8-1-2015

PRIVACY PRESERVATION IN A HYBRID MULTI MESH-LTE AMI NETWORK FOR SMART GRID

Ozan Cakmak
Southern Illinois University Carbondale, cakmak.ozan.35@gmail.com

.

PRIVACY PRESERVATION IN A HYBRID MULTI MESH-LTE AMI NETWORK FOR SMART GRID

by

Ozan Cakmak

B.S., Izmir University of Economics, 2008

A Thesis
Submitted in Partial Fulfillment of the Requirements for the
Master of Science Degree

Department of Computer Science
in the Graduate School
Southern Illinois University Carbondale
August 2015

THESIS APPROVAL

PRIVACY PRESERVATION IN A HYBRID MULTI MESH-LTE AMI NETWORK FOR SMART GRID

By

Ozan Cakmak

A Thesis Submitted in Partial

Fulfillment of the Requirements

for the Degree of

Master of Science

in the field of Computer Science

Approved by:

Kemal Akkaya, Chair

Henry Hexmoor

Mengxia Zhu

Graduate School
Southern Illinois University Carbondale
May 28th, 2015

AN ABSTRACT OF THE THESIS OF

OZAN CAKMAK, for the Master of Science degree in Computer Science, presented on 28 May 2015, at Southern Illinois University Carbondale.

TITLE: PRIVACY PRESERVATION IN A HYBRID MULTI MESH-LTE AMI NETWORK FOR SMART GRID

MAJOR PROFESSOR: Dr. Kemal Akkaya

While the newly envisioned Smart(er) Grid (SG) will result in a more efficient and reliable power grid, its collection and use of fine-grained meter data has widely raised concerns on consumer privacy. While a number of approaches are available for preserving consumer privacy, these approaches are mostly not very practical to be used due to two reasons: First, since the data is hidden, this reduces the ability of the utility company to use the data for distribution state estimation. Secondly and more importantly, the approaches were not tested under realistic wireless infrastructures that are currently in use. In this thesis, a meter data obfuscation approach to preserve consumer privacy is proposed to implement that has the ability to perform distribution state estimation. Then, its performance on LTE and a large-scale Advanced Metering Infrastructure (AMI) network built upon the new IEEE 802.11s wireless mesh standard are assessed. LTE/EPC(Evolved Packet Core) model is used between the gateway and the utility. EPC's goal is to improve network performance by the separation of control and data planes and through a flattened IP architecture, which reduces the hierarchy between mobile data elements. Using obfuscation values provided via this approach, the meter readings are obfuscated to protect consumer privacy from eavesdroppers and the utility companies while preserving the utility companies' ability to use the data for state estimation.The impact of this approach on data throughput, delay and packet delivery ratio under a variety of conditions are assessed.

ACKNOWLEDGMENTS

It is a pleasure to thank the many people who made this thesis possible. First of all, it was a great honor to have had my M.S. studies under the direction of such an awesome advisor, Professor Kemal Akkaya. I would like to express my gratitude to him for all his invaluable guidance and support while conducting this research. I cannot have imagined having a better advisor and mentor for my graduate study. Words are inadequate to thank him.

I also would like to thank Dr. Henry Hexmoor and Dr. Michelle M. Zhu for being members of my committee and giving their insightful comments and suggestions through forming my thesis work.

I am so grateful to Samet Tonyali's help. He has always helped whenever I need help in my thesis. He is one of the greatest person that I have met in my life. I would also like to thank other my friends, Izzet F. Senturk, Mahbubul Arefin Khan, and Abdullah Albuali for their support and helpful comments.

I would also like to thank my family and friends for their unconditional and loving support over all the time.

Finally, I would like to acknowledge the support from National Science Foundation (NSF) during my MS studies. My research was supported by NSF under the grant number 1319087.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1
INTRODUCTION

The future SG is envisioned as a viable solution for finding efficient and economic methods of addressing a combination of several challenges: 1) using electricity more efficiently; 2) reducing the impact of energy production on the environment; 3) integrating renewable energy generated by individuals; and 4) building the framework necessary for the use of electrical vehicles [1]. One part of the SG initiative that is currently being implemented is the AMI, which provides two-way communication between the utility company and consumers' "smart" meters (SMs). The utility companies can use this infrastructure to monitor power demands over short periods, provide more accurate billing as well as utilize dynamic pricing to facilitate the reduction of peak demand. Typically, two-way AMI communication is ensured via a wireless mesh network which can be based on either IEEE 802.15.4 or IEEE 802.11s standards [1][2].

Despite its potential, the implementation of the AMI has arisen concern of consumer privacy, since the fine-grained meter data being collected could be used to infer activities and behavior patterns of consumers [3]. The frequency of the data depends on the application and the premise and can be from 6 secs (for businesses) to 15 mins (for residential) as opposed to once a month in the existing grid [1]. This fine-grained data can reveal the types of activities going on in the house. Obviously, this is against their privacy and can have social impacts. For example, a curious person can run simple signal capturing devices to know what his/her neighbors are doing. A jealous person may be interested in knowing if his spouse invites friends to home when he/she has night shifts or travels. Similarly, at the commercial level, some companies may need to spy on their competitors. For example, smart grid traffic analysis can reveal how long a factory works, the number of workers present in the factor, etc. Revealing this information can cause financial losses, e.g., if a company knows that its competitors are producing too many products, it can work on reducing the price of this product by offering sales on their products. Thus, any effective method of collecting and using fine-grained consumption information from SMs must provide sufficient protection of consumer privacy while preserving the suitability of the data for legitimate uses.

Recently, there has been much research for addressing this privacy issue under different assumptions [4]. While some of the approaches focus solely on the confidentiality of the meter data during its transit, others additionally strive to hide it from utility companies by leaving the data handling to trusted third parties (TTP). In this way, privacy can be provided by only giving the utility company the chance to do monthly billing and thus no access to individual readings; however, the utility companies then cannot perform distribution state estimation of the power grid, which is crucial to them for the reliability of the grid.

To also accommodate the ability to perform state estimation at the distribution level, data obfuscation techniques can be exploited [5]. The idea of data obfuscation is to hide the actual energy usage by randomizing the fine-grained meter data. By perturbing the collected reading values in a linear space, the utility company can still monitor the distribution network and calculate billing for given intervals. Nonetheless, the obfuscation operation necessitates the distribution of obfuscation values to each of the SMs in a secure and efficient way. Such distribution of values within the AMI network is crucial in order to ensure the plausibility of the privacy in addition to providing classical security services such as authentication and integrity. Another major missing component in the existing privacy approaches is the feasibility and scalability of the implementation in a realistic network by considering different parameters regarding the network architecture. This thesis addresses these two issues in a comprehensive manner.

Long Term Evolution (LTE) is a significant evolution of 4G network and the latest standard including its origin in the GSM / GPRS legacy networks family that is presently being deployed worldwide. LTE has high capacity with increased peak data rate of 1 Gbps for the downlink and 500 Mbps for the uplink, higher spectral efficiency, increased number of simultaneously active subscribers, and improved performance at cell edges [6]. It has been thought [7] that the annual LTE-based communication nodes shipment will exceed 5 million units by 2020. According to a report of The Global Mobile Supplier Association (GSA), LTE has been the quickest developing mobile technology ever [8]. It is believed that LTE answers the market's demand. LTE provides high speed, high capacity wireless communication with good Quality of Service (QoS) as well as low latency. These properties make LTE useful for critical applications in SG as well as for Neighborhood Area Networks, Wide Area Networks, and etc. LTE [9] has attracted substantial interest in a Smart Grid environment since there are some reasons: (1) low cost due to the economies of scale associated with a global standard, (2) wide area wireless support across a power transmission and distribution network, (3) efficient use of radio resources because of packet switching, (4) support for applications with different Quality-of-Service (QoS) requirements and built-in security functions.

Specifically, by assuming an 802.11s-based wireless mesh network and LTE for the AMI, secure obfuscation value distribution approaches are proposed in order to implement data obfuscation in an efficient manner based on a number of security goals identified. Then this distribution approach and simulate obfuscated data traffic are implemented on ns-3 [10] by using a draft version of the 802.11s and LTE implementation. Our goal is also to assess the overhead it brings to the network and compare the overhead to that of a regular 802.11s network and LTE which do not provide privacy preservation. The simulation results revealed that such use of obfuscation does not bring any overhead in terms of packet losses or packet delay compared to other approaches. With the ability to perform state estimation, providing consumer privacy and LTE EPC model, the approach is feasible to be used in 802.11s-based AMI networks and LTE.

The rest of the thesis is organized as follows. In the next section, a summary of the literature review is provided. In Section 3, some background on AMI network, state estimation in power grids, LTE EPC model and describe the assumptions, security goals and problem are provided. In Section 4, the approach for obfuscation value distribution and obfuscated reading collection in a secure manner by a wireless mesh network and LTE are presented. In Section 5, the approach is presented for obfuscation value distribution in a secure manner. Section 6 is dedicated to security analysis and experiment results. Finally, this thesis is concluded in Section 7.

CHAPTER 2
LITERATURE REVIEW

This chapter summarizes an overview of communication in the Smart Grid and a summary of approaches to providing security and privacy to Smart Grid environment.

2.1 PRIVACY PRESENTATION IN SMART GRID AMI

In general, the utility companies need fine-grained meter data for each customer to monitor demand and the state of the distribution network as well as utilizing dynamic pricing to reduce peak demand. Also, the utility companies would prefer to have the data to generate the bill on-site rather than relying on each individual SM. Thus, the SG should allow the utility companies to collect and use this fine-grained data while protecting it from being used to monitor or profile an individual consumer's behavior.

Various approaches of providing consumer privacy are surveyed in [4]. This work categorizes these approaches into three groups: approaches that anonymize the fine-grained meter data, approaches that mask or obfuscate the individual consumption, and approaches that focus only on protecting communication of meter data from threats in the communication network. In practice, the third category does not provide any additional mechanism for ensuring privacy other than relying on the trustworthiness of the utility companies.

Two approaches can satisfy the aforementioned requirements: anonymizing the data by using an escrow service [11] and masking usage patterns by using an internal power supply to shape the actual consumption data [12]. Masking usage patterns with an internal power supply enables the utility company to collect the actual data without posing a privacy risk. It can also help reduce peak demand by having the power supply provide power during peak time and recharge when demand is low. However, the cost of implementation and maintenance of these systems make it less than ideal to both the utility company and the consumer.

The approach using an escrow service creates two identities for each customer: one for the SM to use for communication related to billing purposes and another for the collection of fine-grained meter data [11]. While this approach is simple, it has its own problems. First, it requires additional hardware and labor for maintaining two separate IDs to communicate with utility and escrow service independently. More importantly, the approach requires the escrow service to be trustworthy about the actual identities. This could then be circumvented by capturing both types of data and inferring the relationship between an individual SM's two identities. Since the utility company possesses the actual fine-grained meter data, it would then be able to determine the behavior of individual customers.

Our privacy approach in this thesis is based on data obfuscation or hiding. One of these obfuscation approaches in [5] explains the necessary theoretical background on the ability of protecting consumer privacy while also allowing the utility company to perform state estimation, billing and dynamic pricing. However, the adaptation of these ideas in AMI infrastructure and protocols have not been studied. Our work in this thesis makes contributions on these aspects: First, we adapt an IEEE 802.11s-based mesh wireless network and LTE for the implementation of data collection mechanisms. Second, we present two secure protocols for generating and distributing the obfuscation values from the utilities to the SMs in such a large-scale network. Finally, we assess the overhead of obfuscation value distribution protocol in this mesh network and LTE. The main advantage of this approach is the low computation overhead compared to cryptographic mechanisms since only addition and subtraction operations are required to create obfuscated readings.

