SECURITY AND PRIVACY IN CRITICAL INFRASTRUCTURE CYBER-PHYSICAL SYSTEMS: RECENT CHALLENGES AND SOLUTIONS

A Dissertation by

Arash Boustani

Master of Science, Islamic Azad University, 2009

Bachelor of Science, Islamic Azad University, 2005

Submitted to the Department of Electrical Engineering and Computer Science
and the faculty of the Graduate School of
Wichita State University
in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy

December 2016

SECURITY AND PRIVACY IN CRITICAL INFRASTRUCTURE CYBER-PHYSICAL SYSTEMS: RECENT CHALLENGES AND SOLUTIONS

The following faculty members have examined the final copy of this dissertation for form and content, and recommend that it be accepted in partial fulfillment of the requirement for the degree of Doctor of Philosophy with a major in Electrical Engineering.

Murtuza Jadliwala, Committee Chair

Vinod Namboodiri , Committee Member

Visvakumar Aravinthan, Committee Member

Kaushik Sinha , Committee Member

Davood Askari, Committee Member

Accepted for the College of Engineering

Royce Bowden, Dean

Accepted for the Graduate School

Dennis Livesay, Dean

DEDICATION

I would like to dedicate this dissertation to her constant, unconditional support and love for all past years, to my Mom.

Abstract

Cyber-physical systems (CPS) are engineered systems that are built from, and depend upon, the seamless integration of computational algorithms and physical components. CPS are characterized by a strong interconnection of various computing systems (and algorithms) that are used to control, monitor, and interact with physical processes, thus improving the overall capability, adaptability, scalability, resiliency, safety, security, and usability of the associated engineered system. CPS technology has also enabled several critical infrastruc-ture applications, such as smart grid and renewable energy systems, biomedical and health-care, next-generation transportation, industrial automation, and defense systems. However, such Critical-Infrastructure CPS (CI-CPS) are extremely vulnerable to sophisticated cyber-attacks due to their interconnected nature. The consequences of malicious attacks may range from minor variation in performance to absolute inability to control the system, which may lead to catastrophic results for both the system operators and users.

Ensuring security of system components, privacy of system or user data, and availability of services (provided by the system), are some of the most vital requirements of a CI-CPS. Previous works in the literature proposed many solutions to improve security, privacy and availability of CI-CPS. However, several critical open problems in these areas remain unaddressed. In this direction, the first part of this dissertation addresses the problem of securing location discovery of wireless and mobile components (of a CI-CPS) by proposing a novel spread-spectrum-based approach to eliminate incorrect localization data injected by malicious location anchors. The second part of this dissertation presents a framework to increase the capacity (and consequently availability) of existing wireless networks, by utilizing a secondary cognitive radio network based approach. The third part and final part of this dissertation presents a novel framework to enable privacy-preserving smart meter data reporting in a smart grid CI-CPS, with a minimal impact on data utility. The efficiency and effectiveness of the proposed solutions are demonstrated by means of analytical evaluations and empirical results. The outcomes of this dissertation will further our current knowledge and understanding of the security, privacy and availability issues in this upcoming and nationally important area of CI-CPS.

TABLE OF CONTENTS

Chapter **Page**

TABLE OF CONTENTS (continued)

Chapter **Page**

LIST OF FIGURES

LIST OF FIGURES (continued)

Figure **Page**

LIST OF FIGURES (continued)

Figure **Page**

LIST OF TABLES

CHAPTER 1
WHAT ARE CYBER PHYSICAL SYSTEMS AND WHY ARE SECURITY/PRIVACY ISSUES IN THESE SYSTEMS IMPORTANT?

Cyber-Physical Systems (CPS) are systems that are built from, and depend upon, the seamless integration of computational algorithms and physical components. Such systems are, in most cases, tightly integrated with the Internet and its users, and in some cases, with a private communication networks. CPSs are characterized by a strong interconnection of various computing elements (and algorithms) that are used to control, monitor, and interact with physical processes, hence, improving the overall capability, adaptability, scalability, resiliency, safety, security, and usability of the associated engineered system. Examples of CPS include autonomous automobile systems, automatic pilot avionics, smart power grid, medical monitoring, robotics systems, and process control systems.

As a result of the development of the aforementioned CPS technology, critical infrastructure applications, such as smart grid and renewable energy systems, biomedical and healthcare, next-generation transportation, industrial automation, and defense systems have become smarter and interconnected. In this work, we refer to these critical systems as Critical Infrastructure CPS (CI-CPS). Due to the critical nature of CI-CPS, all aspects of system and data security and privacy must be studied in detail. Such systems can be vulnerable to a variety of cyber-attacks due to their interconnected nature. The consequences of malicious attacks may range from minor variation in performance to absolute inability to control the system, which may lead to catastrophic results for both the system operators and users. Hence, studying system and data security, privacy, integrity, and, availability is of paramount importance.

Given the use cases of CI-CPS in monitoring, managing, and controlling critical infrastructure, understanding and protecting the following aspects of system and user privacy and security is of utmost importance. i) Confidentiality: the data transmitted in such systems

should be accessible only by authorized parties. ii) Integrity: mechanisms should be developed to ensure that the critical data sent and received in these systems is not modified or tampered. iii) Availability: data and systems in CI-CPS must be available at all times and at an agreed service level in types of situations, i.e., ranging from normal through highly abnormal (such as in the case of disasters).

An example or use case of such a CI-CPS is the upcoming paradigm of smart power grid or smart grid. In order to provide power reliably and efficiently to consumers, information and communication technologies are being merged into the traditional power grid. A Smart Grid is an electrical grid that leverages communication technologies and information processing to gather, process, and act on collected information to improve reliability, efficiency, economics, and sustainability of the power grid in generation, transmission, and distribution. This two-way communication system enables the utility companies to remotely gather power consumption data from the users at short time intervals. This highly-granular power usage data collected from the users' *smart meters* will equip the utility companies with advanced features such as real time monitoring, fault-detection, self-healing, load balancing, demand-response, demand dispatch, and peak-shaving. The deployment of smart grid will save energy, enable the use of dynamic pricing schemes, integrate renewable resources and electric vehicles into the power grid, and provide greener and cleaner energy [21]. Despite the tremendous promise, security and privacy issues continue to plague the effective operation and adoption of this smart grid technology.In Chapters 3 and 4, we outline a few current privacy and availability related issues in smart grid systems and propose novel approaches in order to overcome them.

Components within a CI-CPS may contain computational entities that are mobile and communicate with each other in a wireless fashion, thus, forming a mobile wireless network. Localization or location discovery is an important protocol in such networks. Range-based localization, where mobile nodes compute distances to static anchors to determine their own location by using trilateration and multilateration techniques, is a well-known concept

in mobile wireless networks extensively used in CI-CPS such as wireless e-health systems. However, secure range-based location discovery in the presence of cheating or untrustworthy anchors (or beacon nodes) is an important, and still open, problem in mobile wireless networks. In order to overcome this problem, earlier research efforts have mostly followed three solution directions: (i) efficient detection and elimination of cheating anchors, (ii) range-based localization without using anchors and (iii) range-based localization in the presence of cheating anchors. On the one hand, despite results that guarantee an upper bound on the localization error [110], the latter two approaches have not fared very well in terms of localization accuracy. On the other hand, most of the malicious anchor detection techniques are based on consensus building or statistical estimation, and are rather restrictive with high false-positive or false-negative rate. Similarly, the issue of elimination of cheating anchors from consideration (once detected) is non-trivial and has not been clearly addressed in the literature. In Chapter 2, we present a novel and deterministic secure localization strategy to overcome the cheating effect of malicious anchors. Our technique employs a "*request confusion*" strategy in order to detect malicious or cheating anchors and a "*DSSS or CDMA-based jamming strategy*" in order to eliminate the (effect of) cheating anchors. By means of simulation experiments, we validate the performance of our secure localization technique under various adversarial strengths and network parameters.

As discussed earlier, the power industry is moving towards the next generation power grid, i.e., the smart grid, by taking advantage of information and communication technologies. This information-based power grid is expected to change the way electricity is generated, distributed, and transmitted to the consumers by enhancing the reliability, efficiency, sustainability, and economics of the grid. However, due to the high volume and high granularity of the data generated by smart electricity meters, careful planning and management of this communication network is necessary. Given the large scale future deployment of smart grid, utility companies face possible network capacity constraints. Due to this scarcity, an efficient spectrum allocation is often difficult, thus resulting in low overall bandwidth utiliza-

tion in smart grid networks. Hence, an efficient utilization of this communication network should be studied. *Cognitive Radio Networks (CRN)* enable *Secondary Users (SU)* to coexist with existing network infrastructures. *Cognitive Smart Grid Networks (CSGN)* use CRN to optimize resource allocation in SGNs. However, efficient utilization of available channel bandwidth by SUs, without interfering with the *Primary Users (PU)*, remains an important open problem in CSGN. In Chapter 3, we focus on CSGN as the *Secondary Network (SN)*, coexisting with a *Primary Network*, and outlining the applicability of Code Division Multiple Access for overcoming the low *Number of SUs (NSU)* in SN. We propose a novel resource allocation technique to improve NSU in CSGN by using a specific kind of *Orthogonal Chip Sequence (OCS)* allocation in spread spectrum communications for SU transmissions. By means of extensive simulations and analysis, we show that our technique improves NSU on SN (or CSGNs) significantly.

The deployment of the smart grid also introduces privacy-related concerns as it can gather highly-granular power consumption data which can reveal sensitive private information about the lives of consumers. Perturbing the actual energy usage data before sharing it with the energy company is a well-known solution to overcome privacy issues associated with smart meters. The degree of correlation between the actual energy usage data and the perturbed data produced by the perturbation technique typically characterizes the trade-off between the privacy requirement (of the customer) and data utility or data usefulness requirement (of the energy company). Our main goal in Chapter 4 is to propose a mechanism to minimize this trade-off, i.e., provide both reasonable levels of privacy protection, as well as, data-utility. We work towards this goal by proposing a novel two-level energy consumption prediction scheme. The first-level prediction at the household level is performed by each SM, and the predicted energy consumption pattern, instead of the actual energy usage data, is reported to a cluster head (CH) or a neighborhood aggregator. Then, a second-level prediction at the neighborhood level is done by the CH which predicts the energy spikes in the neighborhood or cluster and shares it with the EC. Our two-level prediction mechanism is designed such that

it preserves the correlation between the predicted and actual energy consumption patterns at the cluster or neighborhood level and removes this correlation in the predicted data communicated by each SM to the CH. This maintains the utility or usefulness of the cluster- or neighborhood-level energy consumption data communicated to the EC while preserving the privacy of the household-level energy consumption data against the CH (and thus the EC). We further implement and evaluate our two-level prediction mechanism using real smart meter data. Our evaluation results show that our proposed mechanism is successful in hiding private consumption patterns at the household-level while still being able to accurately predict energy consumption at the neighborhood-level.

CHAPTER 2
LOCJAM: A NOVEL JAMMING-BASED APPROACH TO SECURE LOCALIZATION IN WIRELESS NETWORKS

Distributed *localization* or *location discovery* in wireless networks is the problem of determining the location (in a distributed fashion) of a (mobile) device in the network with respect to some local or global coordinate system. Localization protocols in wireless networks can be categorized into two broad types: i) *range-based* and ii) *range-free* protocols [99]. In range-based techniques, a node computes its location by first estimating distances to neighboring nodes, whereas range-free techniques, typically, do not involve any distance estimation by the target node. Range-based techniques can be further classified as (a) *anchor or beacon-based* and (b) *anchor-free* protocols. Anchor-based algorithms such as [250, 14, 185, 28, 175, 228, 32, 144], among others, need special beacon or anchor nodes that are strategically placed in the network and know their own location (for example, by means of GPS). The mobile target node first estimates its distance to a set of neighboring beacon or anchor nodes by using well-known techniques such as *Received Signal Strength Indicator (RSSI)* [159], *Time of Arrival (ToA)* [170], and *Time Difference of Arrival (TDoA)* [258]. The target node then applies constraint satisfaction or optimization techniques, such as, trilateration or multilateration, in order to compute its location. A two-dimensional anchor-based localization process by trilatering distance estimates to three anchor nodes is depicted in Figure 2.1(a). Anchor-free schemes do not involve specifically marked anchor nodes.

Although anchor-based schemes are popular and generally perform well, a majority of these techniques operate under the assumption that anchor nodes behave honestly during the localization process. This assumption is not valid in non-trustworthy wireless environments where anchor nodes could cheat by manipulating the distance estimation process, as shown in Figure 2.1(b), and thus affecting the overall accuracy of the location estimated by the target node. Numerous proposals for overcoming the problem of cheating in range-based

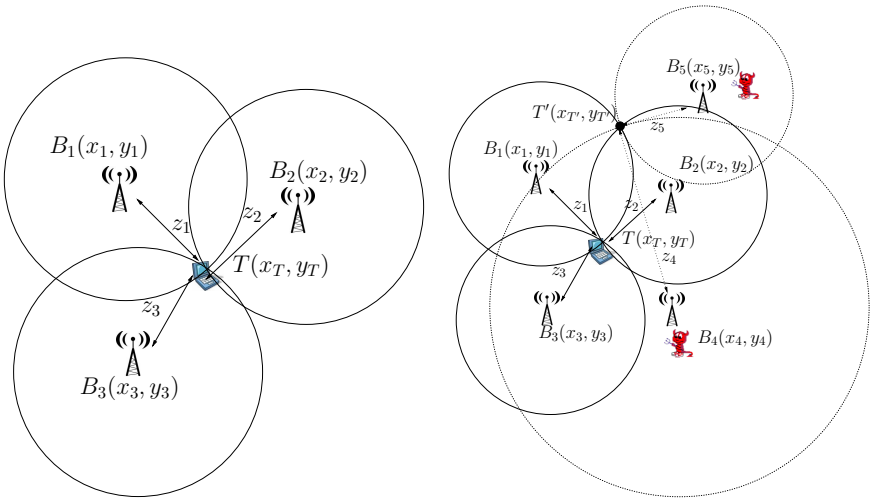


Figure 2.1: Distance-based (range-based) localization (a) Trilateration (b) Cheating anchors.

localization protocols exist in the literature [139, 142, 143, 183, 242, 166, 204, 70, 74, 110, 210]. These proposals have primarily followed one of the following two approaches. The first approach is to localize in the presence of cheating anchor nodes and securely verify that the determined location is within some maximum error bound. The second approach calls for efficiently detecting and eliminating measurements emanating from cheating anchors before location determination. Localization schemes following the first approach often need to satisfy certain necessary network conditions (e.g., in terms of the total number of malicious anchors) and are constrained by the resulting large localization errors. Localization schemes following the second approach suffer from the non-triviality of the detection and elimination process in a distributed networking environment.

