A Fault-tolerant and Efficient Scheme for Data Aggregation Over Groups in the Smart Grid

Fabian Knirsch, Dominik Engel
Josef Ressel Center for
User-Centric Smart Grid Privacy, Security and Control
Salzburg University of Applied Sciences
Urstein Sued 1, A–5412 Puch/Salzburg, Austria
Email: {fabian.knirsch, dominik.engel}@en-trust.at

Zekeriya Erkin
Cyber Security Group
Department of Intelligent Systems
Delft University of Technology
Mekelweg 4, NL–2628 CD Delft, The Netherlands
Email: z.erkin@tudelft.nl

***Abstract*—Aggregating data in the smart grid is an important issue for obtaining the total consumption of a group of households. In order to aggregate data in a privacy preserving manner, it has to be assured that individual contributions are untraceable and only the sum is visible to an aggregator. For billing, network security and statistical analysis data from different types of customers (e.g., industrial, residential) has to be aggregated separately. This paper presents a fault-tolerant and efficient scheme for aggregating data over different groups while preserving the privacy of the households. We propose to build on the Chinese Remainder Theorem for aggregating over groups and on a fault-tolerant and tree-based approach for increasing efficiency. The resulting protocol is evaluated in terms of privacy, complexity and real-world applicability, such as dynamic joins and leaves.**

I. Introduction

The term *smart grid* refers to novel energy grids where both, energy and information are exchanged between customers and utility providers. This information is used for, e.g., billing, demand response or network monitoring. Data from smart meters, intelligent energy meters installed in households, is collected and processed. This often fine grained meter data poses privacy risks to customers and therefore, methods for the secure and privacy preserving aggregation of data have been proposed [1]–[5].

In order to protect customer privacy, the sum over a set of smart meters for one point in time (spatial aggregation) or the sum over time for one smart meter (temporal aggregation) is desired by the aggregator, i.e., the aggregator does not receive fine-grained meter data, but only the total consumption. In addition, aggregation protocols should be capable of dealing with smart meter outages or communication failures, dynamic joins and leaves and provide low complexity in terms of communication overhead.

In this paper we present a novel protocol for the fault-tolerant aggregation over groups in networks with a star topology that also supports dynamic joins and leaves. The protocol is designed for aggregator obliviousness and we show that it is privacy-preserving under a honest-but-curious adversarial model. The approach is based on the Chinese Remainder Theorem (CRT) for calculating the total consumption over a neighborhood in a smart grid as well as the total consumption of subgroups. This scheme draws on secret sharing and the Paillier cryptosystem and reduces the complexity in terms of communication overhead from $\mathcal{O}(N^2)$ to $\mathcal{O}(N^{1+\epsilon})$ while strictly maintaining the star connected network.

Using the CRT for data aggregation with groups is originally proposed by Erkin in [6]. A highly efficient scheme for fault-tolerant aggregation in star networks is proposed by Rane et al. [7]. The approach presented here achieves both, the fault-tolerant aggregation in star networks and aggregation over groups at negligible overhead. The proposed protocol allows for the fault-tolerant aggregation over groups while maintaining a high efficiency of $\mathcal{O}(N^{1+\epsilon})$ in star networks.

The usage of the proposed protocol is not limited to smart metering or smart grid applications, but has manifold applications in domains such as machine learning (e.g., the privacy-preserving clustering of data), data analysis or any application that requires an efficient scheme for data aggregation over groups.

The rest of the paper is structured as follows: Section II introduces the preliminaries such as notation, security and privacy assumptions and cryptographic primitives for this paper. Section III gives an overview of related work in the field of privacy-preserving aggregation in the smart grid. Section IV describes the basic aggregation protocol and Section V discusses the proposed approach with respect to privacy and security, efficiency and real-world applicability. Section VI concludes the paper.

II. Preliminaries

a) Notation: For this paper it is assumed that all smart meters are connected in a star topology with the aggregator $\mathcal{A}$. The set of smart meters is denoted as SM_i with $1 \leq i \leq N$. Smart meters will be grouped in K groups, denoted as G_k with $1 \leq k \leq K$. Without loss of generality and for the sake of simplicity we assume that each group consists of an equal number of $M = \frac{N}{K}$ smart meters. Each smart meter reports a measurement $m_{i,t} \in [0, m_{\max})$ at time t.