2.2 STATE ESTIMATION

SG should monitor the states of the grid to be able to take the most suitable action for the network, generations, and consumers [13]. Therefore, state estimation is indispensable for SG functionality. In power transmission networks, state estimation at the transmission level has been studied since 1970s [14]. At the the distribution level, there has been several studies in recent years. For instance, Dzafic et al. [15] presented a new approach for a three-phase distribution state estimation (DSSE). The approach makes use of real time measurements such as current magnitude, real and reactive power measurements. The weighted least squares error method is used for estimation of the state variables. In another recent study for distribution state estimation, [16] investigate the use of SMs' power injection measurements for low-voltage system observability and controllability. Failure resistance tests demonstrated that the state estimation accuracy can be kept at a good level unless approximately 40% of all the SMs in the network are lost. In addition to this study, [17] develop a state estimator for low-voltage networks with and without distributed generations.

Following the ability of SMs to contribute to the state estimation computations, in this thesis, we also target DSSE where SM meter readings are used. However, since the use of SMs would introduce privacy concerns, we focus on the problem of privacy preserving state estimation which has not been studied before in power grid state estimation research.

2.3 LTE IN A SMART GRID ENVIRONMENT

LTE is a promising alternative for a smart grid [18, 19]. A few LTE scheduler designs are presented for common user equipments as mobile phones. In [20, 21] and [22], these scheduler designs have quite separate requirements in comparison with the services asked by smart grid components. Many commercial service providers has partnered with utilities to provide communications for Smart Grid applications. They have supported many technological improvements such as the general movement toward integrated platforms and open standards for utility communication functions that are facilitated opportunities for improved communication systems [23]. There are a few amount of research that is conducted into the performance of an LTE system in a Smart Grid environment. Most of the traffic with AMI is anticipated to be in the uplink direction. Therefore, LTE Time Division Duplex (TDD) was searched in [24] as a possible solution for AMR/AMI using LTE TDD configurations 0, 1, or 6, that are uplink based. It was compared with LTE Frequency Division Duplex (FDD) in [25]. FDD leads to better uplink performance with respect latency even though TDD can support greater flexibility when the division between uplink and downlink data is unsymmetrical. The main reason is due to the delays caused by the interchange of the uplink and downlink slots in LTE TDD.

A smart grid based 20 MHz LTE Time Division Duplex (TDD) system is implemented and uplink error rate curves are done; however, there is no treatment of latency or channel utilization [26]. The performance of a 5 MHz LTE TDD system that uses uplink and downlink allocation configuration one of them in serving a Smart Grid Distribution Automation (DA) application [27]. It is shown that the best case and average latency are not addressed, and is not channel utilization although the maximum uplink latency is 66 ms (allowing for up to 3 packet re-transmission).

LTE specifies a new all-IP, packet based core network called Evolved Packet Core (EPC). One of the crucial elements of the EPC is that control and data planes have separate interfaces and network entities [28]. This LTE/EPC architecture is performed to provide seamless and ideal IP connectivity between user equipment (UE) and external Packet Data Networks (PDNs). LTE/EPC also provides on-demand connectivity service. This service includes moving active sessions transparently and temporarily from one network equipment to another without causing user session interruption. This service is critical in situations such as network equipment, overload situations, and during energy saving measures. In this thesis, we focus LTE/EPC model between SG gateway and utility. LTE/EPC model is explained in detail in Section 3.4.

CHAPTER 3

PRELIMINARIES

3.1 UNDERLYING AMI NETWORK

We assume that the AMI communication network consists of SMs that are connected via a wireless mesh network with a gateway serving as a relay between the SMs. SM measures mainly the real-time electrical energy consumption of the customers in addition to power quality and instantaneous values such as voltage and current at their connection points. In the current AMI systems, this data is either collected by the utility company control center or by a trusted third party (TTP). In this thesis, we assume the existence of a TTP as well since handling every type of processing in a central control center is not a scalable option [5]. In addition, a TTP can use cloud computing infrastructure which is not only more cost effective but also provide the opportunity to share meter data among utility companies to estimate the state in wide area grids. A typical infrastructure for the considered AMI in this thesis is shown in Figure 3.1.

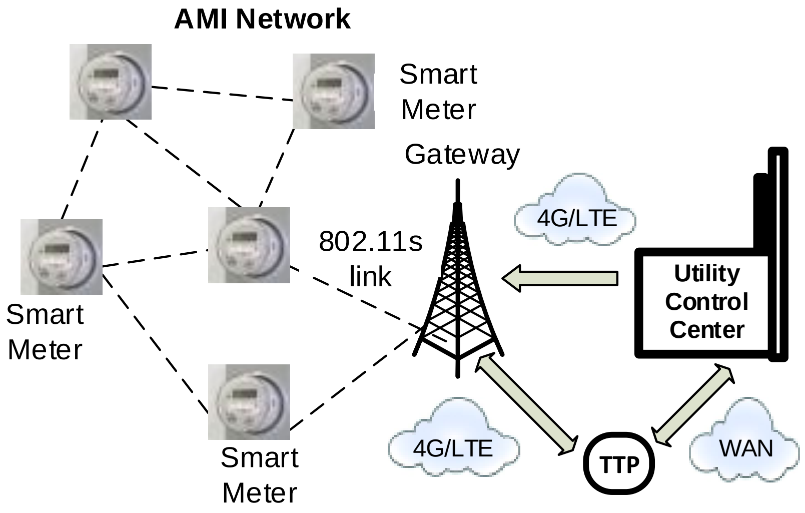

Figure 3.1. A sample AMI communication network, gateway and long-distance communication to a utility company

The mesh network in this figure is created using the new IEEE 802.11s standard which allows mesh networking among the SMs through 802.11 MAC/PHY layer standard [29]. The nodes in 802.11s WMN are given names based on their roles. All nodes are considered as Mesh Points (MP) and are able to provide connectivity at the data link layer between other MPs. If an MP also provides connectivity to the Internet, it is termed a Mesh Portal Point (MPP). In our mesh network, the gateway is the MPP which collects meter readings that are obfuscated at the MPs (i.e., SMs) via multi-hop routes. IEEE 802.11s standard provides a routing protocol called Hybrid Wireless Routing Protocol (HWMP) as its default routing protocol to find a multi-hop path towards the destination.

TTP collects data from the gateway and stores this data for future access but also forwards collected data in form of a data vector (e.g., an array) to the utility. In addition, TTP is responsible for monthly billing computations. The utility is responsible for creating and transmitting obfuscation base information for privacy preservation purposes to the gateway which in turn creates the individual obfuscation values and distributes to the SMs. The gateway communicates with the TTP and utility via LTE. However, the utility and TTP may have other means for communication.

For security and privacy, we assume the availability of public key schemes since symmetric key systems require a lot of overhead in terms of key management. Each SM is initialized with a public/private key based on elliptic curve cryptography (ECC). This was chosen since its overhead is the minimal in comparison to other public cryptography schemes. ECC also uses a key size comparable to current symmetric cryptographic schemes, avoiding the higher computation of other public key schemes due to the larger key size.The gateway knows the public key of every SM in its mesh network. Every SM knows the public key of the gateway.

3.2 WEIGHTED LEAST SQUARES STATE ESTIMATION

A power system consists of a collection of buses, transmission lines and power meters. State estimation is used to monitor the state of a power system (i.e., voltage magnitude and phase angle of every bus) in order to maintain reliable power supply. Recently, there is some interest to do state estimation in low-voltage distribution networks using meters and their instantaneous measurements (real power, reactive power and voltage magnitude) in addition to the measurements collected from the distribution system substation [30]. One of the techniques for this state estimation process is called common weighted least squares (WLS) state estimation.

In this technique, the state of the network is estimated as a vector of variables $x = (x_1, \dots, x_n)^T$ using $z = (z_1, \dots, z_m)^T$ consisting of measurements from the power meters, where n, m are positive integers such that $m > n$ and $x \in \mathbb{R}^n$ and $z \in \mathbb{R}^m$. Then, the state of the system is represented by:

$$z = h(x) + e \tag{3.1}$$

where $h : \mathbb{R}^n \to \mathbb{R}^m$ represents nonlinear dynamics such as the configuration of transformers and buses in the grid and $e \in \mathbb{R}^m$ is measurement errors and unmodeled dynamics. The state x is estimated to be $\hat{x}$ by the following unbiased linear estimation:

$$\hat{x} = (H^T W H)^{-1} H^T W z \tag{3.2}$$

where W^{-1} is the covariance matrix of e.

3.3 PRIVACY-PRESERVING STATE ESTIMATION

Due to data collection from SMs, privacy came to picture in state estimation in distribution networks. One possible solution to this issue is to perturb the collected SM data. To this end, the authors in [5] create a *distortion free obfuscation space* from the span of a basis set $\mathbb{O} = \{o_1, \dots, o_{m-n}\}$ of *kernel* denoted as $ker((H^T W H)^{-1} H^T W)$. Each

$o_i \in \mathbb{O}$ is a vector with m elements that will perturb the SM values. Note that there are $m - n$ such vectors that can be used for perturbation. The authors also create an *obfuscated measurement vector* named z_{obf}, where $z_{obf} = z + o$, $o \in \mathbb{O}$. They show that z_{obf} can be used in place of z to calculate the same estimated state $\hat{x}$. In this way, without having access to actual power readings, the state of the power grid can be estimated.

$\mathbb{O}$ is derived from the state of distribution of the grid such as the configuration of the transformers stepping up or down voltages or buses branching off to multiple distribution lines. It can be sent by the utility company only when any one of these dynamics changes. It can be reused for multiple readings until new ones are provided.

In order to create an obfuscation vector o, some random η weight values need to be generated. At measurement time t_j, the goal is to choose a random weight $\eta_i^j \in \mathbb{R}$ for a vector $o_i \in span(\mathbb{O}) = \{\eta_1^j o_1 + \eta_2^j o_2 + \eta_3^j o_3 + \ldots + \eta_{m-n}^j o_{m-n}\}$. The values of the weights η_i^j at each data collection time t_j in a billing period T are chosen so that the sum of the values of η_i^j in T equals to 0 (i.e., $\sum_{j \in T} \eta_i^j = 0$).

After the above computation is done, an element from the vector o is sent to the corresponding SM. Each SM adds this element to its actual power measurement in order to conceal it and to preserve the privacy by this way.

3.4 LTE/EPC MODEL

EPC is required for end-to-end IP service delivery across LTE and is defined as a generic IP access network. It provides IP Packet transmission only. EPC has a essential specialty that circuit-switched based voice service and packet-switched based data service of earlier 2G and 3G networks are combined under one single IP mobile domain [31]. In addition, data plane and control plane have interface and network entity distinctly which allows operators to optimize their signaling and traffic capacity separately.

The main goal of the EPC model [32] is to implement for the simulation of end-to-end IP connectivity over the LTE model. This model supports for the interconnection of multiple User equipments (UEs) to the Internet, via a radio access network of multiple Evolved Node Bs (eNBs) connected to a single Serving Gateway (SGW) / Packet Data Network Gateway (PGW) node.

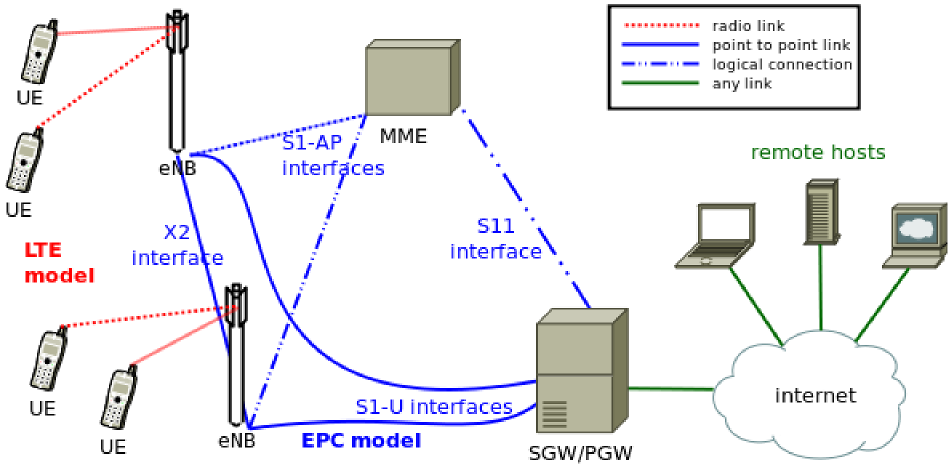

Figure 3.2. Overview of the LTE-EPC Model

As illustrated in Figure 3.2., LTE/EPC consists of the following main network subcomponents:

- *Mobility Management Entity (MME)*: The MME is the main control plane network element and the key control-node for the LTE access-network. It operates user mobility and session management signaling including such as establishment, release and coordination of bearer signaling. It is responsible for tracking, paging, authorization and authentication for UEs. Moreover, it is liable to select S-GW for UEs in initial network attachment and re-location scenarios.