We are motivated by the fact that radio signal jamming has traditionally always been considered as an adversarial tool that is used for disrupting network protocols. In this work, we would like to follow a reverse ideology and use jamming in order to protect network protocols such as location discovery. In this section, we propose a fresh approach to overcome the problem of cheating anchors in distributed range-based localization protocols which implements an asynchronous "request-confusion" mechanism for detecting cheating anchor

nodes and a Direct-Sequence Spread Spectrum (DSSS) or Code Division Multiple Access (CDMA) based jamming mechanism for eliminating range measurements from cheating anchors during location determination by the target node[1]. Distributed Spread Spectrum or Code Division Multiple Access communications by using Orthogonal Chip Sequences (OCS) have several advantages in wireless communications, including higher number of simultaneous transmissions and low interference and data collision [40]. A few CDMA-based localization approaches using OCSs have been proposed in the literature, but these efforts do not address the secure localization problem and have considered a non-hostile networking environment in their work [254, 30]. To the best of our knowledge, this is the first proposal that considers a CDMA-based jamming strategy in order to secure anchor-based distributed localization in wireless networks.

The rest of the chapter is organized as follows. Related work in the literature and background on securing anchor-based localization is outlined in Section 2.2. The network and adversary model assumed in this work is presented in Section 2.3. Our proposed secure localization protocol using CDMA-based jamming is outlined in Section 2.4 and simulation results are discussed in Section 2.5. We conclude the section with a summary of contributions and results in Section 2.6.

2.1 Background and Related Work

In this section, we survey some earlier research efforts towards securing distance-based localization schemes.

2.1.1 Detection and Elimination of Malicious Anchors

The first approach to secure distance-based localization is to detect cheating anchors and eliminate them from consideration. Liu et al. [142] propose a technique for eliminating malicious anchor data, called *attack-resistant Minimum Mean Square Estimation (MMSE)*,

[1]The content of this chapter has appeared in [20].

which leverages on the fact that malicious location references are usually inconsistent with the benign ones. Sastry et al. [210] propose the *Echo* location verification protocol to securely verify location claims by computing the relative distance between the prover and the verifier node using the time of propagation of ultrasound signals. Čapkun et al. [242] outline various attacks on node localization and propose mechanisms such as authenticated distance estimation, authenticated distance bounding, verifiable trilateration and verifiable time difference of arrival, in order to detect cheating anchors. Pires et al. [183] propose protocols to detect malicious nodes in distance-based localization approaches by detecting message transmissions whose signal strength is incompatible with its originator's geographical position. In another similar work by Liu et al. [143], the authors propose techniques to detect malicious anchors by employing special *detector anchors*.

2.1.2 Secure Localization in the presence of Malicious Anchors

The second approach is to design techniques that are robust enough to tolerate the cheating effect of malicious anchors. Priyantha et al. [185] develop the *CRICKET* system that eliminates the dependence on beacon nodes by using communication hops to estimate the network's global layout, and then apply force-based relaxation to optimize this layout. Li et al. [139] utilize statistical methods, such as *adaptive least squares* and *least median squares*, in order to make anchor-based localization attack-tolerant. Alternatively, Doherty et al. [58] outline a secure range-based localization method that employs convex optimization on a set of connectivity constraints. Liu et al. [142] design an intelligent *voting-based scheme* for resisting cheating by anchor nodes during distributed anchor-based localization. In another approach, Yi et al. [217] and Ji et al. [116] apply efficient data analysis techniques such as *Multi-Dimensional Scaling (MDS)* using connectivity information and distances between neighboring nodes to infer target locations. Fang et al. [69] use *Maximum Likelihood Estimation (MLE)* in order to estimate the most probable node location, given a set of neighborhood observations. Lazos et al. [135] present a hybrid secure localization approach,

called RObust Position Estimation (ROPE), which provides robust location computation and verification without centralized management and vulnerability to jamming (from malicious nodes). Misra et al. [166] propose a convex optimization based scheme to secure the distance-based localization process by applying Barrier's method in order to solve the optimization problem. Recently, Jadliwala et al. [110] proved the necessary and sufficient conditions for secure distance-based localization in the presence of cheating anchors and defined a class of algorithms that bound the localization error under these conditions.

2.1.3 Localization using Coding Theory

Concepts from coding theory have also been used to secure distributed range-based localization. For example, Ray et al. [196] propose a framework for providing robust location detection in wireless sensor networks using the theory of *Identifying Codes (ID-Codes)*. In this framework, high powered transmitters are fitted in such a way that each localizable point on the terrain is covered by a unique set of transmitters. Each node localizes itself by mapping the set of neighborhood transmitters to the corresponding location. Similarly, Yedavalli et al. [262] have used the theory of *Error Correcting Codes (ECC)* for robust localization in sensor networks. For each localizable point, the authors used distances from a fixed set of neighboring nodes to that point as a "codeword" for that point such that the "distance" between any two codewords is fixed. Thus, any cheating behavior by the participating nodes can result in an illegal codeword and can be detected and corrected. Contrary to this, in our work, we use orthogonal codes or chips for only eliminating cheating nodes, and not for detecting cheating. Cao et al. [30] outline an OCS and CDMA based technique for mobile location discovery in Line Of Sight (LOS) and Non-Line Of Sight (NLOS) scenarios. In this technique, all anchors are assigned identifiers by using a set of orthogonal codes that are broadcast periodically and synchronously. The mobile target detects the three strongest broadcast signals and estimates its location by calculating the Time Difference of Arrival (TDoA) with respect to these anchors. The authors showed that the use of OCS

for localization helps to cancel the interference at the mobile target caused by simultaneous transmission of the anchors. However, they do not address any security issues related to cheating anchors.

2.1.4 Discussion

Malicious node detection and elimination strategies, as discussed in Section 2.1.1, take into account the inconsistency (caused by cheating behavior) in the measurement of a particular network parameter in order to detect cheating anchors. One shortcoming is that the process of elimination of malicious anchors, once detected, is not clearly defined in most of these approaches. Others [142] propose only passive approaches for detecting and eliminating malicious anchors, for example, collaborative voting in order to blacklist malicious anchors. Although these passive approaches are intuitive and easy to implement, they can be easily circumvented. For example, cheating anchors can regularly change identifiers in order to avoid detection and/or elimination during the localization process. Cheating anchors could also deploy advanced hardware such as sectored antennas in order to avoid any collaborative passive detection mechanism. Collaborative passive detection mechanisms also suffer from an inherent weakness that requires a majority of the honest anchors to be able to detect and verify the cheating behavior of the malicious anchors.

Secure localization schemes discussed in Section 2.1.2 attempt to improve the robustness of distance-based localization procedure by minimizing the effect of inconsistent or erroneous localization data by cheating anchors. Some shortcomings of these solutions include complexity, relatively higher localization errors and/or requirement of specialized hardware.

In this work, we overcome the problems discussed above by following a more *active* approach to detection and elimination of cheating anchors in distributed distance-based localization protocols. In our approach, we employ jamming as a security tool, as opposed to its typical utility as an adversarial tool. Our approach is non-collaborative, and thus does not require a consensus building phase among honest nodes for eliminating cheating

beacons. Only a single honest node is required to eliminate the cheating effect of a malicious anchor. Also, as our approach actively eliminates malicious ranging data, the target node does not have to verify and eliminate these, thus improving the overall performance of the localization process.

2.2 Network Configuration

2.2.1 Network and Communication Model

The network consists of a *mobile* device MT, also referred to as the *mobile target node*, moving over an application area. MT wants to estimate its own location by using distance estimates to a set of *neighboring* (and stationary) anchor nodes who know their own location. In practice, there can be multiple target nodes, but we currently assume a single target node in order to simplify the current exposition. The mobility of the target node is application dependent and we only consider the movement of the target node over the application area. Without loss of generality, we assume that MT is momentarily static during the localization process. Deployed over the application area, are a fixed number (specifically, n) of *stationary* anchor nodes that know their own location and can assist the target node in its location estimation. Let these nodes be denoted as $B_1, \ldots, B_n$. For simplicity, assume that the locations of the target node MT and the anchor nodes can be expressed in the two-dimensional coordinate system as a vector (x, y) where, $x, y \in \mathbb{R}$. Each of the anchor nodes and the MT possesses an omni-directional radio transceiver.

All the anchor nodes in the network are assumed to be synchronized with each other whereas the MT communicates with the anchors in an asynchronous fashion. All communication takes place over two separate channels. The first channel is a CSMA-based control channel for sending and receiving certain control messages and the second channel is a CDMA-based data channel. Packets sent and received over this data channel are used by the target node MT for estimating distances to the corresponding anchor nodes. The chip sequences or OCSs for CDMA-based data transmission are generated using the Golay

[40] code generation algorithm. Various other algorithms for OCS generation in wireless networks also exist, for example, Walsh Hadamard, OVSF, Kasami, etc. [24]. The most important characteristics of OCSs that should be taken into consideration are auto/cross correlation, length of the generated OCSs versus the number of possible OCSs, error correction and fault tolerance. In some algorithms, for instance Walsh Hadamard, only the cross correlation feature exists while in certain others, all features can be observed. One of the advantages of OCSs generated by Golay is that the number of zeros and ones are equal in all chip sequences. Golay OCSs are simple to generate recursively, as shown by the Golay matrix representation in Eqn. 3.1.

$$C_L = \begin{bmatrix} C_{\frac{L}{2}} & \bar{C}_{\frac{L}{2}} \\ C_{\frac{L}{2}} & -\bar{C}_{\frac{L}{2}} \end{bmatrix} \tag{2.1}$$

$$\text{where, } C_L = \begin{bmatrix} A_L & B_L \end{bmatrix}, \bar{C}_L = \begin{bmatrix} A_L & -B_L \end{bmatrix} \text{ and } C_1 = 1$$

In Eqn. 3.1, $L = 2^M$ is the total number of available OCSs, where $M \geq 1$ is the number of bits in each OCS.

In our network model, we will use these chip sequences not only in the physical layer for CDMA data transmission, but also in the data link layer as an identifier (ID) for the corresponding anchor node. We assume that anchor nodes would be tessellated (discussed in details in the following section) or divided into groups. Time can be divided into random periods denoted by a random variable ψ. During each period, each group of anchors randomly choose a subset of available OCSs for use in that period. The subset selected by each anchor group consists of chip sequence with similar bit pattern and low distance. Such a group of *similar* chip sequences is referred to as a *flock* in CDMA literature [40]. A group of anchors using the same flock(s) of OCSs is called a *Grid Cell (GC)*. As the number of available OCSs is limited, flocks are reused throughout the network. A group of GCs in which the flocks are not reused form a *Cluster*. Each anchor will use the OCS uniquely assigned to it in the

time frame ψ in order to transmit data to the MT on the CDMA data channel. It should be noted that it is possible for multiple nodes to use the same OCS for data transmission in different parts of the network, henceforth referred to as *Code Reuse Factor (CRF)*. A CRF of r indicates that a total of r adjacent GCs in a cluster use different flocks of OCS. This is possible if the OCS generation scheme generates OCSs that fall into r categories, $\frac{1}{r}$ of which will be assigned to each GC in a cluster of r adjacent GCs. The OCSs used by an anchor group will be changed or refreshed after time ψ by the group head, which can be appointed as discussed ahead.

Honest anchor nodes are pre-configured with appropriate message authentication and encryption mechanisms for secure communications amongst each other. All honest anchor nodes in the same GC send and receive data signed using a group signature [39, 200]. Hence, each node is able to authenticate the source of any incoming message as being from the same GC or not. We also assume that during each time period, the table of valid OCSs (for that time duration) is exchanged among anchors in a distributed fashion. This can also be accomplished by a group head that is selected in each time period for each GC by using an appropriate group-head election algorithm. For a particular GC, an elected group head during a time period k, denoted by p_{ψ_k}, is also responsible for identifying a flock of OCS that does not conflict with the adjacent GCs. Group head selection is rotated within the GC for both security and energy efficiency reasons. From all possible OCSs in the flock, the group head randomly choose a portion of valid OCSs for the GC (as a function of the number of anchors in that GC) and broadcast the list of valid OCSs to all other anchors in the GC. Let $F_g(\psi_k)$ be the subset of the OCS flock used by a GC g during the time period k. These OCS advertisement messages are encrypted by using appropriate symmetric encryption algorithms. Standard distributed secure key-exchange and symmetric-key cryptographic algorithms can be used for this purpose.

Let us provide more details on the anchor communications over the CDMA data channel. The main concept of CDMA is to spread an information signal with bandwidth δ_s over a

larger bandwidth δ, where $\delta \gg \delta_s$ and $\frac{\delta}{\delta_s}$ is the processing gain. This is achieved by encoding each data symbol (or bit) using an OCS of length L. The OCS $O_i(t)$ assigned to any anchor B_i at any instant in time t can be represented as shown in Eqn. 2.2.

$$O_i(t) = \sum_{j=0}^{L-1} O_{(j,i)} p(t - jT_c) \tag{2.2}$$

In Eqn. 2.2, $p(t)$ is a rectangular pulse which is equal to 1 for $0 \leq t < T_c$ and zero otherwise. T_c is the chip duration of the OCS and $O_{(j,i)}$ is the j^{th} bit (or chip) of the OCS assigned to the anchor B_i (from the set of all OCSs C_L). The signal generated after encoding a data symbol of anchor B_i with the corresponding OCS is given by

$$x_i(t) = f_i \sum_{j=0}^{L-1} O_{(j,i)} p(t - jT_c), 0 \leq t < T_f \tag{2.3}$$

where, f_i is the data symbol of anchor B_i that needs to be encoded and $T_f = LT_c$ is the duration of the encoded data symbol or data frame. The inner product of the sent data with the OCS is done bit-synchronously. Then, the overall transmitted signal $x(t)$ of all n anchors can be given by [40]:

$$x_(t) = \sum_{i=1}^{n} x_i(t) \tag{2.4}$$

The received signal at the receiver (both the MT and the anchors) will be decoded using the OCSs available in the receiver's OCS table. It can be shown that it is impossible to decode the individual signals correctly if atleast one of the signal in this overall transmitted signal has all 1-bits and is encoded with an OCS code of all 1-bits. Such a signal is referred by us as a *jamming signal*.

2.2.2 Network Tessellation

Network tessellation or anchor grouping is an important aspect of our scheme, which we describe briefly in this section. There are many centralized or distributed algorithms

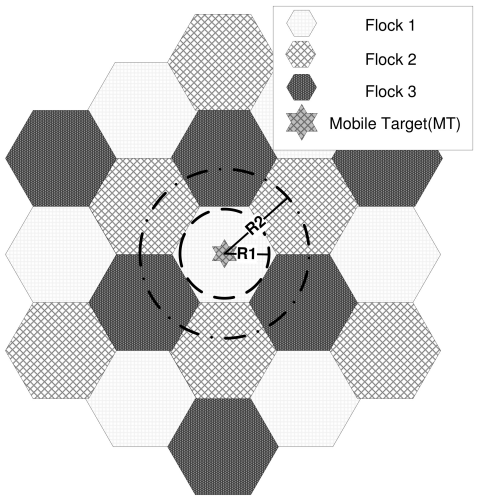


Figure 2.2: Network Tessellation.

in the literature for tessellating distributed wireless networks [214, 19, 122]. Our network tessellation algorithm is based on the Voronoi diagrams and is shown in Figure 2.2. After node placement, we begin from a randomly selected and centrally-located initial anchor node. This anchor sends an invitation message with specific fixed signal strength to all neighboring nodes on the control channel. Nodes within the signal range become members of that particular GC. The nodes at which the received signal strength is less than a given threshold, can attempt to create new GCs and build next layers of GCs and continue to tessellate the entire network. After tessellation, each independent GC is assigned a unique GC number.