WIFS'2017, December, 4-7, 2017, Rennes, France. 978-1-5090-6769-5/17/$31.00

The smart meters report to an aggregator $\mathcal{A}$ that in turn forwards the aggregate to a utility provider $\mathcal{UP}$. The utility provider is capable of computing the total consumption and the consumption of subgroups. $\mathcal{A}$ is only processing the encrypted messages and therefore, does not learn anything about the measurements.

b) Cryptographic Primitives: Let $E_{\mathrm{pk}}(\cdot)$ and $D_{\mathrm{sk}}(\cdot)$ denote the encryption and decryption function, respectively of a semantically secure asymmetric key additively homomorphic cryptosystem, e.g., the Paillier cryptosystem [8], such that $D_{\mathrm{sk}}(E_{\mathrm{pk}}(m_1) \cdot E_{\mathrm{pk}}(m_2)) = m_1 + m_2$ with pk as the public key and sk as the secret key. Note that the Paillier cryptosystem has a plaintext message space of $\mathbb{Z}_n^*$ that expands to $\mathbb{Z}_{n^2}^*$ for the ciphertext. Furthermore, a threshold scheme is needed in order to share a value such that at least $d+1$ participants are needed for reconstructing the secret. Shamir's Secret Sharing [9] is used as follows: Let $p(x) = p_0 + p_1 x + p_2 x^2 + \cdots + p_d x^d$ be a polynomial of degree d with p_0 as the secret. At least $d+1$ participants are required for reconstructing the secret. By the holder of the secret the polynomial is evaluated at $i = 1, 2, \ldots, M$ points with $M \geq d+1$ and $x_i \neq 0$ which yields $q_i = p(x_i)$. These points are shared with the other participants. For reconstruction, polynomial interpolation is used

$$p_i'(0) = \sum_{i=1}^{M} \left(\prod_{\substack{j=1 \\ i \neq j}}^{M} \frac{-x_j}{x_i - x_j} \right) q_i, \tag{1}$$

and therefore, allows to reconstruct the secret by

$$p(0) = \sum_i p_i'(0). \tag{2}$$

c) Chinese Remainder Theorem: Given a system of j congruences in the form $x \equiv a_i \mod p_i$, $1 \leq i \leq j$, with $p_i \in \mathbb{Z}$ and $a_i \in \mathbb{Z}$ and where all p_i are pairwise relatively prime. This system then has a unique solution for x modulo $P = \prod_i p_i$ given by [10]

$$x = \sum_i a_i \frac{P}{p_i} \left(\frac{p_i}{P} \mod p_i \right) \mod P. \tag{3}$$

d) Security And Privacy Assumptions: The purpose of the protocol is to aggregate over groups, such that (i) the aggregator only receives the sum of the predefined subgroups and the total sum (aggregator obliviousness); and (ii) any other participants do not learn about single measurements, the partial sums or the total sum. We assume the following semi-honest, also known as honest-but-curious, adversarial model: $\mathcal{UP}$, $\mathcal{A}$ and the smart meters SM_i follow the protocol faithfully, but attempt to learn additional information. All communication links are encrypted as part of the protocol. Autenthication can be used, but is not within the scope here.

III. Related Work

The problem of privacy preserving data aggregation in the smart grid is discussed in literature by e.g., [1], [2], [4], [11], [12], [13], [14], and [17]. Approaches draw either on differential privacy, homomorphic encryption, masking or a combination of them. The aim of these protocols is to spatially and/or temporally aggregate data over a number of smart meters. The aggregation of such data is an important basis for many smart grid use cases such as (i) billing; (ii) network monitoring; and (iii) electric vehicle charging. For billing, for a single household or smart meter a temporal aggregation over time is desired. Network monitoring use cases require fine grained temporal resolutions, but at little or scalable spatial resolution. This use case requires a flexible and fault-tolerant protocol that allows to aggregate over groups while preserving an individual customer's privacy. For electric vehicle charging data is collected for battery health monitoring and billing [15]. Note that the application of protocols for privacy-preserving aggregation has also a wide range of possible application beyond the aforementioned smart grid use cases.

IV. Aggregation Protocol

In this section, we present our novel, efficient and fault-tolerant aggregation scheme that provides the consumption for different groups as well as the total consumption. Our work was inspired by the protocols of Erkin [6] and Rane et al. [7], however, we introduce a number of changes in the setting and the construction of the protocol to achieve both, the ability to aggregate over groups while maintaining fault-tolerance and high efficiency.