- *Serving Gateway (S-GW)*: S-GW relays user data to and from eNodeB and packet data network. It divides among resources as for each the need of other entities such as MME and P-GW. S-GW is responsible for handovers with neighboring eNodeB's, and also in terms of all packets across user plane for data transfer.

- *Packet Data Network (PDN) Gateway (P-GW)*: The PDN Gateway specifies connectivity from the UE to external packet data networks by being the point of exit and entry of traffic for the UE. It is liable for allocating IP address to the UE.

- *User Equipment (UE)*: UE is any device used directly by an end-user to connect each other. It can be a mobile telephone, a laptop computer with a mobile broadband adapter, or any other device. It connects to the base station Node B/eNodeB.

- *Evolved Node B (eNB)*: It is the hardware which is connected to the mobile phone network. It communicates directly with User Equipments (UEs). Its functionality likes a base transceiver station (BTS) in GSM networks.

LTE S1 interface connects eNodeB to EPC network. S1-C (Control Plane) is designed to assign signaling messages between eNodeB and MME. During this time, S1-U (User Plane) interface transports user data to S-GW. S5/S8 interface transfers data and control signals between S-GW and P-GW. MME communicates with S-GW over S11 interface. LTE utilizes the Evolved Packet System (EPS) bearers to transfer IP data and from UE and P-GW that connect S1-U and S5/S8 interfaces containing air interface [33]. Moreover, LTE S1- flex configuration allows one eNodeB to connect to multiple MMEs by handling S1 interface as a many-to-many interface. X2 interface is presented in LTE for inter-eNodeB communication.

3.5 PROBLEM DEFINITION

Our problem can be defined as follows: "Given an 802.11s-based wireless mesh network and LTE, our primary goal is to distribute the obfuscation values to the SMs and collect the obfuscated values from them in a secure and efficient way via TTP. Our secondary goal is to assess the overhead of this process in a large scale AMI network and LTE, and thus analyze the feasibility of the approach for future SG applications." The security goals are elaborated next along with the corresponding attacks. Note that these security goals are in addition to the goals of preventing misuse of finely-grained meter data and preserving the ability to perform state estimation.

3.6 THREAT MODEL AND SECURITY GOALS

The following attacks to the privacy and security of the collection of fine-grained meter data are identified in the AMI and established the associated security goals. They are organized into two sets: those targeting the consumer and those targeting the utility company. The first set relates to the privacy of a consumer's fine-grained meter data.

- *Attack 1*: The utility company misuses fine-grained meter data it obtains to analyze consumer behavior or shares the data with a third party for this purpose.
- *Security Goal 1*: Obfuscate the collected fine-grained meter data to protect it from misuse by the utility company or related third party.
- *Attack 2*: An eavesdropper monitors the communication channel to capture meter data in messages between a targeted SM and the gateway to determine the behavior of its consumer.
- *Security Goal 2*: Protect communications containing SM readings.
- *Attack 3*: An eavesdropper compromises a gateway to gather the obfuscation basis $\mathbb{O}$ that is stored to re-generate actual meter readings.
- *Security Goal 3*: Limit the amount of obfuscation data that could be obtained if a gateway is compromised.

The second set of attacks relates to accurate state estimation and billing.

- *Attack 4*: An attacker impersonates the gateway and sends fabricated obfuscation values to the SMs to change the state of the power grid.

- *Security Goal 4*: Provide sender authentication to verify the sender and contents of messages.
- *Attack 5*: An attacker captures the obfuscation values and replay them to change the state or billing.
- *Security Goal 5*: Identify and discard replayed messages.

CHAPTER 4

DATA OBFUSCATION ON A WIRELESS MESH NETWORK AND LTE

In this section, the design of a realistic architecture and procedures for obfuscating and collecting SM data is described in detail. The approach has two phases: First, obfuscation values are created by the gateway and distributed to the SMs. Second, each SM creates its obfuscated power reading and transmits it to the gateway. Note that our approach avoids the assumption that each SM has a communication link with the TTP. The gateway transmits all the data to the TTP which is responsible to create the data vector and transmit to the utility control center for state estimation. TTP also performs billing computations at the end of each billing period and stores all obfuscated customer data for archival purposes.

4.1 CREATING THE OBFUSCATION VECTOR

The gateway is responsible for creating the obfuscation vector. To do this, the utility company first sends the basis of the obfuscation space, $\mathbb{O}$, to the gateway directly. The gateway randomly selects weights (η) for each of the vectors in $\mathbb{O}$ and constructs an obfuscation vector. An example for a simple mesh topology is provided in Fig. 4.1. In this example, upon receiving $\mathbb{O}$, the gateway randomly chooses weights for each vector v_i in $\mathbb{O}$ and constructs the actual obfuscation vector o. Note that the row size of this vector will be equal to the size of the SMs in the network. If there is single gateway as assumed in our case, each SM will be assigned one element from this vector o.

Nevertheless, given the large size of the AMI networks, this may not be feasible and the network may need to be divided into multiple clusters of SMs each led by a different gateway. For those cases, our approach will still apply with one difference: Each gateway will get the same $\mathbb{O}$ and will create an obfuscation vector for all the SMs. However, each gateway only serves a subset of all the SMs. Therefore, when distributing the obfuscation

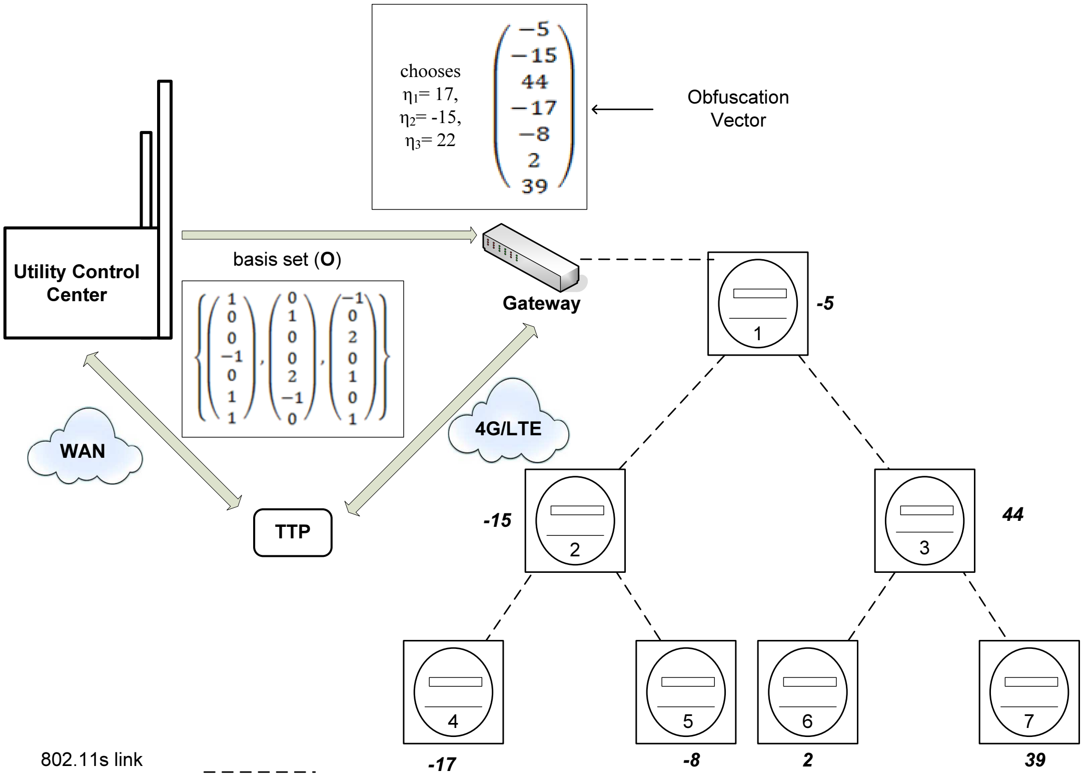

Figure 4.1. A mesh topology of 7 SMs. First, the gateway receives the obfuscation basis from the utility and creates the obfuscation vector by picking random η values and multiplying them with each vector of the $\mathbb{O}$. Then, each obfuscation element (italic) (i.e., $o[j]$ where j : 1 to 7) is communicated to its corresponding SM.

values, only the SMs that are within the cluster of that gateway will be contacted. Nonetheless, this approach may not be efficient in terms of $\mathbb{O}$ transmissions and computations.

Let a sample $\mathbb{O}$ be:

$$\{(1,0,0,-1,0,1,1)^T,$$

$$(0,1,0,0,2,-1,0)^T,$$

$$(-1,0,2,0,1,0,1)^T\}$$

and the gateway randomly generates the weights $\eta_1 = 17, \eta_2 = -15, \eta_3 = 22$ at time t_1. It then constructs the obfuscation vector o as follows:

$$17 \times (1, 0, 0, -1, 0, 1, 1)^T$$

$$+ (-15) \times (0, 1, 0, 0, 2, -1, 0)^T$$

$$+ 22 \times (-1, 0, 2, 0, 1, 0, 1)^T$$

$$= (-5, -15, 44, -17, -8, 2, 39)^T.$$

Let v_i denote the i^{th} vector in every $\mathbb{O}$, the gateway stores a sum of (η) values for all v_i during a particular billing period T where i can get values from 1 to the number of vectors in $\mathbb{O}$. Let T is divided into the following epochs: $\{t_1, t_2, ...t_n\}$, and $\eta_i^{t_j}$ denotes the η value for v_i at time t_j, then the sum of all η values for all v_i is: $sum_i = \sum_{j=1}^{j=n} \eta_i^{t_j}$. When the final meter reading for a billing period is collected, the gateway chooses the weight ($\eta_i^{t_n}$) for a particular vector v_i so that the sum_i becomes zero. Thus, $\eta_i^{t_n}$ is chosen as $-\sum_{j=1}^{j=n-1} \eta_i^{t_j}$ so that $\sum_{j=1}^{j=n} \eta_i^{t_j} = 0$. For the example in Fig. 4.1, if we assume that there are two more collection times, t_2 and t_3, using the same $\mathbb{O}$, the weights can be as follows: At t_2, let us assume that the weights are again randomly chosen as 10, 12, 3. At t_3, the weights must be chosen as -27, 3, -25 so that the $sum_i = \sum_{j=1}^{j=3} \eta_i^{t_j} = 0$.

4.2 SECURE DISTRIBUTION OF OBFUSCATED VALUES

Once the obfuscation vector o is created at the gateway, the next task is to send these values to SMs in a secure and efficient way. To reduce the traffic, one possibility is to broadcast the whole vector within the network and let each SM pick its corresponding obfuscation value. However, there are issues regarding this method. First of all, TCP is used which does not support broadcast. Even if UDP is used without acknowledgements, this creates unnecessary flooding in the network where some SMs receive the same vector multiple times from their neighbors which will be redundant. In addition, the size of the whole vector will grow with the increased SM count and thus may necessitate additional broadcasts due to exceeding maximum transfer unit (MTU) for IEEE 802.11 standard. Given that SMs send readings at regular intervals, we opt to use inter-interval times to distribute the obfuscation values using unicasting capability of IEEE 802.11s standard through its routing protocol HWMP. The gateway prepares a packet for each SM and transmits to each SM separately.