After the tessellation phase, the initial anchor can begin clustering the GCs. This is done by placing all anchors who received the initial invitation (including the initial anchor) and formed a GC into flock1. All other nodes that received the initial signal with RSS lower than some threshold value, and therefore formed another GC, would be considered in flock 2 or flock 3. It should be noted that two neighboring GC's cannot be in the same flock (or have the same flock number). To prevent the nodes in neighboring flocks from having the same flock number, every node that receives the signal weaker than the RSS threshold first

registers a new flock number for itself and then broadcast this number to its neighbors. After clustering and flock assignment, all OCSs from all flocks should be added to an OCS table maintained by the MT. We assume that the network is tessellated once at the beginning when the anchors are configured. It can be repeated every time the distribution of the anchors changes significantly.

2.2.3 Adversary Model

We assume that, amongst a total of n anchors in the network, a maximum of a anchors are malicious or cheating. The set of all the malicious anchors is denoted by $\mathcal{A}$. All anchors that are not malicious are assumed to be honest, i.e., they execute the proposed localization protocol correctly. Although many types of attacks are possible in RF-based positioning systems [242], in this work we focus on distance manipulation attacks. In these attacks, anchor nodes cheat by manipulating the distance between themselves and the target node, for example, by either delaying or manipulating the signal strength of the localization messages depending on the distance estimation technique used in the localization protocol. In addition to acting independently, a malicious anchor can also collude with other malicious anchors. In order to effectively communicate with the MT on the CDMA-based data channel, all anchors (including malicious anchors) need to transmit localization messages by encoding them using one of the OCS known to the MT. Coordinating with each other helps the malicious anchors in selecting *different OCSs* for data transmission, thus avoiding interference and data corruption at the target node. Malicious data transmitted using an incorrect OCS will be directly discarded at the target, and thus not included in the location calculation process. It is also reasonable to assume that the malicious anchors do not possess the secret group keys and other cryptographic materials shared only by the honest anchors. Moreover, the malicious anchors are not able to receive (and maintain) a table of OCSs valid during a particular time period because the OCS updates are encrypted with a group key known only to honest anchors. It is also reasonable to assume that there is no trust between the

MT and the set of honest anchors, and so, even the MT does not know these details. From the point of view of the MT, it will use all localization messages encoded with any OCS (present in its table of known network OCSs) for determining its own location. In order to successfully cheat, the malicious anchors (not belonging to the GC) must first *guess* a set of all unused OCSs from a valid flock (otherwise the communication would be rejected by the MT) and then broadcast malicious localization data using these OCSs in order to disorient the target node MT. We would like the readers to note that, although a variety of Denial-Of-Service attacks are possible and can be executed by the malicious anchors, in this section we focus on overcoming only those attacks where the malicious anchors attempt to successfully disorient the target node MT.

2.3 Securing Localization using Jamming

2.3.1 Proposed Localization Protocol

We propose a simple *request-response* strategy in order to perform secure localization in the presence of cheating anchor nodes. In our proposal, as discussed earlier, the MT and anchors communicate on two separate channels. The MT asynchronously broadcasts localization requests on the CSMA/CA control channel and receives responses from anchors on the data channel that uses CDMA multiplexing. The request frames contain a randomly generated request number (D) which is used by the MT to track the corresponding responses. All communications are *broadcast* and a fixed *source address* (for example, 0) is used to maintain *sender anonymity*. Each time the MT needs to determine its location, it periodically broadcasts request frames at random intervals of time ($\ll$ current OCS validity period ψ) each with a different radio power level (between some permitted E_{min} and E_{max} mW) and with a different random request number. We will soon see how such a redundancy in sending localization requests is useful towards securing localization in our proposal.

The (honest) anchors hear for localization requests on the CSMA/CA control channel and broadcast responses on the CDMA data channel. These broadcast responses contain

position information of the corresponding anchors (in the appropriate coordinate system) which are used by the MT for its own location estimation. Response frames to a request with request number D should contain a number $D+1$. This will help the MT identify and only process responses corresponding to its own requests. Sender's request anonymity is of critical importance in our protocol in order to confuse the malicious anchors and to prevent them from selectively targeting requests from the MT. We accomplish this by following an intelligent "*request confusion*" strategy where dummy request frames, similar to the ones sent by the MT, are periodically broadcast by all the anchors. A request frame similar to the one sent by the MT is used in these dummy requests. Such a request confusion strategy makes it extremely difficult for malicious anchors to distinguish valid requests sent by the MT from the dummy ones sent by the honest anchors.

Location response frames sent by the anchors on the CDMA data channel are encoded with the valid OCS assigned to each anchor (during that time period). Optionally, all anchors could also sign their responses using the group signatures of their corresponding GC, which could be verified by other anchors in their GC. On receiving a response, the MT attempts to decode the received frame using an appropriate OCS (from the OCS table stored in its memory). Response frames that cannot be decoded correctly (because of being encoded with an OCS not in its table) or those that do not pertain to its own request are immediately discarded by the MT. As discussed earlier, the MT sends consequtive requests with different transmission power and each with a different request number. The MT waits for atleast *two* responses encoded with the same OCS within a fixed time duration (T_{res} sec) and from the same fixed location (i.e., from the same anchor) before estimating its distance to each such anchor. We will later see how this will help the MT to avoid using coordinates provided by malicious anchors. In our scheme, the MT uses a time-of-flight based approach for location estimation. Time of Arrival techniques require that the MT is synchronized with all the anchor nodes in the network, which seems difficult to achieve. To avoid the strict

synchronization requirements, the MT employs multilateration by using TDoA estimates which do not need knowledge of the absolute time of transmission.

Now, let us focus on how the malicious anchors would attempt to cheat in this protocol and how this cheating can be overcome. Due to the "request confusion" strategy, where both the MT and the anchors send similar requests with different power levels, malicious anchors are unable to identify the source of the requests. Hence, malicious anchors cannot distinguish if a request was from a MT and are unable to estimate their distance to the MT just from the received location requests. Consequently, it is non-trivial for the malicious anchors to selectively manipulate self location information in the response frames in order to successfully disorient the target node. This leads to two types of cheating behavior by the malicious anchors. First, where it will send random self-locations in the response frames. Second, where it will send fixed false self-locations. It should be noted that the response frames by the malicious anchors still need to be encoded by a valid OCS and we assume that a part of the malicious anchor's attack strategy is to be able to obtain such a valid OCS known only to honest anchors.

Both these kinds of cheating behavior results in inconsistent location information which can be easily detected by honest anchors that know their own locations. Examples of such techniques exist in the literature [143, 183]. The protocol can be made further secure by requiring all anchors to sign the response frames with their group key. Obviously, malicious anchors not possessing the group key would be unable to produce the correct group signature, and thus, would be easily detected by the honest anchors. After cheating is detected, honest anchors will selectively jam all future response packets encoded with that particular OCS by broadcasting a jamming signal (of all 1-bits). We can achieve this by an accurate reactive jamming strategy. Such a strategy prevents malicious anchors from sending multiple responses with false location information encoded with the same OCS. As a result, the MT will never be able to utilize the location information sent by the malicious anchors for

location estimation because it requires atleast two responses encoded with the same OCS (or from the same anchor) within a duration of T_{res} seconds.

Our proposed secure localization technique are outlined in Protocols 1, 2 and 3.

Algorithm 1 Parent Anchor in a GC.

```
Generate OCS table with Golay algorithm;
for each time period ψ_k do
    Randomly select OCSs from the set of OCSs valid for the flock;
    Prepare the valid OCS table for advertisement;
    Sign and Encrypt (with a pre-shared group key) the OCS table;
    Broadcast OCS table on the CSMA control channel;
end for
```

2.3.2 Analysis and Discussion

In this section, we analyze the security provided by our scheme and discuss some of its shortcomings. First thing to note is that the current proposal does not provide any formal guarantees or error bounds for the anchor-based localization process. Provided enough honest anchors are available, such guarantees have been provided in the literature [110]. Rather, our protocol aims to provide a mechanism to actively detect and disable a variety of location or distance manipulation attacks originating from malicious anchors.

Malicious anchors in our protocol could either be *outsiders*, i.e., not belonging to a particular GC or *insiders*, i.e., belonging to a particular GC at network initialization. As malicious outsiders do not possess the shared OCS table currently being used, they first need to determine the valid OCS for the flock they intend to cheat in. The probability of choosing the right flock by the uncoordinated outsider adversary depends on the CRF. This probability decreases as the CRF becomes smaller. The likelihood of picking a valid OCS also depends on the length L and the number of OCSs used in that specific GC. Thus, in a GC g with a current valid OCS table of $F_g(\psi_k)$, during any time period k, and an OCS length of M, this probability is $\frac{|F_g(\psi_k)|}{2^M}$ or $\frac{|F_g(\psi_k)|}{L}$. By using an OCS currently used by another malicious anchor or an honest anchor will corrupt the data received at the MT and cannot be used to disorient it. As the OCSs used by honest beacons are changed periodically, a

Algorithm 2 Honest Anchors.

```
while data on CSMA control channel do
  if data is from parent anchor then
    Verify group signature and decrypt data;
    Identify an OCS to use from the table of valid OCSs;
    Save the table of valid OCSs for the current time duration;
  else if data is a localization request then
    Create and ansynchronously send a dummy location request on the CSMA control channel
    with some probability p;
    Let D be the request number in the received request;
    Create a response packet with response number D + 1 and containing self location coor-
    dinates;
    Optionally, sign the packet with group key;
    Encode packet with the chosen OCS (bit-wise inner product);
    Synchronously send response packet on CDMA data channel;
  else
    drop the data;
  end if
end while
while data on CDMA data channel do
  if data contains location responses then
    if cheating detected in location responses then
      Create a jamming signal (packet consisting of all 1's);
      Broadcast the jamming signal on CDMA data channel;
    else
      Drop the packet;
    end if
  end if
end while
```

brute force type of attack by the adversary would slowly become infeasible. Moreover, as the MT transmits the requests with multiple power levels that extend beyond the current flock or GC, it provides some robustness against the scenario where all OCSs in a particular flock or GC are compromised. Optionally, cryptographic techniques, such as group signatures, can also be used to detect and disable communications from malicious outsiders.

Malicious insiders will be able to effectively communicate with the MT using a valid OCS, but would be easily detected by the honest anchors based on the discrepancy of the location information transmitted in the response packet. The honest anchors will use the proposed CDMA-based jamming approach to prevent further malicious responses from these insiders. Readers should note that even if the malicious outsiders are able to obtain valid OCSs for

Algorithm 3 Mobile Target (MT).

```
while data on CDMA data channel do
  Decode the packet or data frame, i.e., calculate inner product using all valid OCSs;
  if (location response packet) and (flock# and GC# match network plan) then
    if another response encoded with same OCS and from same coordinates received no earlier
    than T_res seconds then
      Save anchor coordinates
    end if
  else
    Drop the packet;
  end if
end while
Select atleast three coordinates;
Perform Multilateration;
```

communication with the MT, discrepancy in location information can be easily verified at the honest anchors who will jam future responses originating from these compromised OCSs.

2.4 Evaluation

In this section, we present simulation results for our proposal on securing anchor-based localization.

2.4.1 Simulation Setup

In our simulations, we consider a $1000m \times 1000m$ network area where anchor nodes (both honest and malicious) are distributed uniformly over the network area. One such distribution of 200 honest and 200 malicious anchors used in our experiments is shown in Figure 2.3. The position of the MT is chosen randomly from the network area. Table 2.1 outlines the various network and protocol parameter values used in our simulations. After deployment, we first tessellate the network and cluster the honest anchors based on the technique outlined in Section 2.2.2. After network tessellation, the MT begins the location discovery process by sending two initial requests on the control channel, and then additional ones as needed. For location computation, we use multilateration by estimating the time difference of arrival of valid coded bits from the various anchors.

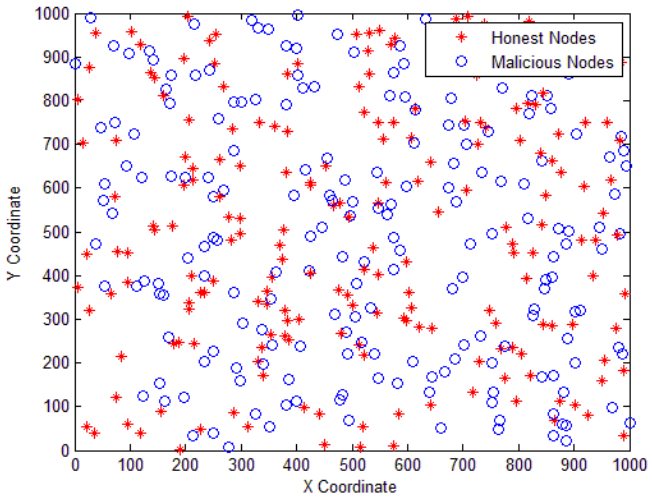


Figure 2.3: Distribution of Anchors in the network.

2.4.2 Simulation Results and Discussion

We evaluate our proposed secure localization protocol under several different network conditions. In our first set of experiments, we deploy 200 honest anchors and 200 malicious anchors (as shown in Figure 2.3). Our first goal is to verify the effectiveness of our protocols in eliminating the cheating effect of malicious anchors. One of the first observation that we make is that, in all our simulation runs, all malicious or cheating anchors are successfully jammed, thus preventing them from disorienting the MT during the multilateration procedure. With only the data received from the honest anchors, the MT is able to accurately localize itself. Thus, we can conclude that the localization error (Euclidean distance between the calculated location and MT's real location) observed in our simulations is not because of the cheating effect of the malicious anchors, but is rather due to the use of CDMA for communication between the anchors and the MT.