(1) This work proposes to use a secret sharing based thresholding scheme that requires $d+1$ smart meters per cohort to participate in decryption, which is similar to Rane et al.'s approach. In contrast to Erkin's approach, the secret value α, which is used by a pair of smart meters to force the aggregator combining their readings, can therefore be omitted. (2) In this protocol there is a single entity $\mathcal{UP}$ that will be able to learn the sum of predefined subgroups based on the CRT approach at the end of the protocol. Here, the utility provider will act as the aggregator as well. The obfuscator nodes prevent $\mathcal{A}$ and $\mathcal{UP}$, respectively, from unintendendly learning partial sums anyway, which is similar to Rane et al.'s approach. This is in contrast to Erkin's setup, where $\mathcal{A}$ and $\mathcal{UP}$ are required to be distinct entities for preserving privacy. (3) For the proposed protocol it is necessary to establish different keys for each SM_i, which is similar to the protocol proposed by Rane et al., but in contrast to Erkin's protocol where a single public key is used for all SM_i. Due to the obfuscator nodes, additional encryption is not necessary in this protocol.

In the following, only $\mathcal{A}$ is used as an aggregator. $\mathcal{A}$ may directly correspond to a utility provider $\mathcal{UP}$. This protocol uses two distinct types of groups. Let therefore C_ℓ with $1 \leq l \leq L$ denote the cohorts (i.e., the groups for efficiently aggregate over all smart meters) and let G_k with $1 \leq k \leq K$ denote the aggregation groups (i.e., the groups for which the aggregator also receives the partial consumption). A particular smart meter SM_i can arbitrarily be assigned to these groups. Therefore, the set of smart meters in any G_k may not at all, partially or fully overlap with any cohort C_ℓ, however, a smart meter may only be assigned to one group, such that all groups are disjoint.

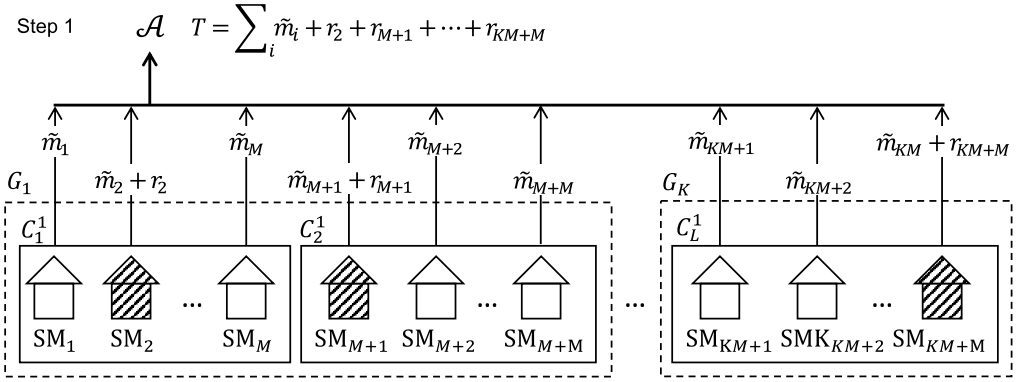

Fig. 1. Basic setup with aggregator $\mathcal{A}$ and smart meters SM_i grouped in L cohorts $C_1, C_2, \ldots, C_L$ with M smart meters in each cohort and K aggregation groups $G_1, G_2, \ldots, G_K$. In this exemplary setup, three cohorts and two aggregation groups are shown. One group spans over two cohorts for illustrative purposes. In step 1, all SM_i apply the CRT approach to their measurement value m_i and execute the secret sharing protocol with their value $\tilde{m}_i$ with $\mathcal{A}$ and the other smart meters in the cohort. The randomly chosen obfuscator nodes (shaded houses) additionally add a random number r_i.

In the following, the proposed protocol is described in detail.