Specifically, the gateway encrypts the elements in the vector with the public key (PU_i) of its corresponding SM_i. The gateway then sends each SM its corresponding element of the obfuscation vector (which is represented as $o[i]$) by signing it with its private key PR_G and adding a timestamp (TS) as follows. This is also illustrated in Fig. 4.1.

$$Gateway \rightarrow SM_i : \{< o[i], TS >\}_{PU_i}, Sig_{PR_G}(\{< o[i], TS >\}_{PU_i})$$

4.3 CALCULATING AND TRANSMITTING OBFUSCATED MEASUREMENTS

When an SM_i receives its element $o[i]$, it calculates its obfuscated power measurement (op_i) by adding its current power reading (p_i) and $o[i]$: $op_i = p_i + o[i]$. SM_i then timestamps (TS) the sum and digitally signs the message for the gateway using its private key, PR_i. SM_i then transmits this to the gateway again by using HWMP:

$$SM_i \rightarrow Gateway :< TS, op_i >, Sig_{PR_i}(< TS, op_i >)$$

Upon receiving the obfuscated measurements from each SM, the gateway verifies the digital signatures and timestamps. It then sends them to the TTP. For simplicity, we assume that the gateway can wait for all the SM readings and send them as a single packet.

TTP prepares the obfuscated measurement vector for the utility. In addition, when the billing period ends, it can sum all the measurements to obtain the total usage for

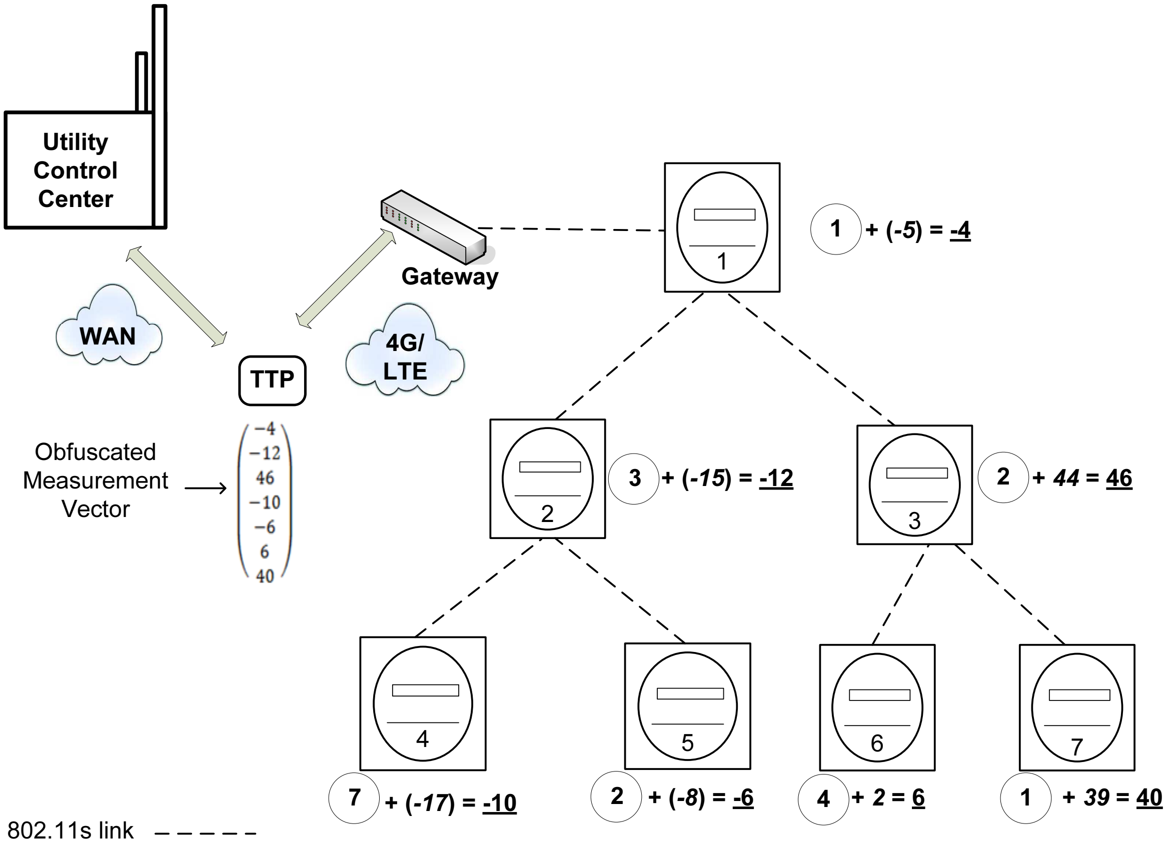

Figure 4.2. Each meter adds its current reading (circled) to the received obfuscation values to calculate its obfuscated reading (underlined). SM_4, for example, has a current reading of 7. It sums it with the obfuscation value it received, -17, obtaining an obfuscated reading of -10. The obfuscated readings are securely communicated back to the gateway which constructs the obfuscated measurement vector, z_{obf}. This is sent to the TTP.

each SM for the billing period to charge the customer. The utility receives the obfuscated measurement vector from the TTP and uses it for performing state estimation. Based on the $\mathbb{O}$ in Figure 4.1, the calculation and collection of the measurements are depicted in Figure 4.2.

4.4 USING LTE/EPC MODEL

As illustrated in Figure 3.2, the goal is to use the Evolved Packet Core (EPC) that is a flat architecture that provides a converged voice and data networking framework to connect users on a Long-Term Evolution (LTE) network.

In this study, Packet Data Node Gateway (PGW) acts as the interface and provides connectivity from the Utility as User Equipment(UE) to gateway in the LTE network. The PGW categorizes each incoming packet from gateway and routes it to a mobility tunnel that reaches the eNB (base station). The eNB maps and manages the data transmission to the Utility (as UsEr Equipment) on appropriate radio bearer channels.

CHAPTER 5
MULTIPLE GATEWAYS FOR INCREASED EFFICIENCY AND SECURITY

5.1 MOTIVATION AND OVERVIEW

The scheduled obfuscation value distribution approach with a single gateway poses some issues regarding security and efficiency. First of all, a single gateway will become a bottleneck when IEEE 802.11s is used with increased node count. This is because of the increased hop counts, congestion and interference in the network. Previous studies suggest using clustering within the AMI network and adjust the size of each cluster based on network conditions [34]. Therefore, it is preferable to use multiple gateways for scalability concerns. However, this is not the only issue. Another motivation to use multiple gateways is due to the security of the distribution of obfuscation values. If a single gateway is being used, it is possible for the entire obfuscation vector to be captured when the gateway is compromised. Since this obfuscation vector is a linear combination of the vectors of $\mathbb{O}$, a linear system of equations could be formed with $\mathbb{O}$ and the obfuscated readings received at the gateway, which can be solved to determine the η weights used to construct o. Since there are $m - n$ possible vectors of $\mathbb{O}$, one would need to have the information to form $m - n$ linearly-independent equations to be able to solve for the weights. These weights could then be used to reconstruct the entire obfuscation vector o, which could be used to calculate the actual readings from the obfuscated measurements. This can be protected against as long as a gateway uses less than $m - n$ vectors. Splitting the process of generating the obfuscation vector among multiple gateways could be used to reduce this vulnerability.

The solution described in the following is to separate the vectors of $\mathbb{O}$ among multiple gateways. Each gateway would then select the η weights for its given vectors and create one part of the obfuscation vector o. These could be sent directly to their associated SMs, but if each gateway is responsible for the SMs in a portion of the grid, each gateway would have to send to the other gateways the elements of their partial obfuscation vector that are associated to the meters for which that gateway is responsible. Since each gateway only has a portion of the basis from which the full obfuscation vector is derived, an adversary compromising one gateway would be limited in the number of actual readings it could acquire.

5.2 MULTI-GATEWAY COMMUNICATION PROTOCOL VIA LTE D2D

Multi-gateway communication protocol is geared for information exchange among the gateways. It is assumed that the AMI network is divided into multiple clusters, where each is led by a different gateway. Each gateway node has two radios one for 802.11s and one for LTE. Each gateway knows the IDs of SMs within its cluster and the public keys of other gateways in advance. The gateways need to communicate with each other to exchange the obfuscation elements using device to device (D2D) communication architecture that is recently being standardized for LTE-Advanced under proximity services (ProSe) studies [35] (also known as LTE-Direct [36]). While direct communications has been possible in unlicensed bands, e.g., using WiFi-direct, there are various challenges, such as the high interference in the unlicensed bands, as well as the difficulty of pairing and synchronizing D2D transmissions. In LTE-Advanced Release-12 standardization, preliminary studies on D2D and ProSe communications were initiated for commercial and public safety applications [37]. A major benefit of ProSe D2D communications in LTE-Advanced is that device pairing, resource allocation, and power control can be coordinated through a base station (which e.g. can be embedded in the utility control center in Fig. 4). This allows improved spectral efficiency and lower scheduling delays compared to completely uncoordinated scheduling of D2D transmissions. Use of licensed bands also allows improved quality of service for D2D transmissions.

Each gateway would run the same protocol explained in the previous section within its cluster once the obfuscation vector is formed. However, the formation of this obfuscation vector is different in this case. The protocol begins when the utility provider derives the obfuscation basis $\mathbb{O}$. It splits $\mathbb{O}$ into g components where g represents the number of gateways involved. Then each components is encrypted with each gateway's public key and sent to that gateway through LTE network.

When a gateway involved in the obfuscation generation receives its *partial* $\mathbb{O}$, it chooses the η weight(s) and calculates its *partial* obfuscation vector. This vector would have obfuscation values for all the SMs in the network but the gateway distributes only the ones that belong to its cluster. The other portions are transmitted to the other gateways (in encrypted form) so that they can distribute corresponding values within their own clusters.

In this way, a reading of an SM in a particular cluster will be obfuscated by adding g obfuscation values as opposed to one in the previous approach. Again, for a particular billing period T, the η weights for the same vector are chosen in such a way that the sum of these weights would be zero. Note that the same security operations of signing and encryption still apply. In Fig. 5.2 you can see an illustration of how two gateways generate obfuscation values, exchange them and distribute them to SMs.

5.3 ALGORITHM AND ANALYSIS

Algorithm 1 gives the pseudo-code for the procedure running at a particular gateway for the processes of generating the weights, creating obfuscation values and distributing them among the gateways.

When an SM receives its obfuscation element, it decrypts it and adds it to its current reading. The obfuscated readings are signed, timestamped and sent back to the gateway, which collects all of readings in this cluster and sends them to a the TTP via LTE. This process is illustrated with a simple example in Fig. 5.3. It is the second phase of what is given in Fig. 5.2

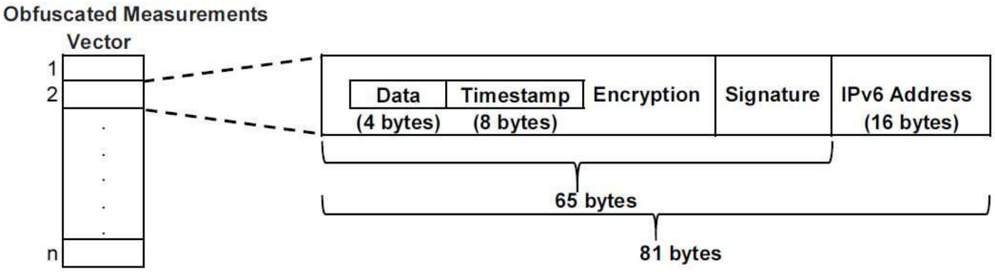

Figure 5.1. Contents of the Obfuscation Vector

If there are g gateways, the utility provider needs to send g messages for transmitting out the partial basis. Each gateway involved in the obfuscation vector generation would need to contact $g - 1$ gateways to send its own obfuscation vector. The remaining steps are similar to those of the single gateway approach and at the end there will be g total messages sent from the gateways to the TTP. Since the inter-gateway communication will be using LTE-Direct, it is argued that this will not put any burden on the gateways. Typically, there will be 648 bits allocated for each entry of a vector as shown in Fig. 5.1. Since LTE-Direct can send up to 3Gbps, this would allow sending up to 5000 SMs' readings at the same time under perfect conditions.