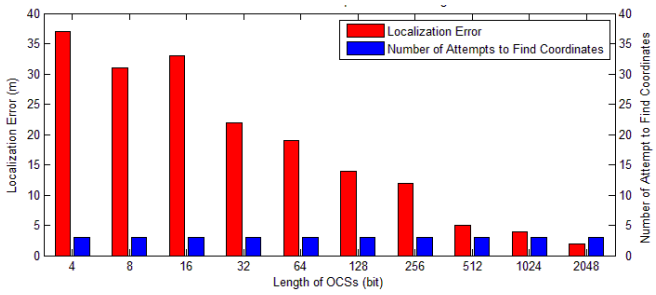
On varying the OCS length (OCSL), we observe that the localization error decreases considerably when the OCS length used by the anchors increases (Figure 2.4 (a)). The localization error is relatively high, for example, roughly 37 meters, when a 4 bit-OCS is used. On increasing the OCS length to 2048 bits, the localization error decreases to less than 2 meters, which is a significant improvement. This behavior is due to the fact that

Table 2.1: Simulation Parameters for LOCJAM

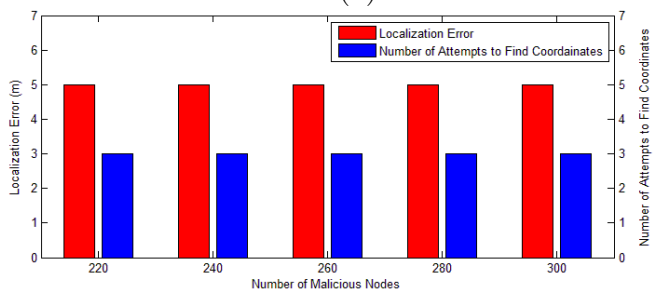
Parameter	**Value**
Simulation area	$1000m^2$
Tx power on CSMA/CA-based control channel	$E_{min} = 1mW$ to $E_{max} = 15mW$
Tx power on CDMA-based data channel	$15mW$
# of honest anchors	200
# of malicious anchors	$200 - 300$
Carrier frequency	$2.4 - 2.48GHz$ (Zigbee)
Bit rate	$250Kbit/sec$
Packet rate	$5208Packet/sec$
Orthogonal Chip Code generator	Golay
Chip Code Size	Varies between 4 to $2048bits$ - Asynchronous OCS
CRF	$\frac{1}{3}$
Radio propagation model	Free Space
Maximum delay spread	$3\mu sec$
Bandwidth Efficiency	84 %
Cluster layout (Honest and Malicious Anchors)	3 GCs
T_c	$0.1 \times 10^{-3} timeslots$

a smaller OCSL means more interference between the communicating anchors, resulting in a relatively larger error in the data received at the MT. This translates to a larger localization error. We can conclude that by choosing an OCS of appropriate length, fairly accurate multilateration-based localization of the MT is possible. However, the number of attempts, defined as the number of distinct localization requests sent by the MT, required by the MT inorder to overcome the cheating effect of malicious anchors and to compute its position does not depend on the OCSL, as seen in Figure 2.4 (a). We also observe that the total number of requests needed by the MT is minimum. The responses from the third request is not even used during the location computation.

When the number of malicious anchors increases and the distribution of the honest anchors is the same, we can see from Figure 2.4 (b) that, for a particular OCSL (512-bits in this case), the increase in the number of malicious anchors has no effect on the localization



(a)



(b)

Figure 2.4: Simulation Results (a) Localization Errors versus OCS Length and (b) Localization Errors versus Number of Malicious Nodes.

accuracy and the number of requests needed for secure localization. This also shows that in our scheme, a smaller number of honest anchors can successfully disable a relatively larger number of malicious or cheating anchors, further proving its robustness in highly unsecure environments.

While the above results are for a single distribution of honest and malicious nodes, we did run our simulations for multiple iterations and distributions of honest and malicious anchors. In our next set of experiments, we simulated 10 different uniform distributions of honest and malicious anchors where each distribution was simulated 100 times. In these simulations, we considered an OCSL of 512 bits and the total number of honest and malicious anchors were fixed at 200. In these simulations, we observed that the average localization error was $4.36m$ and the average number of MT requests before localization was around 4, which is very similar to the earlier results (Figures 2.4 (a) and (b)). These average results show that

our scheme consistently peforms well under various distributions of honest and malicious anchors.

In summary, our simulation results confirm that localization schemes on DSSS or CDMA-based communication schemes are efficient and reactive jamming is an effective strategy to disable cheating anchors in such anchor-based localization schemes, and its effect on the overall localization accuracy is minimal[2].

[2]The content of this chapter has appeared in [20].

CHAPTER 3
OPTIMAL RESOURCE ALLOCATION IN COGNITIVE SMART GRID NETWORKS

Utilizing information and communication technologies, the power industry is moving towards the *Smart Grid*. This next generation power grid is expected to improve the way electricity is generated, distributed, and transmitted to the consumer by enhancing the efficiency, reliability, sustainability, and economies of the power grid. However, due to the high volume and high granularity of the data generated by smart home appliances and smart electricity meters, careful planning and management of the communication network is necessary. Utility companies either lease the communication network [86] or build their own network infrastructure [127]. From utilities' perspective, both bandwidth and network are expensive resources. Thus, researchers have been attracted to the applicability of *Cognitive Radio Networks (CRN)* recently, to overcome the problem of bandwidth scarcity [263, 81, 248, 264, 18, 186, 104, 27, 7]. Instead of having a dedicated communication backhaul, and hence, dedicated bandwidth, *Cognitive Smart Grid Networks (CSGN)* co-exist with existing *Primary Network (PN)*. For instance, Super WiFi or IEEE 802.22 [227] can be used as the *Secondary Network (SN)*. CSGN is also able to transmit as a SN simultaneously with the PN.

Existing static spectrum management and utilization policies in wireless networks do not efficiently utilize the available spectrum and have resulted in acute spectrum shortage for upcoming wireless devices and applications [73]. In the last few years, we have witnessed a tremendous rise of interest in cognitive radio technology or CRN, where the *Secondary Users (SU)* enabled with cognitive radio selectively or opportunistically communicate over certain bandwidth allocated entirely to *Primary Users (PU)*. Cognitive radio technology has the potential of enabling dynamic spectrum access in wireless networks, thereby increasing the capacity and utilization of wireless communications. In CRNs, the PUs that are licensed

to access a certain frequency band, have unconditional access to these bands without any interference from other users, but they might be willing to share their licensed spectrum with the unlicensed SUs.

Based on the channel access and transmit strategies of the PUs and SUs, CRNs can be classified into three broad groups: 1) *interweave* CRNs, 2) *overlay* CRNs, and 3) *underlay* CRNs. In the interweave mode [267, 145, 239], each SU senses for gaps in the spectrum used by a PU and opportunistically uses these gaps to send its own data. In the overlay mode [62, 271, 34], SUs have knowledge of PUs channel access strategy and pattern. The SUs then use this knowledge to transmit data on the shared channel when it is not being used by the PUs. Finally, in the underlay mode [44, 165, 34], SUs access the channel in parallel with the PUs by appropriately adjusting certain SN transmission parameters in order to avoid any interference with PUs in the PN. Recently, CRNs have received a lot of attention from researchers, who primarily focus on secure and efficient techniques for SUs in different network areas to co-exist with PUs without interference and with desired QoS guarantees (for both PUs and SUs). Despite the plethora of proposed access strategies for CRNs, how to improve the *Number of Secondary User (NSU)* of SNs while maintaining a desired level of QoS for the PUs as well as other users in SN, remains an important open research problem. Ideally, we would like to improve NSU without any reduction in QoS of the PUs in a PN.

As a part of *Internet of Things (IoT)*, by 2020, we expect to have more than 800 million *Smart Meters (SM)* and smart appliances globally [89]. Since the number of smart appliances, SMs and other devices with communication capabilities continues to grow, and the necessity to assign available bandwidth for their communication, we need to optimize the network resources. In this section, we propose a novel CDMA-based channel access and resource allocation scheme for static CSGN that improves the NSU in the SN without resulting in high end-to-end *Bit Error Rate (BER)* in the non-CDMA based PN[1]. Our proposed

[1]The content of this chapter has appeared in [21].

method that utilizes a hierarchy of SUs consist of smart appliances (lowest hierarchy), SMs, *Gateways (GW)* and *Super Gateways (SGW)* (highest hierarchy). Additionally, we assume that our SUs are capable of CDMA communication. Our novel architecture assigns carefully chosen orthogonal codes to the SUs in the lowest hierarchy. Then, it assigns substrings of the *Orthogonal Chip Sequences (OCS)*, which have adequate level of orthogonality, to the other SUs in CSGN. Our proposed method, which uses substring of the chip code, to improve the NSU, relies on the fault-tolerance property of CDMA channel multiplexing. Due to this fault-tolerance property, receivers are able to correctly decode the received waveforms despite of errors [21]. Compared to previous studies, where transmission power manipulation and code assignment methods are used to increase NSU, our method provides an improved NSU in the presence of PUs communicating on PN carriers with acceptable BER and interference level. Additionally, SUs in our method would send the data with same bit rate by modulating different length OCSs in proposed SN hierarchy in a shared timeslot.

The rest of the section is organized as follows. Background and related work is described in Section 3.2. Network and system model is outlined in Section 3.3. The Proposed hierarchical resource assignment in SN is represented in Section 3.4. The proposed method's design is discussed in Section 3.5, and evaluated through simulations. The section is concluded with a summary of results and contributions in Section 3.6.

3.1 Background and Related Work

Earlier research on improving number of SUs in a CRN and using CRN in SGNs employ a variety of techniques. Here, we study approaches that are related to our work.

Amin et al. [7] improve data rate for real-time and best effort traffic in CSGNs, utilizeing a global resource controller. Qiu et al. [186] analyze several hardware requirements to achieve secure data communications in CSGNs. Ghassemi et al. [81] propose a WAN communication scheme on CSGNs backhaul to make a fault tolerant and reliable infrastructure on TV primary network. Also, the authors investigate coverage extension mechanisms on their

proposed method. Yu et al. [263] improve QoS level using a hybrid spectrum access model on licensed and unlicensed bands in CSGN. Cost saving in spectrum leasing and network interference optimization are considered in the above-mentioned method. The authors also propose a network scale performance optimization and hierarchical dynamic spectrum access model to share white or free bands in different hierarchies of CSGN [264]. Wang et al. [248] study the technical challenges and solutions in multimedia data transmission in CSGNs for network monitoring. Bicen et al. [18] utilize cognitive sensor networks in CSGNs to build a reliable and low cost remote monitoring system. Huang et al. [104] propose a priority-based traffic scheduling algorithm to decrease spectrum sensing error and system utility optimization performance in CSGNs. In this method, using a differential service algorithm, packets are divided into different groups of priority and are sent based on their priorities. Bu et al. [27] propose a green CSGN using small cells for dynamic pricing optimization, energy efficient power allocation, and interference management.

Dashti et al. [55] study utility-based fairness criteria for radio recourse allocation in underlay SNs by proposing a joint rate and power allocation algorithm. Li et al. [138] introduce a transmission power and code allocation method using a blind identification scheme in a CDMA-based underlay SN. Zhang and Su [267] propose an opportunistic spectrum sharing scheme to improve channel utilization in infrastructure-based CRNs where the SUs can adaptively select either an interweave or underlay mode of operation to transmit data. While operating in the underlay mode, the SUs adapt their transmit power such that the interference caused by these transmissions is below the tolerable noise floor of the spectrum in a PN. CDMA enables PUs and SUs in CRNs to simultaneously co-exist on the same frequency band [40, 12, 247, 203, 187]. Chang [35] propose an efficient code assignment scheme in a hierarchical MC-CDMA-based Cognitive femtocell system for achieving high spectral efficiency and lower interference with PUs.

In this chapter, we propose a novel hierarchical communication architecture for stationary or non-mobile CSGNs. Our main idea is to use an OCS and substring of OCS, called *sub-*

OCS, in underlay SNs while TV network is transmitting its data as PN to its users as PUs. To the best of our knowledge, this is the first work that leverages a sub-string of OCSs to improve the number of supported SUs. Also, we leverage on the power control techniques to further decrease interference and noise ratio, and further improve the number of supported SUs. Hence, utilizing sub-OCSs along with proper power control techniques form the foundation of our architecture towards increasing NSU in CSGNs.

3.2 System Description: Network Model and Configuration

The network model that we consider in this work is shown in Fig. 3.1. We assume that SMs are tessellated into clusters of m SMs per cluster. Every cluster includes a GW which is responsible for gathering the data of the SMs within that cluster. This data gathering can include aggregation, concatenation, or any other kind of operation desired by the utility company. Every k cluster form a *Supercluster*, represents a residential neighborhood. Figure 3.1 represents a supercluster with $k = 7$, where there are seven clusters per supercluster. The data of all the k GWs within a supercluster are gathered at the SGW. This SGW is one of the existing k GWs in a supercluster such that it is within the transmission range of the $k-1$ other GWs. SGW is responsible to transmit the data to *Utility Control Centers (UCC)*. This tessellation process for a supercluster can continue to cover an entire urban area. In this work, we elaborate our protocol for one supercluster. Without the loss of generality, rest of the superclusters operate reusing the same protocols and procedures.

Our underlay CSGN scenario consists of a SN of SUs operating alongside a PN of PUs. The SN always operates in the infrastructure mode. In other words, the SN uses a coordinator or Gateway for session setup and data transmission. The SN will choose an appropriate parameter for data transmission, independent of the PN, as explained in Section IV. SUs use an out-of-band dedicated two-way *Control Channel (CC)* to coordinate the transmission channel and interference avoidance. We assume that the SN is a CDMA-based network employing *Direct Sequence Spread Spectrum (DSSS)*[40], i.e., the transmitted data is modu-

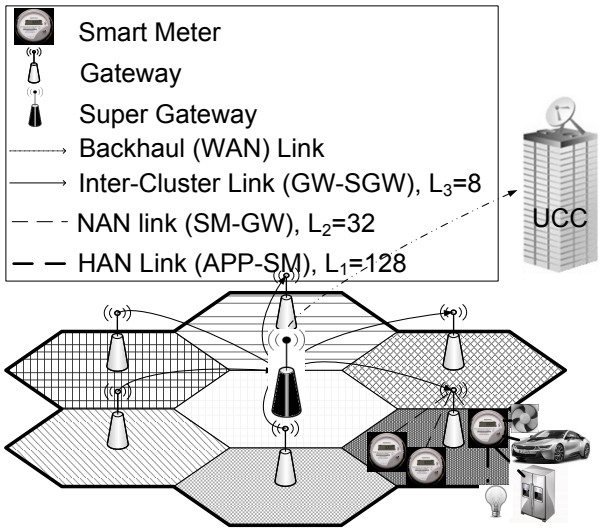


Figure 3.1: Cognitive Smart Grid Network Architecture.

lated or spread using a transmitter-specific spreading code or OCS and the PN is considered to be TV network although it can be any non-CDMA based network. Also, we are using four CDMA communication links in our architecture. APP-SM link, is a link to connect appliances to a SM in a *Home Area Network (HAN)*. To form a connection between SMs and GWs in the *Neighborhood Area Network (NAN)*, we utilize SM-GW links as the second hierarchy in our model. In the third hierarchy, GW-SGW links form an inter-cluster connection between GWs and SGW.

3.3 Proposed Hierarchical Resource Allocation in CDMA-based CSGNs

If standard CDMA-based communication is used in CSGN, the number of supported SUs are limited to the maximum number of available OCSs or the OCS length L [40]. In other words, with an OCS of length L, a total number L SUs can transmit simultaneously. In this work, we plan to increase this restriction of number of OCSs or NSU by reusing substrings of already used OCSs in a hierarchical fashion. In our proposed approach, OCSs of length L are assigned for the data transmission of smart home appliances. Then, substrings of

these OCSs are used on the next levels of our hierarchical design by SMs, GWs, and SGWs. These sub-OCSs are then selected such that they do not interfere with the transmission of other nodes that are simultaneously transmitted with longer OCSs in the SN. Also, utilizing specific power control mechanisms, we can guarantee that SUs in SN will not interfere or collide with the transmission of PUs in the PN.