A. Initialization

The total number of N smart meters is grouped into K groups $G_1, G_2, \ldots, G_K$ with M smart meters in each group. Each smart meter is assigned to exactly one group. These groups are for calculating partial summations and will in practice correspond to, e.g., a group of industry customers, a group of residential customers, etc. Further, the aggregator defines a set of cohorts $C_1, C_2, \ldots, C_L$ for efficient aggregation and randomly chooses an obfuscator node in each cohort. Generally, the number of cohorts can be determined by $N = M^L$. For simplicity, but without loss of generality, we define an equal number of cohorts and groups (i.e., $K = L$) and we assume the number of smart meters in each cohort and group to be $M = \frac{N}{K}$. Note that this simplification is for illustrative purposes only and neither restricts our protocol nor does it impose any privacy issues. An exemplary setup is shown in Figure 1. $\mathcal{A}$ establishes an asymmetric key pair $(\mathrm{pk}_i, \mathrm{sk}_i)$ with each of the smart meters and using an additively homomorphic cryptosystem. $\mathcal{A}$ also creates a prime number $p_k > Mm_{\max}$ for each group $G_1, G_2, \ldots, G_K$ and a large prime $\beta > Nm_{\max}$. Further,

$$P = \prod_{k=1}^{K} p_k, \tag{4}$$

is computed and the triple (p_k, β, P) is sent to all smart meters SM_i in their corresponding group G_k.

B. One Round of Reading

For one round of reading the smart meters report their measurement $m_{i,t}$ to the aggregator, so that $\mathcal{UP}$ can compute $T_{k,t} = \sum_{i \in G_k} m_{i,t}$ and $T_t = \sum_{i=1}^{K} T_{k,t}$, but learns nothing beyond that function. The protocol is run recursively per cohort and obfuscators become the input nodes in the next round.

Therefore, each SM_i, for $1 \leq i \leq N$ and $\mathcal{A}$, respectively perform the following calculations per cohort and level:

(1) SM_i computes $\tilde{m}$ by using the group's prime number p_k and the product of all group primes P as

$$\tilde{m}_{i,t} = m_{i,t} \frac{P}{p_k} \left(\frac{p_k}{P} \mod p_k \right) \mod P. \tag{5}$$

If the particular smart meter has been chosen as the obfuscator node it also adds a random number r to its value $\tilde{m}_{i,t}$ which results in a contribution of $\tilde{m}_{i,t} + r$. In the next round, these obfuscator nodes become the input nodes. The node then contributes $-r$ (see Figure 2).

(2) SM_i creates a polynomial of a degree $d < M$, i.e., a degree less than the number of smart meters in the group. This polynomial is then given by

$$p^{(i)}(x) = \tilde{m}_{i,t} + p_1^{(i)} x + p_2^{(i)} + \cdots + p_d^{(i)} x^d \mod \beta. \tag{6}$$

Note that evaluating the polynomial at 0 does not yield the actual measurement, but the term computed in the first step.

(3) The polynomial is now evaluated by SM_i at M distinct points, e.g., $j = 1, 2, \ldots M$, which corresponds to the number of smart meters in the cohort. These points are encrypted using the public key pk_i of the smart meter SM_i and sent to $\mathcal{A}$.

(4) $\mathcal{A}$ computes the encrypted share A_j by

$$A_j = E_{\mathrm{pk}_j}(r_j) \prod_{i=1}^{N} E_{\mathrm{pk}_j}(p^{(i)}(j)), \tag{7}$$

A_j is sent to the corresponding smart meter SM_j for decryption.

(5) After decrypting A_j, the value is sent back to $\mathcal{A}$ allowing it to subtract r_j and to calculate

$$q(j) = \sum_{i=1}^{M} p^{(i)}(j) \mod \beta, \tag{8}$$

and further

$$q(x) = q_1 x + q_2 x^2 + \cdots + q_d x^d + \sum_{i=1}^{M} \tilde{m}_{i,t} \mod \beta. \tag{9}$$

(6) In order to retrieve the summation term for this cohort,

at least $d+1$ participants need to be online allowing $\mathcal{A}$ to evaluate the polynomial at $d+1$ points. After performing a polynomial interpolation this yields q_0, which is the sum of the $\tilde{m}_{i,t}$ for the cohort. In the next level, the obfucator nodes serve as the contributors. The obfuscator nodes therefore subtract the random value r from the partial aggregate and the protocol is executed again on this level. In the last level, no obfuscators are needed.

C. Computing the Aggregate

Once the aggregate of all cohorts has been computed this results in a final sum T_t over all cohorts and groups. For retrieving both, the actual consumption of the group and the total consumption, $\mathcal{A}$ calculates

$$T_{k,t} = T_t \mod p_k, \tag{10}$$

for the group consumption of group k and

$$T_t = \sum_k T_{k,t}, \tag{11}$$

for the total consumption.

