varepsilon \leq 2\varepsilon'$.*

Proof. We prove by contradiction. Assume that an adversary $\mathcal{A}$ wins the game with an advantage at least $2\varepsilon'$. We construct an algorithm $\mathcal{S}$ such that given $E_{\tilde{b}}$, where $\tilde{b} = 0$ and $\tilde{b} = 1$ with equal probabilities, $\mathcal{S}$ guesses $\tilde{b}$ with an advantage more than ε' by using $\mathcal{A}$ as follows.

$\mathcal{S}$ acts as $\mathcal{C}$ and interacts with $\mathcal{A}$ in the security game. $\mathcal{S}$ embeds elements in $E_{\tilde{b}}$ as random numbers $\{r_{i,1}, r_{i,2}, \cdots, r_{i,l-1}, R_{i,1}, R_{i,2}, \cdots, R_{i,f(\tau)-l+1}\}$. For queries $d_{i,j}$ from $\mathcal{A}$, $\mathcal{S}$ returns $c_{i,j}$, where $c_{i,j} = (d_{i,j} + r_{i,j}) \mod p$. For $j \geq l-1$, $\mathcal{S}$ computes $r_{i,j}$ as $(R_{j-l+2} - \sum_{k=j-l+2}^{j-1} r_{i,k}) \mod p$. For challenges D_0 and D_1 from $\mathcal{A}$, $\mathcal{S}$ chooses D_b, which is either D_0 or D_1 with equal probabilities, and computes $C = \{c_{i,s} | c_{i,s} = (d_{i,s}^b + r_{i,s}) \mod p, s \in [j_1, j_2]\}$. After $\mathcal{S}$ sends C to $\mathcal{A}$, again, $\mathcal{S}$ answers queries from $\mathcal{A}$. Finally, if $\mathcal{A}$ successfully guesses b' for b, i.e., $b' = b$, $\mathcal{S}$ outputs 1. Otherwise, $\mathcal{S}$ outputs 0.

When $\tilde{b} = 1$ (E_1 contains pseudorandom numbers), the simulated environment is identical to our system. Thus, $\mathcal{A}$ outputs b' , where $b' = b$, with a probability at least $1/2 + 2\varepsilon'$ by our assumption. When $\tilde{b} = 0$ (E_0 contains truly random numbers), for each possible set D_b, there exists a unique set of values $r_{i,j_1}, r_{i,j_1+1}, \cdots, r_{i,j_2}$ satisfying that $c_{i,s} = (d_{i,s}^b + r_{i,s}) \mod p$ for $s \in [j_1, j_2]$. The distributions of C conditioned on D_0 and D_1 are identical. Thus, $\mathcal{A}$ has no advantage. Therefore, $\mathcal{A}$ correctly guesses b with probability $1/2$. As a result, $\mathcal{S}$ outputs 1 with probability $1/2$. $\mathcal{S}$ guesses $\tilde{b}$ with an advantage at least ε':

$$\begin{aligned}
&\Pr[\mathcal{S} \text{ correctly guesses } \tilde{b}] \\
&= \Pr[\mathcal{S} \text{ outputs } 0 | \tilde{b} = 0] \Pr[\tilde{b} = 0] + \Pr[\mathcal{S} \text{ outputs } 1 | \tilde{b} = 1] \Pr[\tilde{b} = 1] \\
&= \Pr[\mathcal{A} \text{ outputs } b', b' \neq b | \tilde{b} = 0] \Pr[\tilde{b} = 0] \\
&\quad + \Pr[\mathcal{A} \text{ outputs } b', b' = b | \tilde{b} = 1] \Pr[\tilde{b} = 1] \\
&\geq (\frac{1}{2})\frac{1}{2} + (\frac{1}{2} + 2\varepsilon')\frac{1}{2} \\
&= \frac{1}{2} + \varepsilon'
\end{aligned}$$

It contradicts with the assumption. □

5.2 LMC Privacy Requirement and Analysis

LMC Privacy Requirement. For LMC privacy requirement, we require that LMC only gets an approximate value $\tilde{d}_{i,j}$ for $d_{i,j}$ with a non-negligible probability.

Definition 3. *A metering smart system satisfies LMC privacy requirement if LMC guesses value $\tilde{d}_{i,j}$ for $d_{i,j}$ with $\Pr[\tilde{d}_{i,j} \neq d_{i,j}] \geq \eta$ for some significant probability η.*

LMC privacy requirement is slightly weak. Nevertheless, it is practical enough for smart grids. In smart grid deployments, the load monitoring system is often bundled with the grid operator. The grid operator can physically measure the power consumption at a power substation. By cooperating with the grid operator, LMC can get individual meter readings. Our LMC privacy requirement guarantees that when LMC does not get help from the grid operator, LMC cannot get exactly individual meter readings with a significant probability.

LMC Privacy Analysis.