3.3.1 Generating sub-OCSs and resource allocation in proposed CSGN

We generate the OCSs for the *Spread Spectrum Communications (SSC)* on the SN using *Golay* [40] and *Persian Chip Code (PCC)* [24] algorithms. PCC is an optimized variant of the Golay. The Golay OCS generator algorithm uses the following recursive generator matrix [40]:

$$C_L = \begin{bmatrix} C_{\frac{L}{2}} & \bar{C}_{\frac{L}{2}} \\ C_{\frac{L}{2}} & -\bar{C}_{\frac{L}{2}} \end{bmatrix}$$

$$C_L = [\, A_L \;\; B_L \,], \bar{C}_L = [\, A_L \;\; -B_L \,] \text{ and } C_1 = \bar{C}_1 = [-1] \tag{3.1}$$

where, $L = 2^M$ is the total number of available OCSs, $M \geq 1$ is the number of iterations in OCS generation algorithm, $A_L = \begin{bmatrix} C_{\frac{L}{2}} \\ C_{\frac{L}{2}} \end{bmatrix}$ and $B_L = \begin{bmatrix} \bar{C}_{\frac{L}{2}} \\ -\bar{C}_{\frac{L}{2}} \end{bmatrix}$. OCSs of 16-chip length generated using Golay are shown in Fig. 3.2-a. In recursive OCS generation algorithms such as Golay (or PCC), OCSs can be organized into groups called *flock* based on *chip pattern similarity* and *chip distance* between OCSs. In Fig. 3.2-a, we can see different flocks for 16-chip OCSs, where each flock size is $\frac{1}{4}^{th}$ of the set of all (sorted) OCSs of a particular length.

Both Golay and PCC algorithms are able to produce L OCSs with a length of L-chips. The PCC generator matrix is shown in Eqn. 3.2. Correspondingly, OCSs of 16-chip length generated using PCC are shown in Fig. 3.2-b.

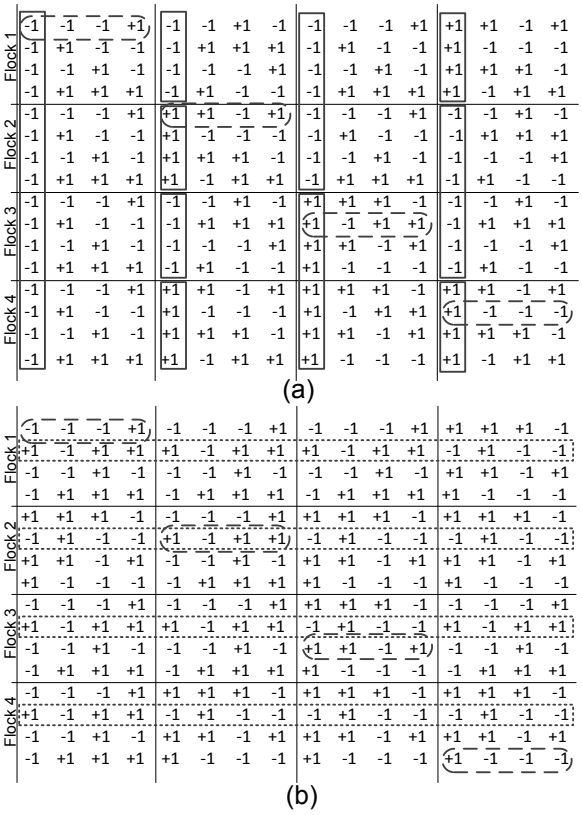


Figure 3.2: a) A 16-chip Golay OCS matrix, solid lines are repetitive patterns and dashed lines are selected sub-OCSs. b) A 16-chip PCC OCS matrix, dotted lines are inverted OCSs after column shifts and dashed lines are sub-OCSs.

$$P_{4^n} = \begin{bmatrix} P_{4^{n-1}} & P_{4^{n-1}} & P_{4^{n-1}} & -P_{4^{n-1}} \\ -P_{4^{n-1}} & P_{4^{n-1}} & -P_{4^{n-1}} & -P_{4^{n-1}} \\ P_{4^{n-1}} & P_{4^{n-1}} & -P_{4^{n-1}} & P_{4^{n-1}} \\ P_{4^{n-1}} & -P_{4^{n-1}} & -P_{4^{n-1}} & -P_{4^{n-1}} \end{bmatrix}$$

$$\forall n \geq 1, P_1 = [-1] \tag{3.2}$$

OCSs generated by PCC have equal number of 1's and -1's, in contrast to OCSs generated by Golay. This property makes data transmission using PCC more fault tolerant than Golay. Among all OCS generator algorithms, PCC and Golay are preferred because of equality in OCS length and number of generated OCSs, and high level of orthogonality. The length of

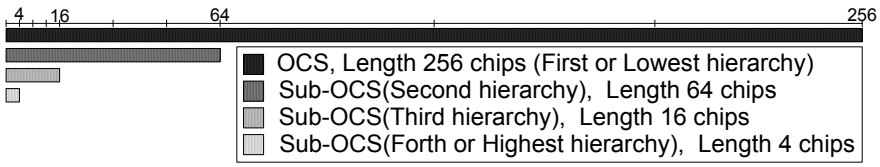


Figure 3.3: One appliance (using OCS with 256 chips length) and three SUs (using sub-OCSs) in three hierarchies in each flock.

OCSs to be used in SSCs depends on various factors such as the number of users, noise, and interference in the network. Golay and PCC have excellent fault-tolerance properties, which increases with the length of the code. The fault-tolerance capabilities of both codes can be shown to be around 37.5%, i.e., the receiver is able to decode the transmitted bit even if the amplitude of received signal is 37.5% less (or more) than the defined thresholds. An analysis of this property is presented in Section 3.3.3.

In our proposal, for each OCS of length L assigned to and used by an appliance (lowest hierarchy SU in CSGN), we assign a substring of those OCSs of length at most $\frac{L}{4}$, also called *sub-OCS*, for simultaneous use by a SU in a higher hierarchy of CSGN. Sub-OCSs can be assigned and used in a hierarchical fashion as shown in Fig. 3.3, where the first level sub-OCS of an OCS of length L could be of length $\frac{L}{4}$ and then decreasing by a factor of four. For example, Fig. 3.3 illustrates a 256-chip OCS that is used by a smart appliance in the HAN and a three level hierarchy of available sub-OCSs that can be simultaneously used by higher hierarchies of SUs. As we can see in Section 3.3.3, every time the OCS length is divided by four (to generate a new level of sub-OCSs), the transmission power amplitude is doubled.

Due to the fault tolerance property of the chip codes used, the upper levels of this chip code hierarchy, with longer OCSs (or sub-OCSs), can tolerate the interference caused by the parts of waveforms that are used by the chip codes in the higher levels SU. However, transmitting with these OCSs and sub-OCSs will impose a tolerable interference on the PN, if the interference is lower than the interference threshold [72]. Also, SU receivers will be able to decode their received data by these signal amplitude increments irrespective of the effect of the lower hierarchies SUs' transmission. Thus, bad effects of sub-OCSs that are used

Table 3.1: Number of Extra Sub-OCSs in PCC and Golay codes

sub-OCSs' Hierarchy(H)	**Length of OCS**	**Number of Golay Extra SUs by sub-OCSs**	**Number of PCC Extra SUs by sub-OCSs**
7, Highest H	16	4_4^2	4_4^2
6	32	8_8^2	N/A
5	64	$4_4^4 + 16_{16}^2 = 20$	$4_4^4 + 16_{16}^2 = 20$
4	128	$8_8^4 + 32_{32}^2 = 40$	N/A
3	256	$4_4^8 + 16_{16}^4 + 64_{64}^2 = 84$	$4_4^8 + 16_{16}^4 + 64_{64}^2 = 84$
2	512	$32_{32}^4 + 128_{128}^2 = 160$	N/A
1, Lowest H	1024	$4_4^{16} + 16_{16}^8 + 64_{64}^4 + 256_{256}^2 = 340$	$4_4^{16} + 16_{16}^8 + 64_{64}^4 + 256_{256}^2 = 340$

by higher hierarchies SUs impose a tolerable interference on lower hierarchies SUs. Selected Golay and PCC sub-OCSs for a 16-chip OCS are as shown in Fig. 3.2-a and Fig. 3.2-b by dashed lines. These selected sub-OCSs will have the lowest interference rate among all other sub-OCSs that reside in the same column and are chosen due to the fact that they have equal number of 1's and -1's in each flock.

The total number of extra SUs that can be supported purely based on the sub-OCSs selected from fixed-length OCSs are shown in Table 3.1. The number of extra users is denoted as $\aleph_\varrho^\delta$, where $\aleph$ is the number of usable sub-OCSs of length ϱ, and δ-times bigger signal amplitude compared with the original OCSs' amplitude. As implied earlier, the number of extra SUs supported in our network is equivalent to the number of employed sub-OCSs. It should be noted that, if one sub-OCS in a flock of Golay is used in a CRN, it interferes with particular chips (as shown by solid lines in Fig. 3.3-a) of other OCSs in all flocks. Contrary to Golay, in PCC, sub-OCSs do not interfere, more than the threshold floor, with other OCSs in the same or different flocks. This is because, unlike Golay, corresponding chips in all flocks are not identical, which will result in less interference at the receiver. Thus, the Golay OCSs will reduce the efficiency of our CRN architecture as the interference would be higher. In the proposed method, all SUs are able to generate PCC OCSs and sub-OCSs independently.

Let us explain this further by means of a simple example. Based on Fig. 4.1, assume all appliances send their data to SM on L_1 (APP-SM link). Then, SM forwards the received data to GW on L_2 (SM-GW link). Also, in the next hierarchy, GWs send the gathered data from the SMs to SGW on L_3 (GW-SGW link). Based on these hierarchies in the network, we can allocate OCSs and sub-OCSs as follows. Connections on L_1 between appliances and SMs use OCSs with length $L = 128$ with minimum required power on one of the TV bands [72, 152]. SUs in proposed CSGN can utilize an adaptable transceiver on different hierarchies with different level of Tx power. For instance, "LC169TR-1" transceiver chipset supports adjustable Tx powers in range of $5mW - 750mW$ on TV bands (IEEE 802.22)[152]. SGW and GW are responsible for sensing the channel and allocating available frequency bands to SMs [263, 264, 104]. Using lower transmission power, e.g. $V_1 = 5\ mW$, PUs will be able to transmit simultaneously on the PN [152]. This power generates small and tolerable noise on the PN in the vicinity of the transmitter. Therefore, if SGWs and GWs coordinate on recognizing and allocating OCSs and sub-OCSs, they can reuse these OCSs and sub-OCSs in other parts of the neighborhood and network. If we assume $L_1 = 128$, each SM can have 128 connections to appliances at the same time. All the SMs are able to reuse OCS, because of low transmission power on L_1 where we cannot reuse sub-OCSs on L_2 and L_3 because of shared mediums on L_2 and L_3. Thus, we should utilize a unique sub-OCS on each link on L_2 and L_3 in a cluster of GWs in range of one SGW. Based on Table 3.1, if we have OCSs with length $L = 128$ on L_1 in lowest hierarchy, and L_2 in the second hierarchy, we can have 32 orthogonal sub-OCSs with length $L = 32$ and transmission power $V_2 = 10\ mW$ [152]. These sub-OCSs can be used to transmit simultaneously and without interference on the network. On the highest hierarchy, for GW-SGW communications and based on Table 3.1, GWs and SGWs should encode their data utilizing sub-OCSs with length $L_3 = 8$ and $V_3 = 20\ mW$. Also, sub-OCSs on L_2 and L_3 connections are orthogonal and unique for each of connections.

Transmission power on L_2 and L_3 are higher than transmission power on L_1. Hence, if sum of transmission power on L_2 and L_3 is smaller than the tolerable noise or noise margin

in PN (Based on FCC(DT-LIC), TV network which is lower than 54.8 dB or 2.511 Watt [72]), all transmitted data on proposed CSGN are delivered to destinations without collision or interference on the PN [72]. Therefore, all the communicating nodes in our proposed CSGN can co-exist with PN on used TV bands using three levels of hierarchical OCSs and sub-OCSs, simultaneously. Also, each SM's, GW's, and SGW's coverage area is related to their transmission power on that network. In high density CSGNs, GW and SGW can allocate longer OCSs to appliances to make connection to SMs. Thus, the proposed model can increase OCSs' length proportional to the number of SMs in each cluster. In the above example, each SM can handle 128 Appliance-SM connections and 32 SM-GW connections in each neighborhood when the CSGN is able to support 8 GW-SGW connections in the proposed cluster. Therefore, based on Table 3.1, GW and SGW are able to choose the right length of OCSs on L_1 proportional to the number of SMs in each cluster. This selection, in addition to increasing the number of OCSs on L_1, can increase the number of sub-OCSs on L_2 and L_3. For instance, for OCSs with length $L_1 = 256$, our proposed CSGN can support 256 Appliance-SM connections, 64 SM-GW connections, and 16 GW-SGW connections.

3.3.2 Transmission Parameters

In this section, we summarize physical layer properties of a CDMA-based CSGN and clarify the received signal and its shape at the receivers. Also, the processing gain, which is directly affected by the number of supported users in our method, is described in this section. The received signal without considering multipath fading is shown in Eqn. 3.3, where $r(t)$ is the received signal at time t, and τ is the decoding delay at the receiver.

$$r(t) = \sum_{i=1}^{NSU} S_i(t - \tau_i) + \eta(t) \tag{3.3}$$

$S_i(t)$ in Eqn. 3.4 denotes the shape of the spread received signal (spread by the i^{th} appliance's OCS), and in Eqn. 3.5, $\acute{S}_i(t)$ illustrates the shape of the spread received signal coded by the corresponding sub-OCS of the i^{th} SU. E_S is the energy symbol in the i^{th} SU for full length

OCSs and $\acute{E}_S$ is the energy symbol in the i^{th} SU for sub-OCSs. Also, $d_i(t)$ is the data sent by the i^{th} user which will be spread either by an OCS or sub-OCS. $OCS_i(t)$ and $subOCS_i(t)$ represent the i^{th} SU's OCS and sub-OCS at time t. f_c is carrier frequency and $\frac{1}{T_c}$ is the chip rate.

$$S_i(t) = \sqrt{\frac{2E_S}{T_S}} d_i(t).OCS_i(t).\cos(2\pi f_c t + \theta_i) \tag{3.4}$$

$$\acute{S}_i(t) = \sqrt{\frac{2\acute{E}_S}{T_S}} d_i(t).subOCS_i(t).\cos(2\pi f_c t + \theta_i) \tag{3.5}$$

Eqn. 3.7 represents the spreading factor, which happens to be equal to OCS length (L) in our architecture, when T_c and T_s are chip duration and symbol duration, respectively.

$$L \triangleq \frac{number\ of\ chips}{symbol\ size} = \frac{\frac{1}{T_c}}{\frac{1}{T_s}} \tag{3.6}$$

$$Spreading\ factor = P_G = \frac{B_{SS}}{B} = \frac{\frac{1}{T_c}}{\frac{1}{T_s}} = \frac{T_s}{T_c} = L \tag{3.7}$$

Sub-OCSs in the proposed method, by design, are employed on the same bandwidth as the PN and do not require increasing CRN's bandwidth and carrier capacity. Thus, it will be able to increase NSU while the processing gain P_G in SSC remains the same as shown in Eqn. 3.7. The processing gain in Eqn. 3.7 represents the number of available OCSs per symbol where B_{SS} is one-sided system bandwidth of spread spectrum signal on overall bandwidth and B is the minimum required bandwidth to spread the data and $B_{SS} \gg B$ [182].