In summary, $\mathcal{A}$ learns the consumption of the K groups and the total consumption. Privacy is preserved as neither the smart meters learn anything about the group consumption, the total consumption or a single measurement from another smart meter, nor does $\mathcal{A}$ learn anything beyond the totals. The groups need to be defined at initialization of the aggregation protocol. The cohorts are for efficient aggregation only, and may be reorganized after each round of reading. For $\mathcal{A}$, in order to be able to receive an aggregate, according to Rane et al. [7] $N-d-1$ smart meters can go offline after step 3 during one round of reading. The presented protocol therefore supports all three of our initial requirements for (i) being capable of aggregating over multiple groups; (ii) being fault-tolerant; and (iii) efficiently aggregating with a complexity of $\mathcal{O}(N^{1+\epsilon})$ while still maintaining a star topology.

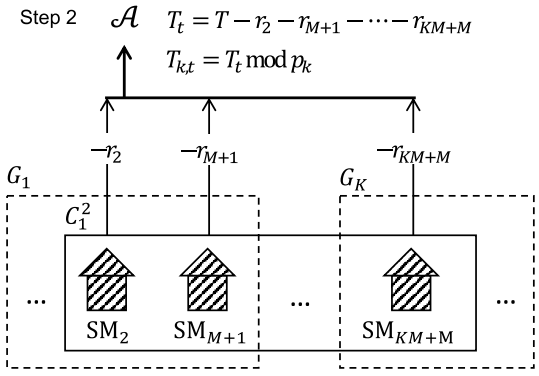

Fig. 2. Step 2 of the basic setup, only the obfuscator nodes act as participants in the secret sharing protocol. They submit their random value used in the previous step as the input value. Once $\mathcal{A}$ received all the values from all cohorts, it can apply the CRT to determine the partial sum $T_{k,t}$ of the previously defined aggregation groups $G_1, G_2, \ldots, G_K$.

V. Evaluation

In this section the proposed protocol is evaluated with respect to its privacy properties, its support for smart meter faults and dynamic joins and leaves and its complexity. Further, the number of groups for different message spaces, given the Paillier cryptosystem, is evaluated and compared to Erkin's approach [6].

A. Privacy Properties

In accordance to our semi-honest or honest-but-curious model, all participants follow the protocol faithfully but attempt to learn additional information. For Erkin's protocol, the security has been proven in [6]. However, the protocol has been modified such that (i) the term α is omitted; and (ii) there is no common public key for all smart meters, i.e., the values $\tilde{m}$ are not encrypted anymore. The term α is originally introduced to force the aggregator to combine at least two previously coupled households, i.e., the aggregator can not just forward a measurement to $\mathcal{UP}$ for decryption. In our modified version of the protocol and in combination with Rane et al.'s approach this term is not required anymore, as smart meters never submit only an encryption of their measurement, but M distinct points of a polynomial. Similarly, the encryption from Erkin's approach can be omitted as the polynomial based secret sharing in combination with the obfuscator nodes do not require additional encryption. In summary, the security and privacy property of our protocol fully relies on Rane et al.'s approach, where Erkin's approach is used for computing subgroups.

For Rane et al.'s protocol, the security has been proven in [7]. The protocol itself is not modified, but instead of sending the actual measurement value m a modified value $\tilde{m}$ is used. This does not hinder the security of the protocol as $\sum_i \tilde{m}_i$ is kept privately instead of $\sum_i m_i$.

In the following, the security and privacy properties for each of the involved participants are investigated in an honest-but-curious model.

1) Smart Meter: During the initialization phase smart meters do not actively participate, but are assigned to cohorts and groups and are provided the parameters for the protocol, i.e., (p_k, β, P). These parameters are publicly known and do not affect the privacy or security of the smart meter.

For one round of reading, each smart meter calculates $\tilde{m}_{i,t}$ (step 1) and computes the polynomial $p^{(i)}(x)$ (step 2). Under the honest-but-curious model, these values will not be forged by the smart meter, but calculated according to the protocol. So far, the private value $m_{i,t}$ has not been released in any kind. In step 3, the polynomial is evaluated at M distinct points. Note that the information theoretical security properties of the underlying Shamir secret sharing scheme hold. Further, these points are encrypted with the public key of the smart meter and sent to the aggregator. The security properties of this step are shown later in this paper.

2) Obfuscator Node: In the course of aggregating the values, smart meters may randomly become obfuscator nodes. Obfuscator nodes add an additive random value to their

measurement in order to prevent the aggregator from learning partial sums (step 1). This addition of a random number can be reduced to a masking scheme as presented in [16].