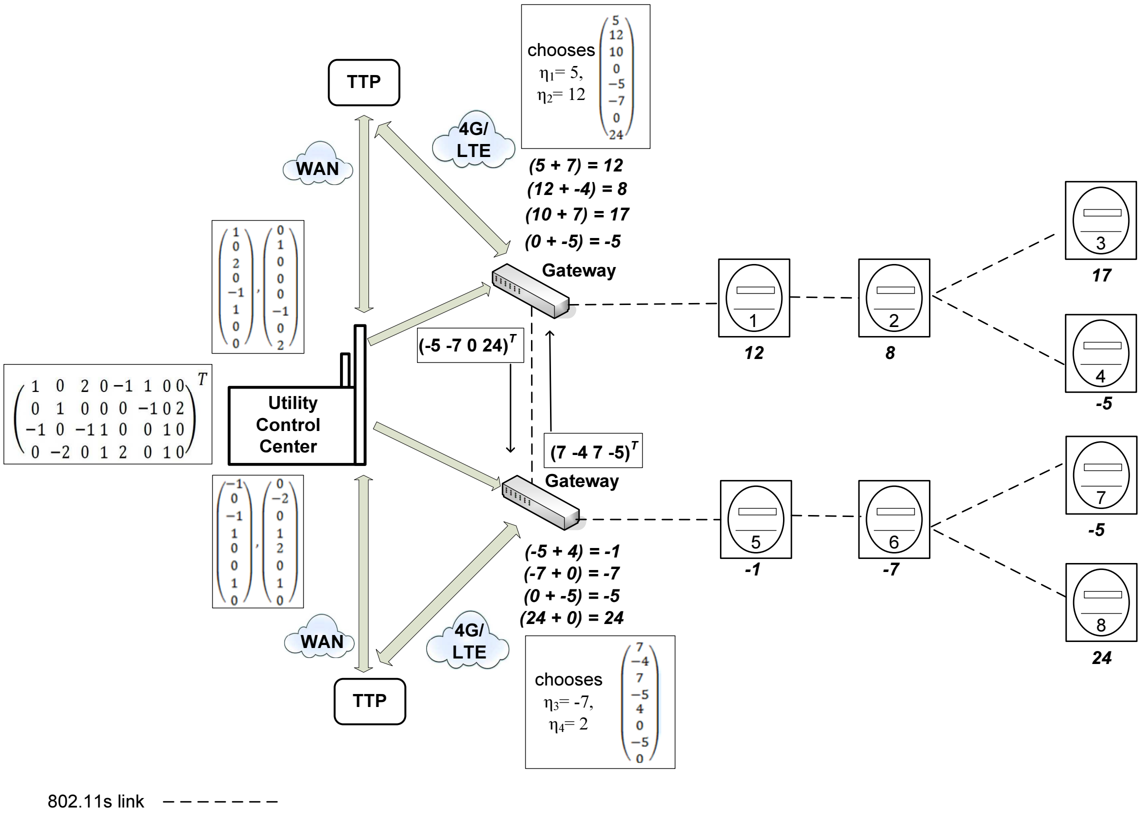
TTP
WAN
4G/ LTE
chooses η1= 5, η2= 12
$(5\ 12\ 10\ 0\ -5\ -7\ 0\ 24)^T$
(5 + 7) = 12
(12 + -4) = 8
(10 + 7) = 17
(0 + -5) = -5
Gateway
$\begin{pmatrix}1\\0\\2\\0\\-1\\1\\0\\0\end{pmatrix}, \begin{pmatrix}0\\1\\0\\0\\0\\-1\\0\\2\end{pmatrix}$
$\begin{pmatrix}1 & 0 & 2 & 0 & -1 & 1 & 0 & 0\\0 & 1 & 0 & 0 & 0 & -1 & 0 & 2\\-1 & 0 & -1 & 1 & 0 & 0 & 1 & 0\\0 & -2 & 0 & 1 & 2 & 0 & 1 & 0\end{pmatrix}^T$
Utility Control Center
(-5 -7 0 24)^T
(7 -4 7 -5)^T
Gateway
$\begin{pmatrix}-1\\0\\-1\\1\\0\\0\\1\\0\end{pmatrix}, \begin{pmatrix}0\\-2\\0\\1\\2\\0\\1\\0\end{pmatrix}$
(-5 + 4) = -1
(-7 + 0) = -7
(0 + -5) = -5
(24 + 0) = 24
WAN
4G/ LTE
TTP
chooses η3= -7, η4= 2
$(7\ -4\ 7\ -5\ 4\ 0\ -5\ 0)^T$
1
12
2
8
3
17
4
-5
5
-1
6
-7
7
-5
8
24
802.11s link

Figure 5.2. Two gateways case with a mesh topology of 8 SMs. First, the gateways receive their obfuscation basis from the utility provider and create the obfuscation vector by picking random η values and multiplying them with each vector of the $\mathbb{O}$. The gateways exchange some obfuscation elements in their obfuscation vector according to the SMs they control. Then, each obfuscation element (italic) (i.e., $o[j]$ where $j\ :\ 1$ to 8) is transmitted to its corresponding SM by each gateway.

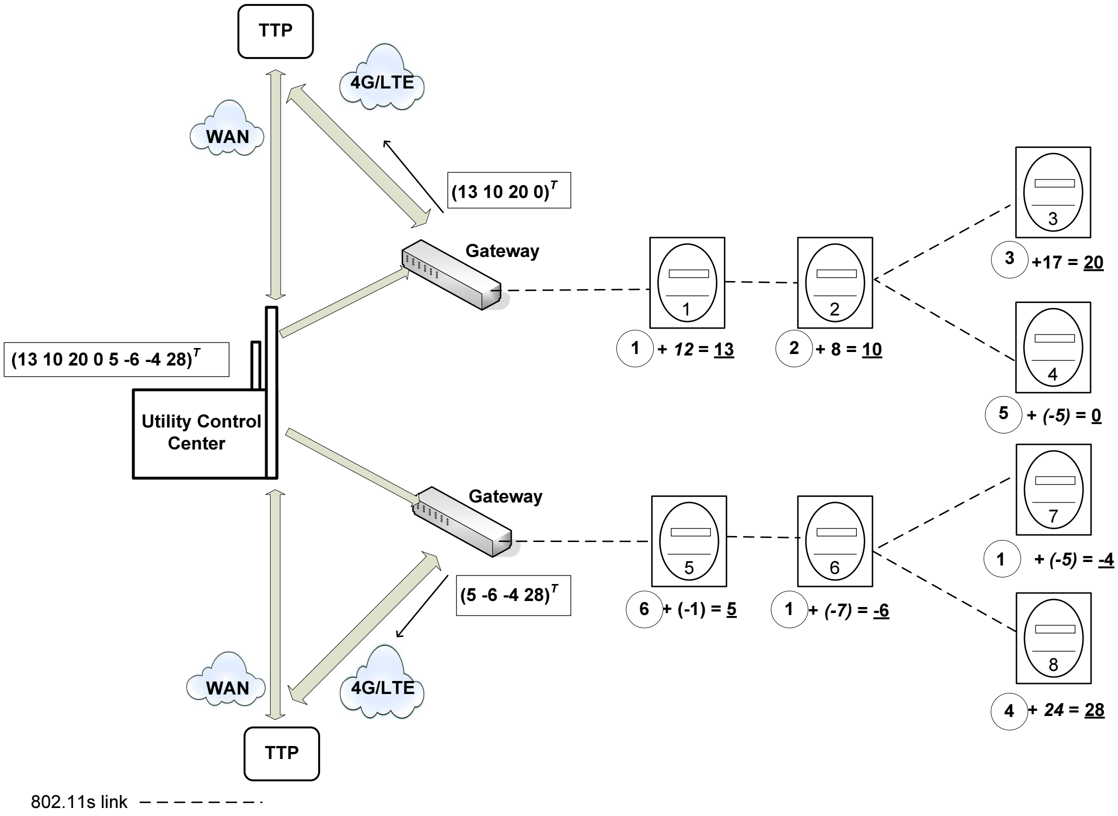

Figure 5.3. Data transmission with two gateways: Each SM adds its current reading (circled) to the received obfuscation values to calculate its obfuscated reading (underlined). SM_7, for example, has a current reading of 1. It sums it with the obfuscation values it received, 0 and -5, obtaining an obfuscated reading of -4. The obfuscated readings are securely transmitted back to the gateway from which the obfuscation values are received. The gateways construct the obfuscated measurement vector, z_{obf} and send their vectors to the TTP.

Algorithm 1 distributeObfVals(obfBasis[][], collTimes)

```
numOfAllSMs = obfBasis[0].$length$
weights[collTimes][collTimes]
initObfVect[collTimes]
obfVectFrGWs[$numOfAllGWs$ - 1][$numOfNSMs$]
ultimateObfVect[$numOfNSMs$]
i = 0
for all  $i \leq collTimes$ - 1 do
    if (i == $collTimes$ - 1) then
        weights[i] = generateComplementWeights(weights)
    else
        weights[i] = generateRandomWeights($collTimes$)
    end if
    j = 0
    for all $j \leq numOfAllSMs$ - 1 do
        initObfVect[j] = 0
        k = 0
        for all $k \leq collTimes$ - 1 do
            initObfVect[j] += (weights[i][k] * obfBasis[k][j])
            k++
        end for
        j++
    end for
    deliverObfValsToGWs(initObfVect)
    obfVectFrGWs = receiveObfValsFromGWs()
    ultimateObfVect =
         combineVects(initObfVect, obfVectFrGWs)
    sendObfValsToSMs(ultimateObfVect)
    waitForNextCycle($waitTime$)
    i++
end for
```

CHAPTER 6

PERFORMANCE EVALUATION

6.1 SECURITY ANALYSIS

In this section, the *scheduled obfuscation* approach based on the security goals listed in Section 3.6 are evaluated.

- *Security Goal 1*: Since the fine-grained meter data is obfuscated, the actual reading cannot be determined at any time. Because of this, it cannot be analyzed to determine any activity or behavior of the consumer.

- *Security Goal 2*: The obfuscated reading that the SM sends to the gateway does not reflect the actual reading. Therefore, even if an eavesdropper captures this reading, its inference about the activity in the house will be wrong. Also, since the gateways disseminate different obfuscation values at each reading collection period, the eavesdropper cannot extract a pattern of the consumer's power consumption.

- *Security Goal 3*: If a gateway is compromised, the obfuscation information regarding that cluster could be obtained. However, since there are other obfuscation values coming from the other gateways, the attacker needs to have access to all other gateways as well. Therefore, obtaining actual meter readings is not possible with a single gateway being compromised.

- *Security Goal 4*: Since all the SMs use digital signatures for messages containing obfuscation information and measurements, the digital signature can be verified to confirm the identity of the message sender. In addition, since the messages are digitally signed, they cannot be modified without invalidating the signature, providing message integrity.

- *Security Goal 5*: Since all messages are timestamped and digitally signed, the timestamp can be checked to verify that the received message is for the current reading.

6.2 EXPERIMENTAL SETUP

The *scheduled obfuscation* approach is implemented under the widely used network simulator ns-3 [10], which has an implementation of IEEE 802.11s and LTE. Randomly connected AMI network topologies were created containing 25, 36, 49, 64 and 81 nodes in an area of size 1200m x 1200m. The area mimics the size of a neighborhood which uses a single gateway to communicate with the utility company.