In the next Section, we will use one level of *Signal-to-interference-plus-noise ratio (SINR)* in our simulation. Based on the analysis of the fault-tolerance property in Section 3.3.3, the desired SINR of SUs to correctly decode the required bit is calculated by Eqn 3.8. In this equation g^{ρ}_{ji} indicates the gain of the uplink (ρ), between the j^{th} SU (transmitter) and i^{th} BS (receiver). Also, η_i is the interference plus noise power in the network and $SINR_i$ represents the SINR in the i^{th} SU. The $SINR_i$ must be greater than $Q^*(t)$ (the required link quality at SN at time t) to be able to decode SU's waveform in the SU receiver. $Q^*(t)$ could be set

by the SUs based on the environmental noise, given a SINR threshold. $P_G^S = P_G.\gamma$, where γ is a factor that increases the SUs' gain in order to compensate the use of sub-OCSs.

$$SINR_i = P_G^S \frac{P_i g_{i,i}^{(\rho)}}{\sum_{j=1,\, j\neq i}^{N_S} P_j g_{j,i}^{(\rho)} + \eta_i} \geq Q^*(t) \tag{3.8}$$

3.3.3 Fault tolerance threshold in CSGNs

The fault tolerance threshold depends to the model of data transmission in CSGNs. The data transmission models on spread spectrum are synchronous or asynchronous. Since in synchronous data transmission model all senders spread their data on channel at the same time, synchronous data transmission can handle more fualt than asynchronous data transmission model.

3.3.3.1 Fault tolerance threshold in synchronous CSGNs

In this section, we explain the fault tolerance property thresholds in SSC that we assumed in our proposed CSGN. Garg [119] shows that DS-CDMA receivers will interpret any amplitude between [0+ε,$\mathcal{A}$] as 0 and [-$\mathcal{A}$,0 $-$ ε] as 1. Hence, their scheme will tolerate up to length of OCSs minus noise power, or $L - \varepsilon$, fault without considering idle senders (As we will show, L is proportional to $\mathcal{A}$). Gerakoulis [80] also elaborates that orthogonal Gold codes can tolerate up to 50% of the OCS length of jitter. We are assuming that the sender can be silent at times. Thus, our scheme consists of three voltage levels representing the sender sending 1, nothing, and -1 respectively. Based on what we will show in this analysis, the receiving node will be able to tolerate an error of lower than 37.5% (of the received decoded value) in asynchronous and 50% in synchronous transmissions.

In the following equations in this section, $C(t)$ is the channel state in the synchronous CSGN at time t and b_i represents the desperad signal using i^{th} OCS by a smart appliance in the synchronous CSGN. Also, $\eta(t)$ represents AWGN at time instant t. The $\mathcal{A}_i(t)$ and $\mathcal{A}_g(t)$ represent the amplitude share of the data spread by the i^{th} OCS and the g^{th} sub-OCS,

respectively at time t. Let b_i represent the received despread decoded bit in the i^{th} receiver. $OCS_i(t)$ shows the t^{th} spreading chip in the i^{th} OCS.

$$C(t) = \sum_{i=1}^{L} \mathcal{A}_i(t) + \sum_{g=1}^{NSU=(log_4 L)-1} \mathcal{A}_g(t) + \eta(t) \tag{3.9}$$

$$b_i = \sum_{t=1}^{OCSLength=L} C(t).OCS_i(t). \tag{3.10}$$

Let $Q^*(t)$ represent the defined threshold of the required $\mathcal{A}$ in a synchronous CSGN at time t. Thus, in Eqn. 3.11 $\mathcal{A}$ represents the maximum signal amplitude (voltage) in the receiver that will be proportional to L or length of the OCSs in the maximum state when all appliances in CSGN are using their OCSs. Maximum peak-to-peak amplitude can be $2\mathcal{A}$ while $\mathcal{A}$ is proportional to L. Thus, if we want to define a threshold, we can decode the despread bit b_i to 1 in the i^{th} receiver, if b_i is greater than $\frac{\mathcal{A}}{2}$, and to -1, if b_i is smaller than $\frac{-\mathcal{A}}{2}$. $Q^*(t)$ represents this threshold in a synchronous transmission when the receiver is able to decide for choosing 1 or -1 where the despread waveform is greater or smaller than 50% of the maximum amplitude $\mathcal{A}$. Figure 3.4-a illustrates the case of the synchronous CSGN.

$$if \begin{cases} \frac{+\mathcal{A}}{2} < b_i < +\mathcal{A} & b_i = +1 \\ \frac{-\mathcal{A}}{2} \leq b_i \leq \frac{+\mathcal{A}}{2} & b_i = 0 \\ -\mathcal{A} < b_i < \frac{-\mathcal{A}}{2} & b_i = -1 \end{cases} \tag{3.11}$$

Hence, based on Eqn. 3.11, in order to decode a despread signal correctly in synchronous CSGN, the maximum fault that can be tolerated at the receiver in terms of $\mathcal{A}$ is:

$$Q^*(t) = |1 - b_i| \leq \frac{\mathcal{A}}{2}. \tag{3.12}$$

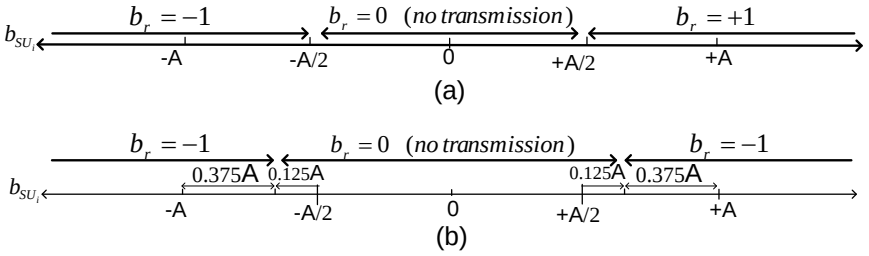


Figure 3.4: Convert received despread amplitude to bit in receiver in a) synchronous CSGN, b) asynchronous CSGN.

3.3.3.2 Fault tolerance threshold in asynchronous CSGNs

However, in an asynchronous mode shown in Fig. 3.4-b, using the proposed sub-OCSs will cause, at most, $\frac{\mathcal{A}}{4}$ of the amplitude to be altered in a time instant. This will affect both positive and negative amplitudes equally ($\frac{\mathcal{A}}{8}$ or $0.125\mathcal{A}$ in positive voltages and $\frac{\mathcal{A}}{8}$ or $0.125\mathcal{A}$ in negative voltages). This effect, which is indicated by the interferer voltage by asynchronous OCS and sub-OCSs at time instant t in the figure, makes the acceptable decoding amplitude interval smaller, from 50% in the synchronous mode, to 37.5% in the asynchronous mode.

In the asynchronous CSGN, all the symbols and notations used are similar to the synchronous case except for the term $\mathcal{A}_p(t)$ that represents the interference caused by asynchronous transmission, where N_{asy} represents the number of nodes causing jitter interference. Throughout the proposed CSGN, the noise power is assumed to be 1 (Fig. 3.4-b).

$$C(t) = \sum_{i=1}^{L} \mathcal{A}_i(t) + \sum_{g=1}^{NSU=(log_4 L)-1} \mathcal{A}_g(t) + \sum_{p=1}^{N_{asy}} \mathcal{A}_p(t) + \eta_t \tag{3.13}$$

$$b_i = \sum_{t=1}^{OCSLength=L} C(t).OCS_i(t). \tag{3.14}$$

Let $f_{b_i}^{min}$ denote the minimum required amplitude in an asynchronous receiver to decode the data bits from the despread signal correctly. Then,

$$f_{b_r}^{min} = b_i + \sum_{p=1}^{N_{asy}} \mathcal{A}_p(t) = 0.5\mathcal{A} + 0.125\mathcal{A}$$

$$= 0.50\mathcal{A} + 0.125\mathcal{A} = 0.625\mathcal{A}. \tag{3.15}$$

Let $Q^*(t)$ denote the maximum fault tolerance in an synchronous transmission. Then,

$$if \begin{cases} 0.625\mathcal{A} < b_i < +\mathcal{A} & b_i = +1 \\ -0.625\mathcal{A} \leq b_i \leq +0.625\mathcal{A} & b_i = 0 \\ -\mathcal{A} < b_i < -0.625\mathcal{A} & b_i = -1 \end{cases} \tag{3.16}$$

Thus, based on Eqn. 3.16, considering time shift and jitter in OCSs and sub-OCSs, a despread signal can be decoded at a receiver in an asynchronous CSGN, if the fault is bounded by Eqn. 3.17 in terms of $\mathcal{A}$.

$$Q^*(t) = |1 - b_i| \leq 0.375\mathcal{A}. \tag{3.17}$$

To bring up a tangible example, assume that sending a data symbol with an OCS of length L results in an amplitude of $\mathcal{A}$ or $-\mathcal{A}$ at the receiver. Then, any corruption in a pattern of L will result in a change in $\mathcal{A}$, relative to the corrupted fraction of L. There is a linear relation between L and $\mathcal{A}$. Additionally, as we have seen, there is a way for the primary and secondary receivers to decode the received signal if the total noise and interference result in atmost 37.5% corruption of the decoded waveform (or 37.5% of the chip length), based on the fault-tolerance threshold presented above. In other words, if the decoded value (the value which is decoded at the receiver from the received signal) is between ($\mathcal{A} - 0.375\mathcal{A}$, $\mathcal{A} + 0.375\mathcal{A}$), it will be interpreted as $\mathcal{A}$ providing an error tolerance of 37.5% of the OCS length. For instance, if $\mathcal{A} = 256$, any decoded value at the receiver ranging from ($\mathcal{A} - 0.375\mathcal{A}$)= 160 to ($\mathcal{A} + 0.375\mathcal{A}$) = 352 will be interpreted as 256 at the receiver. Likewise, any decoded value ranging from ($-\mathcal{A} - 0.375\mathcal{A}$) = -352 to ($-\mathcal{A} + 0.375\mathcal{A}$) = -160 will be interpreted as -256 at the receiver. In case of getting a decoded value between -160 to

160, the receiver will ignore the received data (equivalent to receiving 0) assuming that the sender has been silent (the received signal could be generated due to environmental noise). It should be noted that any value more than ($\mathcal{A}+0.375\mathcal{A}$) or less than (-$\mathcal{A}-0.375\mathcal{A}$) will be mapped to $+\mathcal{A}$ and $-\mathcal{A}$, respectively. It is worth mentioning that similar orthogonal codes, such as Gold, can also tolerate a timing jitter of 50% of the OCS length [80].

3.4 Evaluation

3.4.1 Network Setup

In this section, we evaluate the proposed scheme for improving number of SUs in underlay CSGNs by considering non-CDMA-based PNs. In the proposed CSGN simulation, all SMs are static and are equipped with omni-directional antennas with radius r in a square planar area. It is also assumed that SMs are randomly placed based on a uniform probability distribution function with a planar node density χ, where χ is the number of SMs per square meter. We also know that the number of SMs (n) in an area of πr^2 square meters has a Poisson distribution of [246]:

$$p(\chi, n, r) = \frac{(\chi \pi r^2)^n e^{-\chi \pi r^2}}{n!} \tag{3.18}$$

When a communication channel in the PN is being utilized by a PU, the SUs have to spread their data by utilizing full OCSs or sub-OCSs with a power below the PUs' tolerable noise floor. Thus, the sum of the SUs' transmission power should be in the noise margin of TV which is lower than $2.511W$ [72, 152]. When the PN is considered non-CDMA-based, it can use any multiplexing scheme such as TDMA or FDMA. It goes without saying that, as PN is not CDMA-based, PUs do not use OCSs or sub-OCSs for data transmission. Therefore, all OCSs and sub-OCSs will be available to be utilized by the SUs, resulting in considerably more supported SUs. Each SN's GW selects an available OCS or sub-OCS for each SU in the respective hierarchies. Lower hierarchies are locally significant and each SM adjusts

appliances' transmission power based on its distance to appliances which should be less than 5 mW. This level of power would be enough to transmit data between appliances and SM in indoor environments [152]. SUs can set their transmission power during the coordination phase on the CC based on the attenuation of the received signal [263, 264, 104]. Despite the challenge of power control in CDMA-based communication [182], we will show by means of our simulation results that the NSUs can be increased by utilizing multi level of power in a hierarchical structure of SN.

SMs assign one OCS to each appliance to prevent interference. Also, in second hierarchy GWs assign one sub-OCS, from existing sub-OCSs in second hierarchy, to each SM. In a same way, SGWs assign one sub-OCS to each GW. Hence, there is a particular synergy in OCS and sub-OCS assignment to devices in each supercluster.

3.4.2 Simulation Results

We evaluate our proposal in an underlay CSGN that supports TV network as a PN. We would like to argue that the number of SUs, and consequently the number of OCSs and sub-OCSs, are independent from the infrastructure and are solely bounded by the parameters of the proposed architecture. The OCS length used by SUs will determine the hierarchy of possible sub-OCSs in the SN. However, in order for SUs not to interfere with PUs, SUs transmission power plays a key role [101]. The task of assigning an OCS or sub-OCS to a SU, which is done independently by SM or GW, will add extra set up and processing time on the network. GW assigns a unique sub-OCS to each SM periodically when SM has data to send to higher hierarchy in each time interval. As each SM in a cluster is assigned a unique sub-OCS by the GW, there is no interference among SM communications within the cluster. Also, neighboring GWs collaborate with each other but duplicate sub-OCSs are not assigned to neighboring clusters.

We evaluate our proposed CSGN architecture using MATLAB simulations under several different network conditions. In one set of experiments, we simulate our network in a $1000m$

Table 3.2: Simulation Parameters for Proposed SU

Parameter	**Value**
Simulation area	1000 m × 1000 m
GWs' sensing range radius	100 m
SUs' power control model	Out-band CC
Tx power on CSMA/CA-based CC	Coordinated model by GW and SGW
Tx power on CDMA-based data channel	$V_1 \leq 5mW$ (3kbps in building), $V_2 \leq 10mW$ (5kbps, $\leq$ 100 m), $V_3 \leq 20mW$(5kbps, 100-200 m)[152]
Number of available secondary users in CSGN	$L + \sum_{i=1}^{(log_4 L)-1} \frac{L}{4^i}$
Carrier frequency	TV band (VHF), 54-72 MHz, 76-88 MHz
Channel bandwidth	18 MHz
Bit rate on shared network	64 Kbps
Bit rate for each connection	3 kbps
Orthogonal Chip Code generator	PCC and Golay (Asynchronous OCSs)
OCS and sub-OCS Size in CDMA based SN, non-CDMA based PN	Varies from 4 to 1024 bits
Node Placement	Random
Radio propagation model	Free Space
ReceiverGWs hight	$h_r = 8m$
Maximum PN Interference tolerance	$3.2(\log_{10}(11.57(h_r)))^2 - 4.97\ dB$
TV Noise Margin (tolarable noise floor)	$<$ 2.511 W
Number of iterations	100

$\times\ 1000m$ area. Our first goal is to verify the effectiveness of our approach in increasing the number of supported SUs. Simulation parameters can be found in Table 3.2 and results are available in Fig. 3.5-a and Fig. 3.5-b. For simplicity, we currently only assume the free space propagation model.