3) Aggregator: The aggregator $\mathcal{A}$ is involved in this protocol in all phases. First, during initialization $\mathcal{A}$ is responsible for defining groups and determining p_k and β. Second, $\mathcal{A}$ is involved in one round of reading (steps 3-6). Finally, $\mathcal{A}$ calculates the aggregate.

The group setup and the initialization parameters (p_k, β, P) are publicly known and can be verified by all the participants. If $\mathcal{A}$ attempts to choose too small groups for privacy-preserving aggregation (i.e., in the worst and trivial case only one smart meter per group), the smart meters can decline sending values. Similarly, the prime numbers p_k, as well as the corresponding product P can be verified. The same holds for β, which needs to be large enough ($\beta > N m_{\max}$). The latter, however, does not affect the privacy properties of the protocol, but is required for the protocol to work and produce the correct aggregate.

During one round of reading, $\mathcal{A}$ is responsible for relaying and processing the encrypted secret shares sent by the smart meters. Note that these shares stem from Shamir's Secret Sharing and that the operations by $\mathcal{A}$ are performed in the encrypted domain where the Paillier cryptosystem is used and the obfuscator nodes add random values in order to prevent the aggregator from learning partial sums, respectively. The Paillier cryptosystem is proven to be semantically secure if the Decisional Composite Residuosity Assumption holds [8] and Shamir's Secret Sharing is semantically secure.

Finally, the CRT approach has a unique solution for the prime numbers used, which prevents $\mathcal{A}$ from learning any other partial consumption than defined at the beginning of the protocol. After one round of reading $\mathcal{A}$ therefore only learns the total consumption and the consumptions of the subgroups.

4) Summary: In summary, for an honest-but-curious smart meter it is impossible to learn the measurement value of any other smart meter or the sum of measurement values of a group or cohort. This is guaranteed by secret sharing and the obfuscator nodes, where only $\mathcal{A}$ is able to recover the total sum. Even an obfuscator node does not learn anything beyond its own measurement value, since it only acts as a participant in the next level and prevents $\mathcal{A}$ from learning the partial sum. According to Rane et al. [7], in case of a collusion of honest-but-curious smart meters, the protocol is secure if less than d participants collude. Further, Rane et al. point out that choosing the obfuscator at random is crucial, since a collusion of $\mathcal{A}$ and obfuscator node reveals the partial sum to both.

B. Fault Tolerance and Dynamic Joins/Leaves

A crucial aspect of smart meter aggregation protocols is fault-tolerance and the support for dynamic joins and leaves. The protocol must be capable of handling newly added smart meters and it must handle smart meters that suddenly fail. Failure can be due to a smart meter fault or a broken communication link.

A new smart meter joining a group is provided the triple (p_k, β, P) by $\mathcal{A}$ and assigned to a cohort. According to [7], a smart meter can either join the original set of all smart meters or the set of obfuscator nodes at a later stage. Note that this does not require the entire protocol to be reinitialized.

If a smart meter fails during operation or a communication link breaks, the following two conditions may apply: (i) if the respective node is an obfuscator node and it fails after step 3 of the final protocol, the aggregate can still be calculated; and (ii) at most $N - d - 1$ smart meters must be online after step 3 in order to complete the protocol.

C. Complexity

The complexity of the proposed protocol in terms of communication is $\mathcal{O}(N^{1+\epsilon})$. This is in contrast to Erkin's approach which comes at a lower complexity of $\mathcal{O}(N)$, but has the disadvantage of not being fault-tolerant. This increase in complexity buys from the capability of having a threshold scheme based on secret sharing and allowing the failure of smart meters during the protocol while still being capable of computing the aggregate. The latter also applies for the final protocol, since communication occurs in our protocol exactly as originally presented by Rane et al. [7]. Any additional steps for calculating the aggregation over subgroups does not come at additional communication. The complexity is calculated as follows [7]: Given $N = M^L$, with N as the total number of smart meters, M as the number of smart meters in each cohort and L being a positive integer, the complexity can be summarized as $\mathcal{O}(N^{1+\epsilon})$, where $\epsilon = \frac{1}{L} < 1$.

Table I compares the three protocols in terms of the number of encryptions, decryptions, multiplications, exponentiation, and inbound and outbound messages with respect to the total number of smart meters N, the number of groups K and the number of smart meters in each group M.