The transmission range of each SM is set to 100m [34]. The underlying MAC protocol is IEEE 802.11g. TCP protocol is used to ensure reliability compared to UDP. The data frequency of the SMs is set to 10sec. [34].The simulation is run for 300 secs. Each run for 10 different topologies are tested and reported the average of these topologies for significance of the results.

For encryption, crypto++ library [38] is used. The Elliptic Curve Digital Signature Algorithm (ECDSA) is an approved signature algorithm for the US government use [39] and the Elliptic Curve Integrated Encryption Scheme (ECIES) is a well-known scheme having several standards [40]. ECDSA is used when only signature is required and ECIES is used when encryption and signature are required. In both cases, the ASN.1 secp128r1 standard curve with SHA1, having a key length of 256 bits is used.

6.3 BASELINES AND PERFORMANCE METRICS

Three baselines are considered for comparison with the *scheduled obfuscation* approach. The first baseline (represented as “baseline” in the graphs) sends meter readings in clear, providing no privacy. The second baseline (represented as “baseline sign”) provides authentication but does not provide any confidentiality in transmission and the utility provider has access to the fine-grained meter data. The third baseline (represented as “baseline sec” in the graphs) provides authentication as well as confidentiality, but the utility provider still has access to the fine-grained meter data.

For the *scheduled obfuscation*, the scheduled transmission in which a SM sends its obfuscated reading value to the gateway at every 10sec. even if the meter receives the obfuscation value earlier. The gateway sends obfuscation values to the SMs for the next reporting time, simultaneously. This is the default mechanism used in all experiments.

In analyzing the results, three metrics: throughput (i.e., the amount of data received at the transport layer by the gateway per second), data delay (i.e., the total time it takes for a reading to reach the gateway) and packet delivery ratio (PDR) (i.e., the ratio of packets that are delivered to the gateway compared to the number of packets sent by the SMs) are considered.

6.4 SIMULATION RESULTS

The results of the experiments conducted for comparing the performance of the *scheduled obfuscation* approach with those of the other three baseline approaches in between Smart Meters-Gateway and Gateway-Utility are shown in Figure 6.1, 6.2, 6.3, 6.4, 6.5 and 6.6. Each of the metrics is discussed separately below.

6.4.1 Smart Meters and Gateway

6.4.1.1 PDR

The PDR as shown in Figure 6.1 decreases slightly for all approaches as the number of SMs increases. This is due to increasing number of packets transmitted throughout the network from SMs, which increases the collisions. It is seen that PDR is the highest for the *baseline* because it has the smallest packet size compared to the other three approaches. The *baseline sign* and *baseline sec* approaches achieve a little different PDR, but mostly the *baseline sign* is slightly better because its packet size is smaller than that

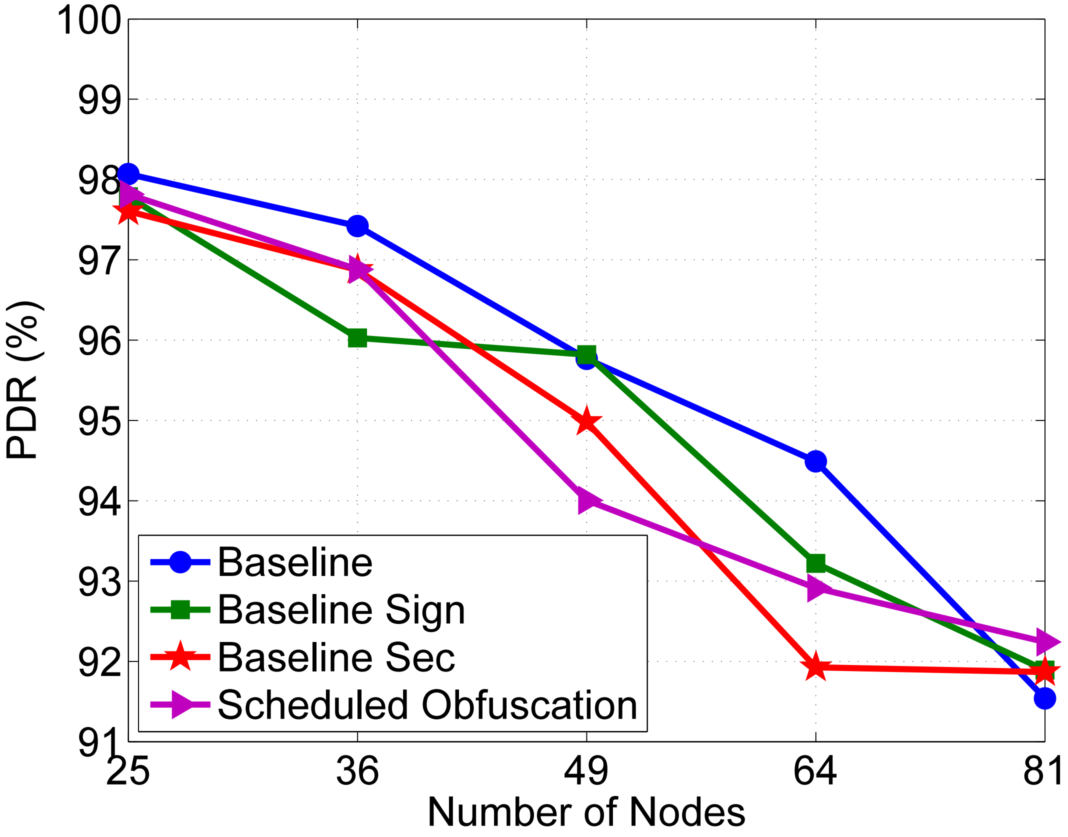

Figure 6.1. PDR at different number of nodes.

of the *baseline sec*. The *baseline sign* and the *scheduled obfuscation* approach generate 44-byte packets whereas the *baseline* and the *baseline sec* generate 12-byte and 65-byte packets, respectively.

Although the *scheduled obfuscation* approach and *baseline sign* generate the same size packets, the former mostly achieves lower PDR. This is due to the two-way packet traffic between the SMs and the gateway. Contrary to the other baselines, the the *scheduled obfuscation* approach needs to have obfuscation values coming from the gateway. The *scheduled obfuscation* is scheduled and waits for the end of 10secs to send all the SM readings. At the same time, the gateway starts sending obfuscation values for the next cycle. The obfuscation values coming from the gateway are signed and encrypted and thus the packet size becomes 65 bytes which is even greater than the readings' size (i.e., 44 bytes after signing). Therefore, the obfuscation values sent for the next data cycle may collide with some of the SM readings. This slightly increases the number of dropped packets and results in a decrease in the PDR.

6.4.1.2 Throughput

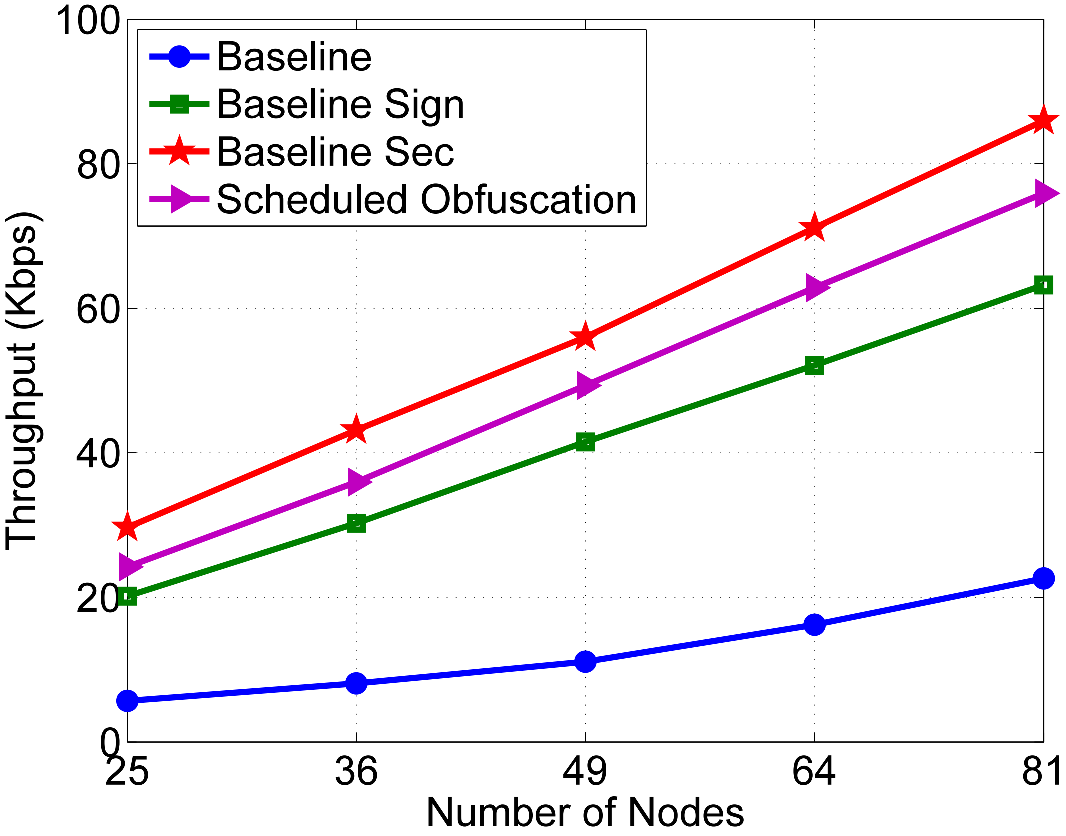

Figure 6.2. Throughput at different number of nodes.

As seen in Fig. 6.2, the throughput increases as the network size grows due to the contribution of more nodes as expected. Although the throughputs of the *baseline sign* and the *scheduled obfuscation* approach are similar as expected because of the almost near sending packet size, there seems to be a little difference between *scheduled obfuscation* approach and *baseline sign*. This is again due to the crossing of traffic when the obfuscation values are being sent from the gateways to the SMs and shortly after the readings are being sent from SMs to the gateway through the same paths. While these transmissions are happening one after another, there may still be some traffic in the network during the transmission of obfuscation values to the leaves (i.e., the nodes at the far end of the network) of the network topology. This can cause some interference and keep the channel busy at certain parts of the network which eventually causes the rate of increase in throughput to reduce slightly. However, overall these results indicate that there is no major adverse effect of the proposed distribution and obfuscation approach in terms of throughput.

It is also seen in Fig. 6.2 that the *baseline sign*, *the baseline sec* and the *scheduled obfuscation* approach have significantly higher throughput than the *baseline* has. This can be attributed to the sizes of the packets they send. Even if the *baseline* has the highest PDR, the sizes of the packets the other three approaches generate compensate this difference and cause higher throughput.

6.4.1.3 Delay

The impact of the approach on ETE delay is an important metric for some of the AMI applications such as demand and response. The delay of all of the baselines is almost close when the network size is smaller (i.e., up to 36 nodes), and it increases because of the increase in congestion cases as the network size grows. The *scheduled obfuscation* approach is expected to experience almost the same delay with the *baseline sign* for all of the number of nodes, but it demonstrates difference delays after 25 nodes.

There are 2 reasons. First reason is that *the baseline*, *the baseline sign* and *the baseline sec* can send their readings around the same time which more nodes become involved in message sending at the same time and thus channel access delay increases significantly due to heavy contention and interference. As for *scheduled obfuscation*, they cannot send their readings around the same time since the SMs are waiting for obfuscation values from the gateway. The obfuscation values reach the destinations at different times due to the topological structure of the network and SMs are scheduled to send their readings at the next sending time. Since ns-3 does not schedule the sending operation to an exact time value but schedules it so as to be performed after a given time interval, the SMs cannot be scheduled to exactly the same time for transmission. They are scheduled between the same seconds, but there are time lags (some milliseconds) between each scheduling. This apparently reduces the contention among the nodes for accessing the channel in the network and thus MAC layer delay is reduced.