As the hierarchical organized SN is CDMA-based, while the PN is not CDMA based, SUs are able to utilize all OCSs ($=L$) and sub-OCSs ($= \sum_{i=1}^{(log_4 L)-1} \frac{L}{4^i}$). Thus, the total NSU in the proposed method is $L + \sum_{i=1}^{(log_4 L)-1} \frac{L}{4^i}$.

Our simulation results for NSU and delay in the proposed CSGN consider three different time synchronization states that have an effect on interference on SN's channels: 1) All data

transmissions on SN are synchronous. 2) All data transmissions on SN are asynchronous. 3) All data transmissions on SN are asynchronous with a cyclic shift of OCSs and sub-OCSs. More synchronization difference creates more interference in the SN. Each state is analyzed with $SINR = 20\ dB$ and four different OCS lengths 16, 64, 256, and 1024. Simulation results confirm our analysis in Section 3.3.3 showing that using sub-OCSs by SUs will not increase BER on SUs' transmission in lower hierarchies that are using longer OCS or sub-OCSs. It can be seen from the simulation results that asynchronous transmission by SUs in different hierarchies will impose more interference on the SN than the interference caused by synchronous data transmission by SUs. Figure 3.5-a illustrates the average number of secondary users in our proposed underlay CSGN in synchronous, asynchronous, and asynchronous with cyclic shift for 100 times iteration when $SINR = 20\ dB$. Each average value contains information about NSU for all iterations with $SINR = 20\ dB$. In Fig. 3.5-b, number of dropped packets is shown when fixed number of SUs ($NSU = L + \sum_{i=1}^{(log_4 L)-1} \frac{L}{4^i}$, $SINR = 20\ dB$) are utilizing SN for data transmission. Number of dropped packets (in all synchronous, asynchronous, and asynchronous with cyclic shift states) in PCC and Golay algorithms as shown in Fig. 3.5-b are simulated for full load of data in SCGN. In fully loaded simulation scenarios, all OCSs and sub-OCSs are utilized by SUs. Thus, results in the figure indicate the number of dropped packets when all OCSs and sub-OCS are used by SNs in the proposed CSGN. For instance, when $L = 1024$, CSGN has 1024 active OCSs and 340 active sub-OCSs when different scenarios have various numbers of dropped packets. In Fig. 3.5-b, we can see the number of dropped packets is increased when we increase the length of OCSs. The interference among OCSs and sub-OCSs is the reason for increase in the dropped packets, because there are more simultaneous SUs transmissions. As shown in the Fig. 3.5-b, the number of extra sub-OCSs that are generated by longer OCSs is always bigger than number of dropped packets. Thus, increasing the length of OCS in SNs that need to support more SUs, provide more resource in SNs.

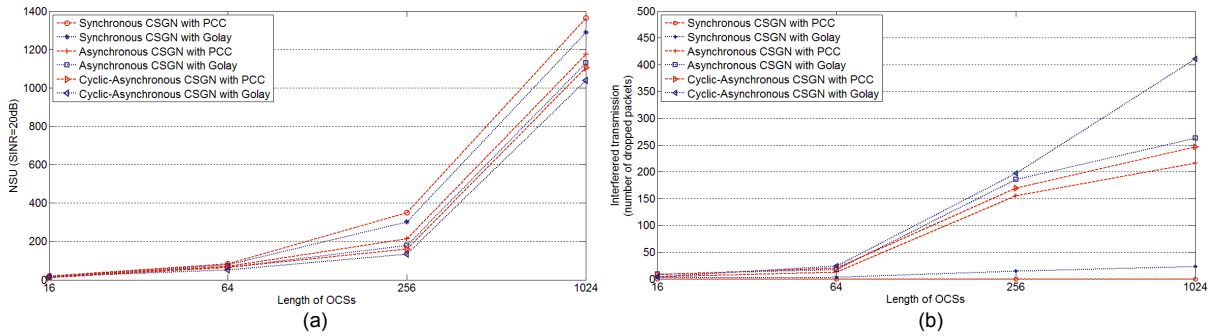


Figure 3.5: a) Number of SU (OCSs plus sub-OCSs) in synchronous SN, asynchronous SN, and asynchronous SN with cyclic shifts in SUs' data transmissions. b) Number of dropped packets when NSU is same (NSU=L) with and without utilizing proposed CSGN.

Low latency for CSGN data transmission is an important feature of *Quality of Services (QoS)* in future smart grid networks. Real-time communications and transmitting large data are requirements in future SGNs, although these requirements will increase data transmission latency in SNs, concurrently [248]. Thus, latency can be one of the drawbacks in such networks that is increased with size of data transmission and QoS requirements. All SUs who utilized OCSs and sub-OCSs in proposed underlay CSGN are able to spread data simultaneously. Thus, increasing NSU and parallel transmission on SN decrease the latency by means of parallel transmission on the same channel. Also, this simultaneous data transmission makes channel available to other users when SUs have already transmitted their data in the same time. In this part of simulation we assume that the number of SU in CSGN is fixed. Also, CSGN utilized sub-OCSs or extra SN's capacity to transmit data by the existing SUs instead of using sub-OCSs to increase NSU. Therefore, each round of simulation, for various numbers of SUs, is simulated with and without utilizing sub-OCSs. For instance, when CSGN has 1024 SUs and each SU wants to transmit ten megabyte of data simultaneously with other SUs, CSGN without sub-OCSs has 1024 OCS (1024 simultaneous SUs). Therefore, in proposed CSGN, SN is able to use 340 sub-OCSs, when $L = 1024$, to increase bit-rate in SNs. The increasing bit-rate in SNs' data transmission will relatively decrease the propagation delay. Therefore, the extra 340 sub-OCSs will share among all SUs and

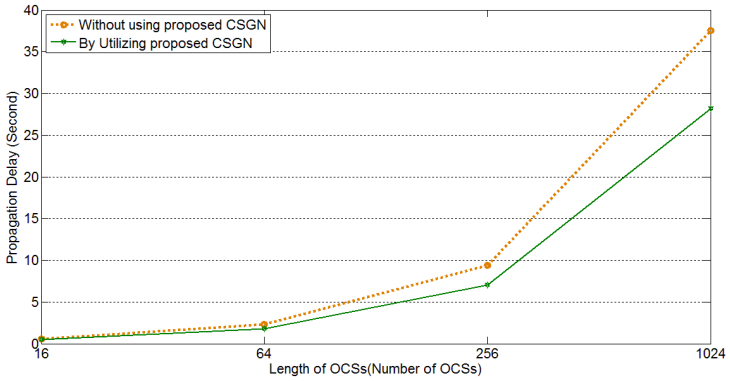


Figure 3.6: The comparision on saved time in proposed CSGN with utilizing and a plain CDMA-based SN without utilizing sub-OCSs to send 10 megabyte of data by SM to UCC.

approximately each sub-OCS will be shared among three SUs to transmit data and decrease the propagation delay for a fixed amount of data. In this part of simulation, we assume the bit-rate in each connection from appliances to UCC is 3 *kbps* on CSGN CDMA-based channel. Figure. 3.6 illustrates the time saved using the proposed method without considering environmental noise to reduce latency and sending ten megabyte of data from appliances to UCC when network is worked fully loaded.

Based on Fig. 3.5-b and Fig. 3.6, we can see the propagation delay will be much better in all situations even as the number of dropped packets increases. However as the value of L increases, there is an increase in the number of dropped packets which in turn affects the transmission rate resulting in an increased propagation delay even when sub-OCSs are used.

We would remark that our proposed method increases NSU by the number of utilized OCSs and sub-OCSs, and it does not depend on the network infrastructure. Changing the carrier will not affect the number of SUs although it might alter the noise and the interference level[2].

[2]The content of this chapter has appeared in [21].

CHAPTER 4
SEER GRID: PRIVACY AND UTILITY IMPLICATIONS OF TWO-LEVEL LOAD PREDICTION IN SMART GRIDS

As part of the future smart electricity grid initiative, a smart grid communication network (SGN) is a large-scale integration of information and communication technologies within the electricity generation, transmission, and distribution systems of the traditional electricity grid. A combination of various *smart* technologies at different levels of the SGN promotes efficiency, reliability and stability in operations of the smart grid. One indispensable piece of technology in a SGN is a smart meter (SM) which collects and periodically reports the energy usage or consumption information of the customers to the electric (a.k.a. utility) company (EC), which in turn facilitates highly efficient energy generation and distribution and helps the EC to cope with changes in energy demand and supply. The monetary and natural resource savings due to the improved efficiency is a major factor in the fast growing adoption of SMs, with predictions that about 800 million SMs will be in use globally by 2020 [232]. Despite their tremendous importance in a SGN, SMs can also be easily exploited by malicious adversaries (including the EC) who may attempt to infer private customer information from reported energy consumption patterns, such as occupancy of the house [129], specific appliances being used [253], and even daily routine of the residents [202][141].

Various techniques for overcoming privacy issues due to the energy usage information generated and shared by SMs have been proposed in the research literature, and these solutions have primarily followed one of the following two approaches: (i) completely obscure the individual SM data from the perceived adversary, or (ii) hide privacy-sensitive signatures or patterns from the individual SM data by perturbation or down-sampling. In the first direction, protocols that take advantage of the homomorphic properties of public-key cryptographic algorithms to perform neighborhood-level aggregation of SM data have been proposed in the literature [137][78]. These protocols enable the EC to learn the actual aggre-

gated energy consumption information (at a neighborhood level) without leaking individual customer-specific information to the aggregator. In the second direction, many approaches have been proposed to efficiently perturb energy consumption data in order to meet certain privacy requirements. In-residence storage batteries have been employed to flatten or mask variances in the load or electricity usage information [120][162]. Similarly, controlled perturbation [98][191][207] and down-sampling [59][160] of the energy consumption data to mask specific load signatures of appliances have also been attempted.

However, as pointed out by [207] and [59], the degree of correlation between the actual energy consumption and the data output by a privacy-preserving technique typically characterizes a trade-off between privacy and utility (or usefulness). Higher correlation with the actual ground-truth makes the perturbed data more useful but reveals private information, whereas lower correlation (or increased perturbation) is good for privacy but reduces data usefulness or utility. As protocols following the first approach do not really perturb the electricity consumption data, the utility of the data (or any function computed from the data) is high. Also, as this data is cryptographically obscured from the aggregator, there is no leakage of private customer information. However, protocols using public-key cryptography are non-trivial to implement in practice and have very high computation and communication overhead [56]. Perturbation mechanisms, such as the ones using storage batteries [120][162], are effective in masking private usage patterns, but installing and maintaining large capacity batteries in every household is shown to be economically non-viable [9]. Similarly, Dong et al. [59] show that performance of smart grid operations can degrade due to reduction in sampling frequency. Other perturbation mechanisms, such as [207], that attempt to strike a good balance between privacy and data-utility by masking or suppressing specific appliance signatures assume that individual appliance electricity consumption information is readily available (or can be easily separated from the overall data) which may not always be feasible. Given the above state-of-the-art, we feel that both data hiding and data perturbation

approaches have inherent limitations, which motivates us to explore alternate paradigms (beyond hiding and perturbation).

Our goal in this work is to explore alternate practical designs for privacy-sensitive generation and sharing of energy consumption information from the SMs to the EC which enables effective operation of the EC in terms of accurately predicting future demand and electricity generation and distribution. In order to achieve this goal, we move away from the classical perturbation/data-hiding techniques and use learning-based prediction mechanisms to generate (or predict) energy consumption patterns shared by SMs. Our prediction mechanism will replace variances in the individual household-level actual energy consumption patterns (which is typically indicative of loads) with relatively smoother patterns that are free of load signatures but accurate enough to be useful in predicting energy consumption at the neighborhood level (which is the one that is actually used by the EC). Due to this, privacy-sensitive inference attacks will be much harder on the household-level data shared by the SM without significantly impacting the demand-response and electricity generation/distribution calculations at the EC.

With Seer Grid[1], we propose a household-level prediction scheme comprising of a statistical learning algorithm (trained using past consumption pattern of the household) which predicts an entire day's electricity consumption pattern one day in advance[2]. This prediction can be performed locally on the SM, on a local energy management unit or on a computing device that connects to such a unit. The household electricity consumption pattern predicted locally at the SM, with the load or appliance signatures masked or flattened, is then reported to an aggregator or data concentrator (referred here as a cluster head or CH) at the beginning of each day. All SMs within a neighborhood or cluster report their energy consumption predictions to their respective CH who in turn forwards an aggregated pre-

[1]According to Oxford Dictionary, *seer* is "a person who is supposed to be able, through supernatural insight, to see what the future holds." Through two-level energy prediction we enable insight into future demands, while simultaneously promoting consumer privacy.

[2]The content of this chapter has appeared in [23].

diction (as described below) to the EC. As our localized prediction flattens or eliminates sharp variations (which may indicate load signatures) in the predicted consumption at the SM or household level, this difference can add up significantly while aggregating predictions for multiple households in a neighborhood or a cluster. This can reduce the accuracy of the aggregated prediction, thereby adversely impacting its utility or usefulness to the EC. In order to restore this utility lost due to prediction at the SM level, we introduce a second level of energy load prediction at the CH for compensating the difference in the aggregate of predicted and actual energy usage of individual SMs in the cluster. CH performs the spike prediction based on past energy consumption pattern of the entire neighborhood or cluster, and then reports the result of the second level prediction, in addition to summation of first level predictions, to EC just before beginning of each day. EC can then use this cluster or neighborhood-wide load prediction to efficiently control electricity generation and distribution. To ensure fail-proof operation of the SGN in case of major prediction errors, we also incorporate a privacy-preserving reporting of the aggregated variance between actual and predicted energy consumption of all SMs in the cluster. Also, the entire framework is computationally efficient, allowing it to work in real-time (we assume intervals of 5 minutes for SM reporting, but can be easily applied to shorter intervals).

Seer Grid's two-level prediction mechanism offers several advantages over traditional privacy-preserving energy data reporting schemes in the literature. Unlike data hiding schemes that require several encryption operations at the SM or household level per day (once every reporting interval), our prediction and reporting operation is performed just once per day. Moreover, Seer Grid is communication-efficient (as no additional data or overhead needs to be communicated), does not require any specialized hardware (e.g., storage batteries) and does not need access to appliance-level consumption patterns. In rest of the chapter, we first describe the generic SGN architecture and capabilities of the assumed adversary in Section 4.2. In Section 4.4, we discuss the details of the proposed Seer Grid architecture and it's operation. In Section 4.5, we evaluate the Seer Grid architecture by

performing extensive experimental simulations using real smart meter data. We empirically measure the correlation between predicted and actual consumption patterns at each level, using standardized metrics, to support our proposition of a practical SGN architecture which improves both privacy and utility of SM data. Comparative evaluation shows that Seer Grid's two-level prediction provides better privacy and utility, compared to just SM level perturbation techniques.

4.1 Related Work

Multiple schemes have been proposed for short term [37] [112][111] and long term [132] load prediction at cluster level. Sevlian and Rajagopal [216] proposed short term electricity load forecasting on varying levels of aggregation, and concluded that aggregating more customers improves the relative forecasting performance only up to a specific point. Recently, smart meter based short-term load forecasting was proposed [82][224], as a household's historic energy consumption pattern is a better predictor of peak load than any other observable variables. In contrast, Seer Grid uses two level of prediction to retain the privacy benefits of aggregation, and utility benefits of individual household prediction.