Theorem 2. *Let a noise be the floor of a randomly chosen number from the normal distribution $N(0, \sigma^2)$ in our system. Our smart metering system satisfies LMC privacy requirement, where $\delta = 1/2 - 1/(4\pi\sigma^2)$*

Proof. We analyze $\Pr[\tilde{d}_{i,j} \neq d_{i,j}]$. The event of $\tilde{d}_{i,j} \neq d_{i,j}$ implies that the noise $n_{i,j}$ is not 0. Since the noise $n_{i,j}$ is the floor of a randomly chosen value n from $N(0, \sigma^2)$, the event $n_{i,j} = 0$ implies that $0 \leq n < 1$. Since $\Pr[n_{i,j} \neq 0]$ and $\Pr[n < 0] = (1 - \Pr[n = 0])/2$,

$$\Pr[n_{i,j} \neq 0] > \frac{1}{2} - \frac{1}{4\pi\sigma^2}$$

Thus, for a meter reading $d_{i,j}$ and a noised meter reading $\tilde{d}_{i,j}$, we have $\eta = \frac{1}{2} - \frac{1}{4\pi\sigma^2}$, which is significant for properly chosen σ. It concludes the proof for the LMC privacy requirement. □

When σ is larger, LMC has less probability to get a correct meter reading. Nevertheless, when σ is small, LMC has a better guess $\tilde{\xi}$ for ξ. Based on the previous chosen condition of $\epsilon = 10\%$ and $\delta = 10\%$ for LMC correctness, we consider $\sigma = 25$ and $m = 600$ for achieving that $\Pr[\omega \leq 10\%] \geq 99\%$. Under this setting, we have $\Pr[n_{i,j} \neq 0] > 0.4998$. Similarly, when $\epsilon = 7\%$ ($\sigma = 18$) and $\epsilon = 5\%$ ($\sigma = 12$), we have $\Pr[n_{i,j} \neq 0] > 0.4997$ and $\Pr[n_{i,j} \neq 0] > 0.4994$, respectively.

5.3 Storage Security

We show that meter readings are computationally securely stored in the storage system. Note that ESP has more information than the storage system does and our smart metering system satisfies the ESP privacy requirement.

Table 1. Summary of performance analysis

Actor	Storage system	Meter				
Subject	Storage	Computation			Communication	
		for storing	for ESP	for LMC	for ESP	for LMC
Result	9MB	$7(l+1)$ms	$14a$ms	14ms	64 bits	64 bits

We define the security requirement of storage in a security game $\mathcal{G}'$. The security game $\mathcal{G}'$ is the same as $\mathcal{G}$ except that the attacker $\mathcal{A}$ in $\mathcal{G}'$ represents the storage system. Thus, the security game $\mathcal{G}'$ captures that the storage system colludes with ESP. The storage security requirement is then defined:

Definition 4. *A metering system satisfies secure storage requirement if for any probabilistic polynomial time algorithm $\mathcal{A}$ and a negligible function ε, $|\Pr_{\mathcal{A}}[b' = b] - 1/2| < \varepsilon$.*

Theorem 3 states that our smart metering system satisfies secure storage requirement.

Theorem 3. *Let g be a pseudorandom number generator. Our smart metering system satisfies secure storage requirement, where $\varepsilon \leq 2\varepsilon'$.*

Proof. Since the proof is the same as the proof of Theorem 1, here we refer readers to the proof of Theorem 1. □

6 Performance Analysis

We use the previous setting of $\lceil \log_2 p \rceil = 64$ and set a time unit as 5 minutes. We evaluate the storage cost, computation cost, and communication cost in the following. Table 1 gives a summary.

Storage cost. Inside each meter $\mathcal{M}_i$, l pseudorandom numbers $r_{i,z-l+1}, r_{i,z-l+2}, \cdots, r_{i,z}$ are stored. The total storage size is $l\lceil \log_2 p \rceil$, i.e. $8l$ bytes.

For the storage system, each household uses $\lceil \log_2 p \rceil$ bits per time unit. Let a time unit be 5 minutes. The total storage size for meter readings of a household over 10 years is about 9 MB.

Computation cost. Computation operations of ESP and LMC are modular additions, which are efficient in modern computers. We focus on the computation cost of a meter. For a meter $\mathcal{M}_i$, to store a meter reading $d_{i,j}$, one pseudorandom number $R_{i,j}$ is generated by the TPM and l modular additions are performed by the driver of $\mathcal{M}_i$. To reply a query of al continuous time units $t_\beta, t_{\beta+1}, \cdots, t_{\beta+al-1}$ from ESP, a pseudorandom numbers $R_{i,\beta+kl}$ for $0 \leq k \leq a-1$ are generated and a modular additions are performed by the driver of $\mathcal{M}_i$. To reply a query of a recent time unit t_j from LMC, the driver of $\mathcal{M}_i$ generates a random noise $n_{i,j}$ and performs a modular addition to compute the noised meter reading $\tilde{d_{i,j}}$.