The other reason is that computation time for encryption and signature process can be high when the *baseline sign* signs their readings and the *baseline sec* encrypts their readings. Therefore, delay time for the *baseline sign* and the *baseline sec* is higher than other approaches in Figure. 6.3

6.4.2 Gateway and Utility

6.4.2.1 PDR

The PDR as seen in Figure 6.4 decreases slightly for the *baseline sign*, the *baseline sec* and the *scheduled obfuscation* because the number of SMs increases. Increasing number of packets transmitted throughout the network from SMs augments the collisions. It is seen that PDR is constant for the *baseline* because it has the smallest packet size compared to the other three approaches and all collected reading values coming from gateway sends to utility one time with small packet size. After node 49, PDR increases slightly for the *baseline sign*, the *baseline sec* and the *scheduled obfuscation* because of packet segmentation. These three approaches' packet size are high because of signature, encryption and obfuscation process. If the data packet is larger than the maximum transmission unit supported by the network, packet segmentation divides a data packet into smaller units for transmission over the network. Therefore, the number of dropped packets decreases and result in a increase in the PDR.

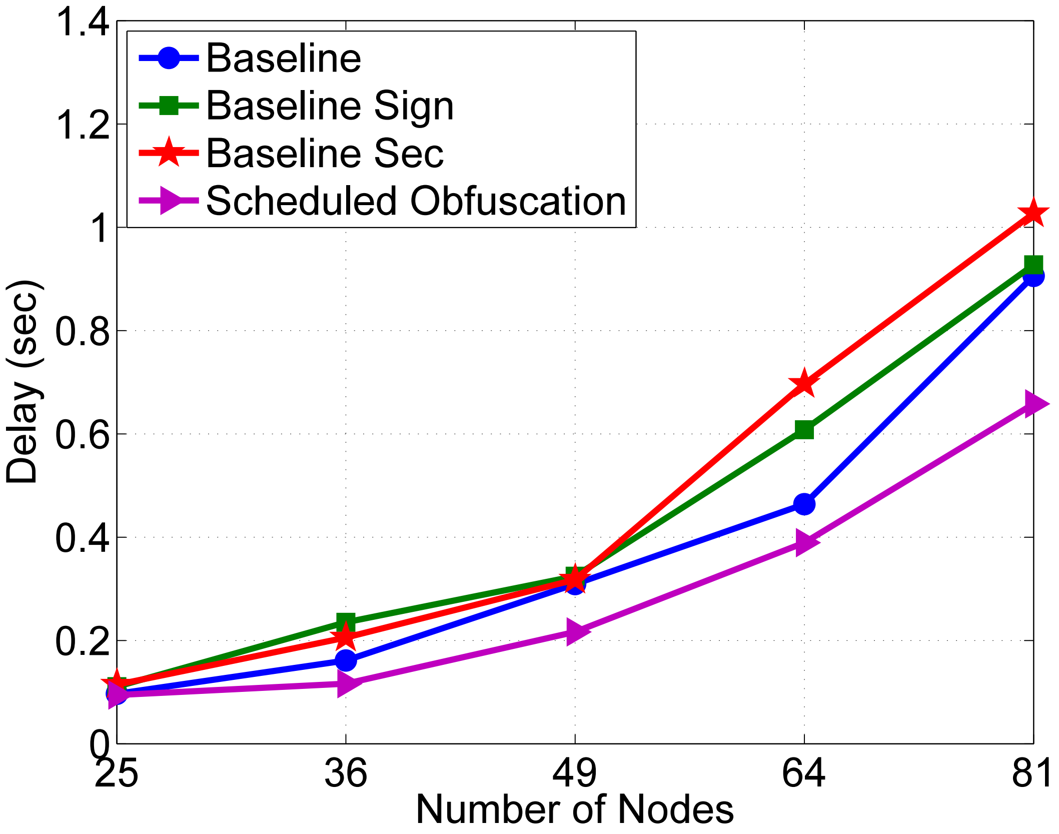

Figure 6.3. Delay at different number of nodes.

6.4.2.2 Throughput

As illustrated in Figure. 6.5, the throughput increases because the network size grows due to the contribution. The throughputs of the *baseline sign* and the *scheduled obfuscation* are same for all of the number of the number of nodes since the main reason is the packet size of all collected the obfuscated or signed values are being sent from gateway to utility are same for both of these approaches. In addition, these obfuscated or signed values are sent just one time so there is less crossing of traffic and less contention among the nodes for accessing the channel in the network. It is also seen that the *baseline sign*, *baseline sec* and *scheduled obfuscation* have higher than throughput than *baseline* has because of the size of the packets they send.

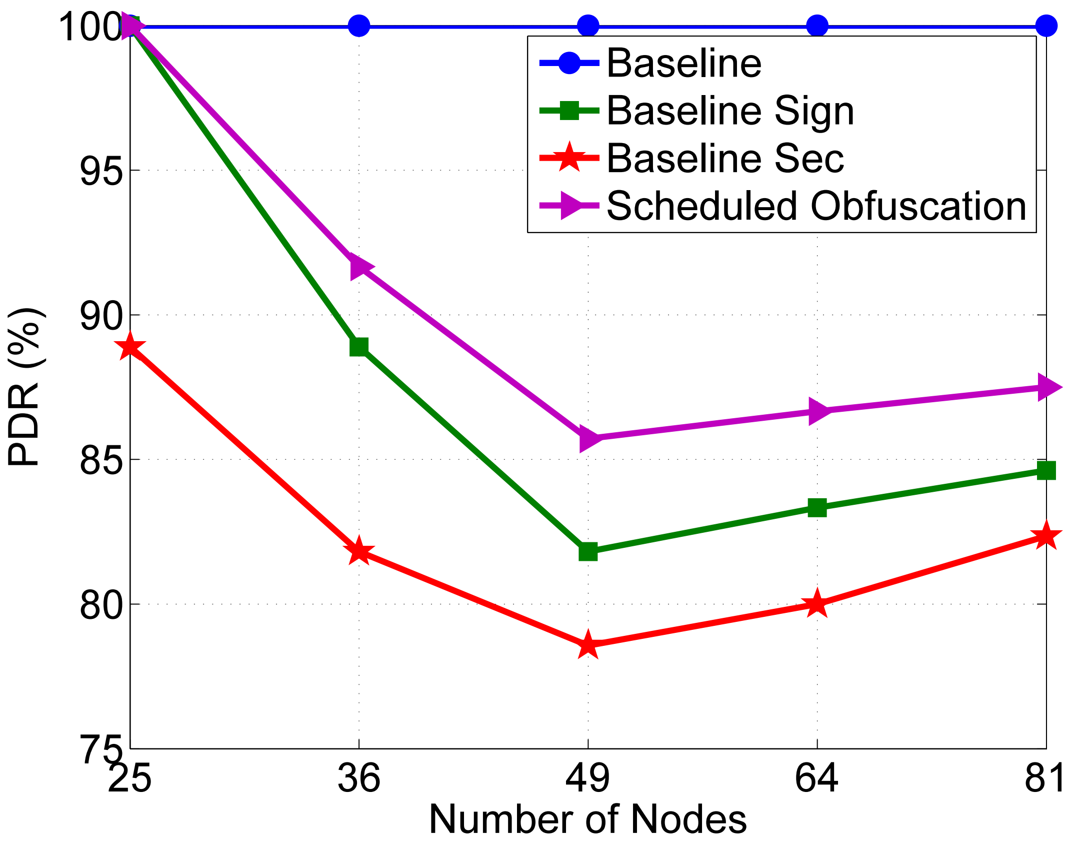

Figure 6.4. PDR at different number of nodes.

6.4.2.3 Delay

As illustrated in Figure 6.6, the delay of all approaches increases due to the increase in congestion cases as the network size grows. The *scheduled obfuscation* approach experiences the same delay with the *baseline sign* for all of the number of nodes since the *baseline sign* and the *scheduled obfuscation* generate 44-byte packets. In addition, there is not heavy crossing of traffic or contention among the nodes for accessing the channel in this network topology since all collected reading values that are signed or obfuscated, are sent just one time from the gateway to utility. It is shown in Figure 6.6 that *baseline sec* has significantly higher delay than the *baseline* due to the size of the packets they send. The *baseline sec* has the highest delay since *baseline sec* generate 65-byte packets. The

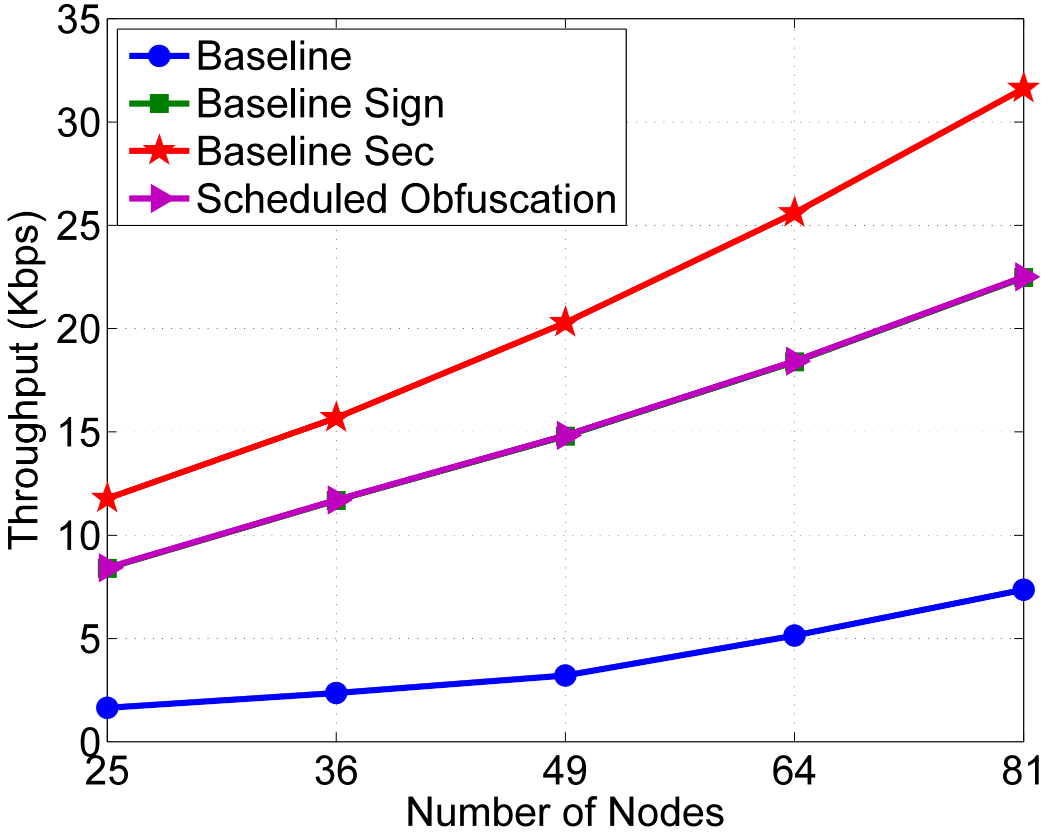

Figure 6.5. Throughput at different number of nodes.

baseline generate 12-byte packets. Therefore, the delay of *baseline* is the lowest.