D. Groups and Message Size

The total number of households supported by the protocol depends on the message space n of the underlying cryptosystem. This issue is originally discussed by Erkin [6]. Table II compares the number of households for Erkin's approach, Rane et al.'s approach and this paper. For a better comparison we follow [6] and assume the number of bits for a single measurement value $m_{i,t}$ to be $\bar{m} = 16$. This value is motivated by typical annual electricity consumption of residential buildings.

For Erkin's approach the value of N is calculated following the method presented in [6] as $N = KM = \frac{n}{\bar{m}}$. Note that all the N smart meters share a public key, regardless of the group they are in, and therefore, the result is independent of the number of groups K. For Rane et al.'s approach, a public-private key pair is established between the aggregator and each smart meter. The aggregator only needs to calculate the sum of the cohort in each step. However, each smart meter needs to evaluate the polynomial at M (i.e., the number of smart meters in the cohort) points. The number of supported households is therefore calculated as follows: Given the condition $\beta > N\bar{m}$, it can be concluded that $\frac{N}{K} = M = \frac{n}{\bar{m}}$. Therefore, the total number of supported smart meters can be increased by increasing the number of groups. In our protocol we draw on the approach from Rane et al. for securely aggregating

TABLE I

COMPARISON OF COMPLEXITY FOR THE NUMBER OF ENCRYPTIONS, DECRYPTIONS, MULTIPLICATIONS, EXPONENTIATION, AND INBOUND AND OUTBOUND MESSAGES. N IS THE TOTAL NUMBER OF SMART METERS, K IS THE NUMBER OF GROUPS AND M IS THE NUMBER OF SMART METERS IN EACH GROUP. THE PROTOCOL BY RANE ET AL. AND THIS WORK DO NOT DISTINGUISH BETWEEN AGGREGATOR AND UTILITY PROVIDER.

	Erkin [6]			Rane et al. [7]		This work	
	SM	$\mathcal{A}$	$\mathcal{UP}$	SM	$\mathcal{A}$ / $\mathcal{UP}$	SM	$\mathcal{A}$ / $\mathcal{UP}$
encryption	$\mathcal{O}(1)$			$\mathcal{O}(M)$	$\mathcal{O}(K)$	$\mathcal{O}(M)$	$\mathcal{O}(K)$
decryption			$\mathcal{O}(1)$	$\mathcal{O}(M)$		$\mathcal{O}(M)$	
multiplication		$\mathcal{O}(N)$			$\mathcal{O}(N)$		$\mathcal{O}(N)$
exponentiation	$\mathcal{O}(1)$		$\mathcal{O}(K)$				
messages inbound	$\mathcal{O}(1)$	$\mathcal{O}(N)$	$\mathcal{O}(N)$	$\mathcal{O}(M)$	$\mathcal{O}(N)$	$\mathcal{O}(M)$	$\mathcal{O}(N)$
messages outbound	$\mathcal{O}(N)$		$\mathcal{O}(1)$	$\mathcal{O}(M)$	$\mathcal{O}(N)$	$\mathcal{O}(M)$	$\mathcal{O}(N)$

TABLE II

NUMBER OF HOUSEHOLDS N FOR A GIVEN MESSAGE SPACE n OF THE UNDERLYING CRYPTOSYSTEM.

n	Erkin [6]	Rane et al. [7]	This work
1024	64	$64 \cdot K$	$64 \cdot K$
1536	96	$96 \cdot K$	$96 \cdot K$
2048	128	$128 \cdot K$	$128 \cdot K$

the measurement values. Therefore, the number of supported households is equivalent to Rane et al. [7].

VI. CONCLUSION

In this paper we present a novel protocol that allows an aggregator to efficiently aggregate over groups of smart meters. We achieved this by using the CRT for defining subgroups and an approach based on Shamir's Secret Sharing and the Paillier cryptosystem for efficient aggregation in a network with star topology. Erkin [6] originally proposed the use of the CRT for retrieving group aggregations and the total aggregation, while Rane [7] et al. proposed a tree-like approach for aggregating over cohorts for an increased efficiency in star networks. The latter approach also supports fault-tolerance. In this paper, we describe a new protocol inspired by those two approaches but do have significant advantages in terms of computational efficiency. Furthermore, we show that our protocol supports group aggregation, a complexity of $\mathcal{O}(N^{1+\epsilon})$, fault-tolerance and dynamic joins and leaves. Future work will focus on adding an additional level of error-resilience that allows to recover the total aggregate even if one of the obfuscator nodes fails and the addition of integrity checks to verify the aggregate.