A recent commercial TPM chip[2] consists of a cryptographic accelerator capable of computing a 1024-bit RSA signature in 100 ms. Since generating a 1024-bit random number by using the pseudorandom number generator is no slower than the task of generating a 1024-bit RSA signature, each 64-bit random number can be generated in less than 7 ms. Similarly, we assume that a modular addition over $\mathscr{Z}_p$ can be done in less than 7 ms. Thus, the smart meter can store a meter reading in less than a time unit (5 minutes) when $l < 42856$ and reply a query in less than a time unit when $a < 21428$. The numerical results show that the computation of our system is well supported by current hardware technologies.

The *communication cost* between a meter and ESP or LMC for a query is $\lceil \log_2 p \rceil$ bits. That is, for a query from ESP or LMC, a meter transmits a 64-bit sum B of random numbers or a 64-bit noised random number $\tilde{r}_{i,j}$.

7 Conclusion and Future Works

We proposed a smart metering system that simultaneously supports the billing and load monitoring applications in a privacy preserving manner. ESP can only query for consumption of a household over a time period. LMC can only query an approximate consumption in an area at a recent time unit. Our construction is based on the layered meter model and uses the pseudorandom number generator in the TPM. According to our performance analysis, based on current TPM technologies, our construction is a practical and feasible solution to privacy preserving smart metering systems.

In addition to the billing and load monitoring applications, fine-grained consumption data contribute to other intelligent smart grid applications, such as demand prediction and power distribution planning. It is interesting to design a secure smart metering system that supports more applications.

Acknowledgement. We thank anonymous reviewers for their comments. The research was supported in part by projects ICTL-101-Q707, ATU-101-W958, NSC 101-2218-E-009-003-, NSC 98-2221-E-009-068-MY3, and NSC 100-2218-E-009-006.

References

1. Ontario energy board, smart meters and time of use (tou) prices, webpage available at `http://www.ontarioenergyboard.ca/OEB/Consumers/Electricity/Smart+Meters`
2. Smart grid cyber security strategy and requirements. National Institute of Standards and Technology (U.S.) Interagency Reports, DRAFT-NISTIR-7628 (2009)
3. Ács, G., Castelluccia, C.: I Have a DREAM (DiffeRentially privatE smArt Metering). In: Filler, T., Pevný, T., Craver, S., Ker, A. (eds.) IH 2011. LNCS, vol. 6958, pp. 118–132. Springer, Heidelberg (2011)
4. Bohli, J.M., Sorge, C., Ugus, O.: A privacy model for smart metering. In: Proceedings of the 1st IEEE International Workshop on Smart Grid Communications (2010)

[2] Atmel AT97SC3203S.

5. Chan, T.H.H., Shi, E., Song, D.: Privacy preserving stream aggregation with fault tolerance. In: Proceedings of the 16th International Conference on Financial Cryptography and Data Security - FC 2012. LNCS. Springer (2012)
6. Garcia, F.D., Jacobs, B.: Privacy-Friendly Energy-Metering via Homomorphic Encryption. In: Cuellar, J., Lopez, J., Barthe, G., Pretschner, A. (eds.) STM 2010. LNCS, vol. 6710, pp. 226–238. Springer, Heidelberg (2011)
7. Hart, G.W.: Nonintrusive appliance load monitoring. In: Proceedings of the IEEE, pp. 1870–1891. IEEE Press (1992)
8. Inc., A.P.L.S.: Aclara ami industry glossary (2008)
9. Jawurek, M., Johns, M., Kerschbaum, F.: Plug-in privacy for smart metering billing. Computing Research Repository - CoRR (2010)
10. Kursawe, K., Danezis, G., Kohlweiss, M.: Privacy-Friendly Aggregation for the Smart-Grid. In: Fischer-Hübner, S., Hopper, N. (eds.) PETS 2011. LNCS, vol. 6794, pp. 175–191. Springer, Heidelberg (2011)
11. Laughman, C., Lee, K., Cox, R., Shaw, S., Leeb, S., Norford, L., Armstrong, P.: Power signature analysis. IEEE Power and Energy Magazine 1, 56–63 (2003)
12. LeMay, M., Gross, G., Gunter, C.A., Garg, S.: Unified architecture for large-scale attested metering. In: Proceedings of the 40th Hawaii International International Conference on Systems Science - HICSS 2007, p. 115. IEEE Computer Society (2007)
13. Metke, A.R., Ekl, R.L.: Security technology for smart grid networks. IEEE Transactions on Smart Grid 1(1), 99–107 (2010)
14. Molina-Markham, A., Shenoy, P., Fu, K., Cecchet, E., Irwin, D.: Private memoirs of a smart meter. In: Proceedings of the 2nd ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Building - BuildSys 2010, pp. 61–66. ACM (2010)
15. Petrlic, R.: A privacy preserving concept for smart grids. In: Sicherheit in vernetzten Systemen: 18. DFN Workshop, pp. B1–B14. Books on Demand GmbH (2010)
16. Rial, A., Danezis, G.: Privacy-preserving smart metering. Microsoft Technical Report, MSR-TR-2010-150 (2010)
17. Shi, E., Chan, T.H.H., Rieffel, E., Chow, R., Song, D.: Privacy-preserving aggregation of time-series data. In: Proceedings of the 18th Annual Network and Distributed System Security - NDSS 2011. The Internet Society (2011)