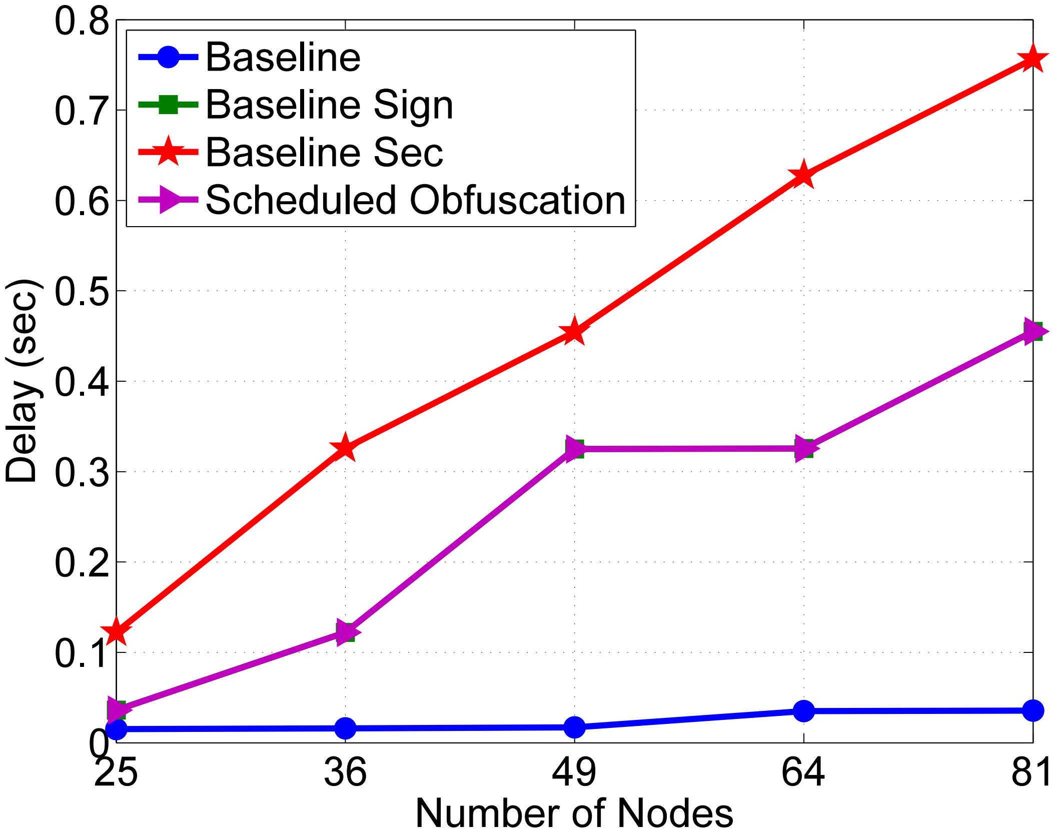

Figure 6.6. Delay at different number of nodes.

CHAPTER 7
CONCLUSION

In this thesis, the problem of user privacy preservation is tackled in SG that will also enable distribution state estimation. A data obfuscation mechanism is followed and proposed secure and efficient algorithms to distribute obfuscation values within an AMI and LTE network. Specifically, ECC is used for hiding and authenticating the obfuscation values that are distributed within the network. Using LTE EPC model between gateway and utility is to enable operators to deploy and operate one common packet core network for 3GPP radio access (LTE, 3G, and 2G) and non-3GPP radio access (HRPD, WLAN, and WiMAX). Multiple-gateway implementations are also considered for increased security. A protocol that utilizes LTE-Direct for exchanging of data among multiple gateways is proposed.

All the proposed approaches are implemented in ns-3 using LTE and a draft version of 802.11s for a 802.11s-based mesh network to assess their overhead. Simulation results showed that the obfuscation approaches are promising in terms of ETE delay without introducing additional overhead on PDR and throuhput compared to other existing approaches. The approaches are analyzed in terms of the security goals they provide and showed that they can ensure consumer privacy while still allowing state estimation and billing.

REFERENCES

[1] Wenye Wang, Yi Xu, and Mohit Khanna. Survey paper: A survey on the communication architectures in smart grid. *Comput. Netw.*, 55:3604–3629, October 2011.

[2] Nico Saputro, Kemal Akkaya, and Suleyman Uludag. A survey of routing protocols for smart grid communications. *Computer Networks*, 56(11):2742–2771, 2012.

[3] Klaus Kursawe, George Danezis, and Markulf Kohlweiss. Privacy-friendly aggregation for the smart-grid. In *Proceedings of the 11th international conference on Privacy enhancing technologies*, PETS'11, pages 175–191, Berlin, Heidelberg, 2011. Springer-Verlag.

[4] Nico Saputro and Kemal Akkaya. On preserving user privacy in smart grid advanced metering infrastructure applications. *Security and Communication Networks*, 7(1):206–220, 2014.

[5] Younghun Kim, E.C.-H. Ngai, and M.B. Srivastava. Cooperative state estimation for preserving privacy of user behaviors in smart grid. In *Smart Grid Communications (SmartGridComm), 2011 IEEE International Conference on*, pages 178 –183, oct. 2011.

[6] Jolly Parikh and Anuradha Basu. Lte advanced: The 4g mobile broadband technology. *spectrum*, 5(2.5):30, 2011.

[7] Navigant Research. Lte networks for smart grid applications. 2013.

[8] *http://www.gsacom.com/downloads/pdf/GSA_evolution_to_lte_report_310811.php4.*

[9] 3GPP TR 25.913 V8.0.0 (2008-12) Requirements for Evolved UTRA (E-UTRA) and Release 8 Evolved UTRAN (E-UTRAN).

[10] ns 3. ns-3: network simulator 3. Release 3.19, 2013.

[11] C. Efthymiou and G. Kalogridis. Smart grid privacy via anonymization of smart metering data. In *Smart Grid Communications (SmartGridComm), 2010 First IEEE International Conference on*, pages 238 –243, oct. 2010.

[12] G. Kalogridis, C. Efthymiou, S.Z. Denic, T.A. Lewis, and R. Cepeda. Privacy for smart meters: Towards undetectable appliance load signatures. In *Smart Grid Communications (SmartGridComm), 2010 First IEEE International Conference on*, pages 232 –237, oct. 2010.

[13] Ali Keyhani and Mohammad Albaijat. *Smart power grids 2011*. Springer, 2012.

[14] Sourav Mallick, SP Ghoshal, P Acharjee, SS Thakur, et al. Optimal static state estimation using hybrid particle swarm-differential evolution based optimization. *Energy and Power Engineering*, 5(04):670, 2013.

[15] I Dzafic, S Henselmeyer, and H-T Neisius. High performance state estimation for smart grid distribution network operation. In *Innovative Smart Grid Technologies (ISGT), 2011 IEEE PES*, pages 1–6. IEEE, 2011.

[16] Ahmad Abdel-Majeed and Martin Braun. Low voltage system state estimation using smart meters. In *Universities Power Engineering Conference (UPEC), 2012 47th International*, pages 1–6. IEEE, 2012.

[17] Ahmad Abdel-Majeed, Stefan Tenbohlen, Daniel Schollhorn, and Martin Braun. Development of state estimator for low voltage networks using smart meters measurement data. In *PowerTech (POWERTECH), 2013 IEEE Grenoble*, pages 1–6. IEEE, 2013.

[18] Stefania Sesia, Issam Toufik, and Matthew Baker. *LTE: the UMTS long term evolution*. Wiley Online Library, 2009.

[19] P. Lescuyer and T. Lucidarme. *Evolved Packet System (EPS): The LTE and SAE Information of 3G UMTS*. Wiley & Sons Ltd., January 2008.

[20] Hongkun Yang, Fengyuan Ren, Chuang Lin, and Jiao Zhang. Frequency-domain packet scheduling for 3gpp lte uplink. In *INFOCOM, 2010 Proceedings IEEE*, pages 1–9. IEEE, 2010.

[21] Suk-Bok Lee, Sayantan Choudhury, Ahmad Khoshnevis, Shugong Xu, and Songwu Lu. Downlink mimo with frequency-domain packet scheduling for 3gpp lte. In *INFOCOM 2009, IEEE*, pages 1269–1277. IEEE, 2009.

[22] Suk-Bok Lee, Ioannis Pefkianakis, Adam Meyerson, Shugong Xu, and Songwu Lu. Proportional fair frequency-domain packet scheduling for 3gpp lte uplink. In *INFOCOM 2009, IEEE*, pages 2611–2615. IEEE, 2009.

[23] US Department of Energy. Communications requirements of smart grid technologies. October 5, 2010.

[24] Jason Brown and Jamil Y Khan. Performance analysis of an lte tdd based smart grid communications network for uplink biased traffic. In *Globecom Workshops (GC Wkshps), 2012 IEEE*, pages 1502–1507. IEEE, 2012.

[25] Jason Brown and Jamil Y Khan. Performance comparison of lte fdd and tdd based smart grid communications networks for uplink biased traffic. In *Smart Grid Communications (SmartGridComm), 2012 IEEE Third International Conference on*, pages 276–281. IEEE, 2012.

[26] Hu Dan Zhao Feng, Liu Jianming and Zhang Yuexia. Study on the application of advanced broadband wireless mobile communication technology in smart grid. In *2010 International Conference on Power System Technology (POWERCON)*, 2010.

[27] Peng Cheng, Li Wang, Bin Zhen, and Shihua Wang. Feasibility study of applying lte to smart grid. In *Smart Grid Modeling and Simulation (SGMS), 2011 IEEE First International Workshop on*, pages 108–113. IEEE, 2011.

[28] 3GPP TS 23.401 V10.2.0:General Packet Radio Service (GPRS) enchancements for Evolved Universal Terrestrial Radio Access Network (E-UTRAN) access.

[29] The status of ieee 802.11s standard.

[30] A. Abdel-Majeed and M. Braun. Low voltage system state estimation using smart meters. In *Universities Power Engineering Conference (UPEC), 2012 47th International*, pages 1–6, Sept 2012.

[31] Alcatel-Lucent White Paper Introduction to Evolved Packet Core. 2009.

[32] Ns-3 lte module. *https://www.nsnam.org/docs/models/html/lte.html*.

[33] Alcatel-Lucent White Paper The LTE Network Architecture. 2009.

[34] Nico Saputro and Kemal Akkaya. PARP-S: A secure piggybacking-based ARP for IEEE 802.11s-based smart grid AMI networks. *Computer Communications*, 2014.

[35] Dimitris Tsolkas, Eirini Liotou, Nikos Passas, and Lazaros Merakos. Lte-a access, core, and protocol architecture for d2d communication. In *Smart Device to Smart Device Communication*, pages 23–40. Springer, 2014.

[36] Gwenael Poitau, Benoit Pelletier, Ghyslain Pelletier, and Diana Pani. A combined PUSH/PULL service discovery model for LTE direct. In *Vehicular Technology Conference (VTC Fall), 2014 IEEE 80th*, pages 1–5. IEEE, 2014.

[37] 3gpp technical report 23.703 v0.5, rel 12 in technical specification group services and system aspects. *http://www.3gpp.org/*, 2013.

[38] cryptopp. *http://www.cryptopp.com/*.

[39] NIST. Fips 186-3: Digital signature standard. 2009.

[40] Vctor Gayoso Martnez, Luis Hernndez Encinas, and Carmen Snchez vila. A survey of the elliptic curve integrated encryption scheme. *Journal of Computer Science and Engineering*, 2(2), 2010.

[41] Zhuo Li, Qilian Liang, and Xiuzhen Cheng. Emerging wifi direct technique in home area networks for smart grid: Power consumption and outage performance. *Ad Hoc Networks*, 22:61–68, 2014.

VITA

Graduate School
Southern Illinois University

Ozan Cakmak

ozancakmak@siu.edu

Izmir University of Economics, Izmir, Turkey
Bachelor of Science, Software Engineering, June 2008

Thesis Title:
Privacy Preservation In a Hybrid Multi Mesh-LTE AMI Network for Smart Grid

Major Professor: Dr. Kemal Akkaya

Publications:

[C1] **Ozan Cakmak**, Abe Kazemzadeh, Dogan Can, Serdar Yildirim And Shrikanth S. Narayanan, *"Root-Word Analysis Of Turkish Emotional Language"*, Proceedings Of 4th International Workshop On Corpora For Research On Emotion Sentiment And Social Signals (LREC), Istanbul, Turkey, 2012.

[C2] **Ozan Cakmak**, Abe Kazemzadeh, Serdar Yildirim And Shri Narayanan, *"Using Interval Type-2 Fuzzy Logic To Analyze Turkish Emotion Words"*, Proceedings Of APSIPA Annual Summit And Conference 2012, Hollywood, CA, Dec 2012.