ACKNOWLEDGMENT

The financial support by the Austrian Federal Ministry of Science, Research and Economy, the Austrian National Foundation for Research, Technology and Development and the Federal State of Salzburg is gratefully acknowledged.

REFERENCES

[1] K. Kursawe, G. Danezis, and M. Kohlweiss, "Privacy-friendly aggregation for the smart grid," in *PETS*, 2011, pp. 175–191.

[2] G. Acs and C. Castelluccia, "I have a DREAM! (DiffeRentially privatE smArt Metering)," in *Proc. Information Hiding Conference*, 2011, pp. 118–132.

[3] E. Shi, R. Chow, T.-h. H. Chan, D. Song, and E. Rieffel, "Privacy-preserving aggregation of time-series data," in *Proc. NDSS Symposium 2011*, 2011.

[4] F. Gomez Marmol, C. Sorge, R. Petrlic, O. Ugus, D. Westhoff, and G. Martinez Perez, "Privacy-enhanced architecture for smart metering," *International Journal of Information Security*, vol. 12, no. 2, pp. 67–82, 2013.

[5] G. Danezis, C. Fournet, M. Kohlweiss, and S. Zanella-Béguelin, "Smart Meter Aggregation via Secret-sharing," in *Proceedings of the First ACM Workshop on Smart Energy Grid Security*, ser. SEGS '13. New York, NY, USA: ACM, 2013, pp. 75–80.

[6] Z. Erkin, "Private Data Aggregation with Groups for Smart Grids in a Dynamic Setting using CRT," in *2015 IEEE International Workshop on Information Forensics and Security (WIFS)*. Rome, Italy: IEEE, 2015.

[7] S. Rane, J. Freudiger, A. E. Brito, and E. Uzun, "Privacy , Efficiency & Fault Tolerance in Aggregate Computations on Massive Star Networks," in *2015 IEEE International Workshop on Information Forensics and Security (WIFS)*. IEEE, 2015.

[8] P. Paillier, "Public-Key Cryptosystems Based on Composite Degree Residuosity Classes," in *Advances in Cryptology — EUROCRYPT '99: International Conference on the Theory and Application of Cryptographic Techniques Prague, Czech Republic, May 2–6, 1999 Proceedings*, ser. Lecture Notes in Computer Science, J. Stern, Ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 1999, vol. 1592, pp. 223–238.

[9] A. Shamir, "How to Share a Secret," *Communications of the ACM (CACM)*, vol. 22, no. 11, pp. 612–613, 1979. [Online]. Available: http://doi.acm.org/10.1145/359168.359176

[10] H. C. van Tilborg and S. Jajodia, Eds., *Encyclopedia of Cryptography and Security*, 2nd ed. Springer, 2011.

[11] V. Rastogi and N. Suman, "Differentially private aggregation of distributed time-series with transformation and encryption." in *Proceedings of the 2010 ACM SIGMOD International Conference on Management of data.*, 2010.

[12] F. Li, B. Luo, and P. Liu, "Secure Information Aggregation for Smart Grids Using Homomorphic Encryption," in *Proceedings of First IEEE International Conference on Smart Grid Communications*, Gaithersburg, Maryland, USA, 2010, pp. 327–332.

[13] G. Danezis, M. Kohlweiss, and A. Rial, *Differentially private billing with rebates*, T. Filler, T. Pevny, S. Craver, and A. Ker, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, vol. 6958 LNCS.

[14] T. H. H. Chan, E. Shi, and D. Song, "Privacy-preserving stream aggregation with fault tolerance," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 7397 LNCS, pp. 200–214, 2012.

[15] L. Langer, F. Skopik, G. Kienesberger, and Q. Li, "Privacy issues of smart e-mobility," in *39th Annual Conference of the IEEE Industrial Electronics Society, IECON 2013*, 2013, pp. 6682–6687.

[16] C. Castelluccia, A. C. Chan, E. Mykletun, and G. Tsudik, "Efficient and Provably Secure Aggregation of Encrypted Data in Wireless Sensor Networks," *ACM Transactions on Sensor Networks (TOSN)*, vol. 5, no. 3, 2009.

[17] J.-H. Hoepman, "Privacy Friendly Aggregation of Smart Meter Readings, Even When Meters Crash," *2nd Workshop on Cyber-Physical Security and Resilience in Smart Grids*, pp. 37, 2017.