There have also been extensive research efforts that attempt to address privacy issues related to SM data release. Li et al. [137] proposed using Paillier cryptosystem's homomorphic property for distributed energy consumption data aggregation from SMs, where the EC is able to know only the aggregated data upon decryption of the aggregated cipher. Garcia and Jacob [78] proposed the combination of additive secret sharing algorithms and cryptosystems with homomorphic property, in order to compute the aggregated energy consumption of a given set of users (for example, in a cluster) in a privacy preserving fashion. However, cryptosystems with homomorphic properties induce a large computational overhead on the SMs, and real-time reporting in short time interval is impractical [56]. Alternatively, McLaughlin et al. [162] proposed a non-intrusive load leveling model by using large capacity batteries. Large capacity batteries smoothen the energy consumption pattern and effectively help in

hiding load signatures contained in actual consumption pattern. However, large batteries are economically inconvenient [9] due to their high capital cost and low energy-efficiency.

Privacy through anonymization tries to unlink the energy usage data from individual SMs [61]. However, anonymization may turn out to be ineffective, as Jawurek et al. [115] and Faisal et al. [68] demonstrated the feasibility of using household anomaly detection and behavior pattern to de-anonymize SM data. With limited computational capabilities and practicality in mind, researchers suggested the use of perturbation techniques for hiding load signatures. Consumer privacy can be preserved by deliberately introducing error into the energy usage data [98][191][161][219], and such perturbation techniques often try to achieve differential privacy in order to reduce the privacy-utility trade-off [207]. However, the privacy-utility trade-off of SM data perturbation techniques can still be significant [211], which may not be admissible to ECs. In this chapter, the proposed Seer Grid architecture aims to decrease the privacy-utility trade-off by using a two-level prediction scheme.

4.2 The Traditional SGN Architecture

We base our work on one of the most popular SGN architecture consisting of a three-level hierarchical network (Figure 4.1). At the lower level are the SMs, physically located in households of end-users or customers. At the middle level, each neighborhood has a CH, and SMs report energy consumptions to CH. Situated at the higher level is the EC, to which all CHs report aggregated load of their respective neighborhood. The load reporting from all CHs aids EC in optimizing generation and distribution of electricity. In real-life implementation, CH may be owned and operated by a third party or by the EC itself.

We assume a passive adversary who may try to infer personal information of customers based on accessible energy consumption data. Motivations can vary widely, such as financial gain from advertising agencies, health insurance companies trying to find unhealthy lifestyle of insurees, etc. If given access to actual energy consumption data, the adversary is computationally capable of carrying out inference attacks by analyzing the data. We also assume

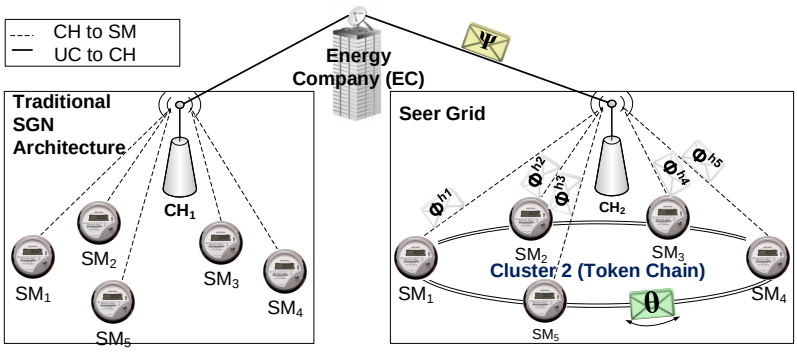


Figure 4.1: Traditional SGN architecture on the left, and our proposed SGN architecture (details in Section 4.4) on the right.

that the adversary can access energy consumption data reported to the CH and/or to the EC. However, CH and EC must be honest and cooperative for the protocol to function properly. Thus, CH and EC can be considered as honest but curious adversaries. We also consider any eavesdropper (eavesdropping communication between SM and CH, or CH and EC) as an adversary. All SMs are assumed to honestly and correctly follow the proposed protocol. As a result, we do not consider collusion attacks between SMs and CH, or between SMs and EC.

4.3 Technical Background

We carefully analyzed various statistical learning algorithms for predicting energy consumption patterns, in order to identify the algorithm apposite for preserving only the desired characteristics of the consumption pattern data. In this section we first detail the constituents and properties of the consumption pattern data, followed by a discussion on how we select prediction algorithms for SM and CH.

4.3.1 Prediction at SM Level

Traditional SMs report energy usage data to EC in short time intervals, where each reporting conveys the energy used since last reporting. Let us denote the actual daily SM

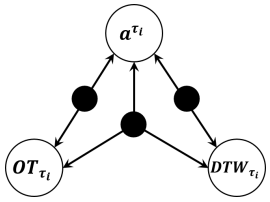


Figure 4.2: Interaction between $a^{(\tau_i)}$ and OT_{τ_i} is 2-way. Interaction between $a^{(\tau_i)}$ and DTW_{τ_i} is also 2-way. And there exists a 3-way interaction between $a^{(\tau_i)}$, OT_{τ_i} and DTW_{τ_i}. The prediction model must learn these interactions in order to make effective predictions.

energy consumption pattern of a household h_k as $\mathcal{A}^{h_k} = \{a^{(\tau_1)}, a^{(\tau_2)}, \cdots, a^{(\tau_n)}\}$, where $a^{(\tau_i)}$ is the energy consumed since $a^{(\tau_{i-1})}$. The goal of using a prediction model at the SM is to predict a pattern $\Phi^{h_k}_{day_j} = \{p^{(\tau_1)}, p^{(\tau_2)}, \cdots, p^{(\tau_n)}\}$, such that there is a high overlap between $\Phi^{h_k}_{day_j}$ and $\mathcal{A}^{h_k}_{day_j}$, but $\Phi^{h_k}_{day_j}$ is free of specific load signatures (such as spikes and plateaus). Predictive modeling leverages statistics to predict outcomes, i.e., the forecast of a day's consumption pattern is based on collection of past $\mathcal{A}^{h_k}$ (let's say for m days). After analyzing various factors that affect consumers' energy usage, we identified the input variables critical to the outcome of the prediction model as [i] power usage history in each time interval ($a^{(\tau_i)}$), [ii] outdoor temperature in each interval (OT_{τ_i}), and [iii] day and time of the week (DTW_{τ_i}). Each day of the week is considered differently so as to improve prediction based on weekly routines [235]. All interactions present between these three variables is represented in Figure 4.2.

Popular time series forecasting uses a statistical model for predicting future values based on previously observed values. However, basic time series forecasting does not capture the complex interaction between different input variables, resulting in inferior forecasting. Due to the highly complex interactions and some dependencies between input variables, multi-class classification and regression analysis will also result in non-optimal prediction. To achieve better prediction results, we employ structured prediction using supervised machine learning techniques. Among candidate machine learning techniques for structured prediction, we decided to use multi-layered perceptron (MLP) because it is specifically designed to discover

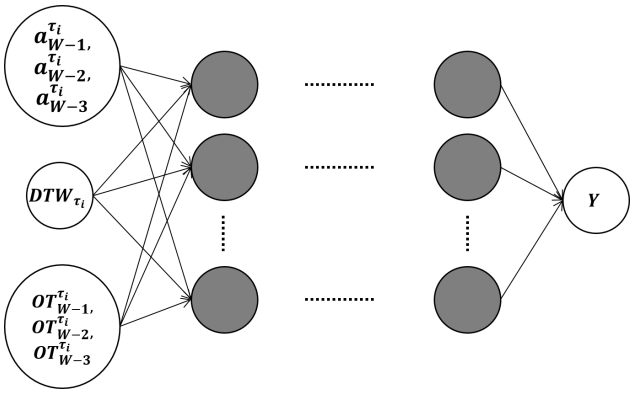


Figure 4.3: The abstract structure of the MLP used of learning and prediction. $a^{(\tau_i)}_{W-1}$, $a^{(\tau_i)}_{W-2}$ and $a^{(\tau_i)}_{W-3}$ is the power usage in the τ_ith interval from last 3 weeks; DTW_{τ_i} represents day and time of the week,; and $OT^{(\tau_i)}_{W-1}$, $OT^{(\tau_i)}_{W-2}$ and $OT^{(\tau_i)}_{W-3}$ are the outdoor temperature (in Fahrenheit) in the τ_ith interval from last 3 weeks.

the complex interactions among input variables. MLP is a feed-forward artificial neural network (ANN) model that uses a nonlinear activation function to map sets of input data onto a set of appropriate outputs. MLPs consisting of three or more layers (input, output, and one or more hidden layers) is called a deep neural network, where each node in one layer connects with a certain weight w_{pq} to every node in the following layer. The error in output of a node q in the nth training data point is represented as $e_q(n) = d_q(n) - y_q(n)$, where d is the target value and y is the value produced by the perceptron. The calculated error for each training data point is used to make corrections to the weights of the node as $\mathcal{E}(n) = \frac{1}{2}\sum_q e^2_q(n)$, which in turn minimizes the error in the entire output of the ANN. Change in each weight during an epoch is calculated as $\Delta w_{qp}(n) = -\eta \frac{\partial \mathcal{E}(n)}{\partial v_q(n)} y_p(n)$, where y_p is the output of the previous neuron and η is the learning rate.

In the learning phase of our MLP execution, for each epoch we input power usage history of last three weeks recorded in 5 minute intervals. Outdoor temperature for the corresponding interval and day of the week is also fed in each epoch (Figure 4.3). The output of the ANN is a structured object (Y) containing multiple possible $\Phi^{h_k}_{day_j}$ for next day. Given that the next day's temperature forecast and day of the week is known, the structure object is parsed

for the matching $\Phi^{h_k}_{day_j}$. More details about the MLP specifications used in our simulation experiments can be found in Section 4.5. Additionally, we use a low-pass filter over SM training data which shaves any load pattern for energy consumption above 4kW. Shaving spikes in the energy consumption pattern eliminates outliers in the training data.

4.3.2 Prediction at CH Level

The purpose of using prediction at SM is to remove specific load signatures (such as spikes and plateaus) from $\mathcal{A}^{h_k}_{day_j}$. Although the missing spikes and plateaus from the SM of one household represent a minuscule amount of energy for the grid, spikes and plateaus from multiple households in a cluster can add up to a significant amount of unpredicted energy, which can endanger proper functioning of the electricity grid. Thus, we introduce another level of statistical prediction at the CH based on historical load profile of the cluster, while also factoring in individual predictions from all SMs in the cluster $\{\Phi^{h_1}_{day_j}, \Phi^{h_2}_{day_j}, \Phi^{h_3}_{day_j}, \ldots\}$. The proposed algorithm (Algorithm 5) uses average of difference between past load predictions and actual loads of the entire cluster ($\Lambda_{day_d} = \{\lambda^{(\tau_1)}, \lambda^{(\tau_2)}, \cdots, \lambda^{(\tau_n)}\}$), in order to complement missing loads. The output of the algorithm $\Psi_{day_j} = \{\psi^{(\tau_1)}, \psi^{(\tau_2)}, \cdots, \psi^{(\tau_n)}\}$ is the prediction for the whole cluster reported to CH, where $\psi^{(\tau_i)} = \delta^{(\tau_i)} + \sum_k p^{(\tau_i)}$ and $\delta^{(\tau_i)} = \frac{\sum_{d=j-m}^{d=j-1}\{\lambda^{(\tau_i)}_{day_d} - \sum_k p^{(\tau_i)}_{day_d}\}}{m}$. Although trivial, the algorithm can achieve high accuracy.

4.4 Seer Grid

The primary distinction between the traditional SGN and Seer Grid is that, in Seer Grid SMs never report their actual energy consumption data; they report predicted energy consumption pattern instead. Similar to the traditional SGN architecture, Seer Grid also consists of a three level hierarchical network (Figure 4.1). At the lower level are the SMs, physically located in households. At the middle level, each neighborhood has a CH, and SMs report next day's predicted energy consumption patterns to CH. A second level prediction is performed by CH on the aggregated predicted patterns reported by all SMs belonging

Algorithm 4 Prediction algorithm executed by CH.

Prediction Function (for day j)

Define new $\Psi_{day_j} = \{\psi^{(\tau_1)}, \psi^{(\tau_2)}, \cdots, \psi^{(\tau_{288})}\}$

for $k = 1$ to K (K households in the cluster) **do**
  $\sum p^{(\tau_i)}_{day_j}$
end for

for $i = 1$ to 288 (5 minutes time intervals for 24 hours) **do**
  for $d = 1$ to m (m days to historical data) **do**
    $\delta^{(\tau_i)} = \lambda^{(\tau_i)}_{day_d} - \sum_k p^{(\tau_i)}_{day_d}$
  end for
  $\delta^{(\tau_i)} = \frac{\delta^{(\tau_i)}}{m}$
  $\psi^{(\tau_i)} = \delta^{(\tau_i)} + \sum_k p^{(\tau_i)}$
end for

Report Ψ_{day_j} to CH

to the cluster. At the higher level is the EC, to which all CHs report the second level predicted energy forecast for their respective neighborhood. The predicted forecast from all CHs aids EC in optimizing generation and distribution of electricity. We assume that the CH is capable of measuring the actual electricity usage of the whole cluster for a given time interval[3], which is used to the form statistics used in the cluster level prediction. This is a reasonable assumption because all cluster level energy forecasting schemes [37] [112][111] [132] rely on readings from a cluster level electricity meter. We also consider billing once as a month event, which can be done independently.

We carefully analyzed various statistical learning algorithms for predicting energy consumption patterns, in order to identify the algorithm apposite for preserving only the desired characteristics of the consumption pattern data. We first detail the constituents and properties of the consumption pattern data, followed by a discussion on how we select prediction algorithms for SM and CH. Readers should note that we use the following prediction algo-

[3]CH is assumed to be equipped with a cluster level meter which measures energy being withdrawn by the entire cluster. When CH measures the electricity usage of the entire cluster, it does not violate privacy of individual household because of aggregation.

rithms as an example, in order to demonstrate the benefits of Seer Grid. Other suitable prediction algorithms can be used as well.

4.4.1 Prediction at SM Level

In a traditional SGN, SMs report energy usage data in short time intervals, where each report conveys the energy consumed since the last reporting. Let us denote the actual daily SM energy consumption pattern of a household h_k as $\mathcal{A}^{h_k} = \{a^{(\tau_1)}, a^{(\tau_2)}, \cdots, a^{(\tau_n)}\}$, where $a^{(\tau_i)}$ is the energy consumed since $a^{(\tau_{i-1})}$. In Seer Grid, the goal of using a prediction model at the SM is to predict a pattern $\Phi^{h_k}_{day_j} = \{p^{(\tau_1)}, p^{(\tau_2)}, \cdots, p^{(\tau_n)}\}$, such that there is a high overlap between $\Phi^{h_k}_{day_j}$ and $\mathcal{A}^{h_k}_{day_j}$, but $\Phi^{h_k}_{day_j}$ is free of specific load signatures (such as spikes and plateaus). Predictive modeling leverages statistics to predict outcomes, i.e., the forecast of a day's consumption pattern is based on collection of past $\mathcal{A}^{h_k}$ (let's say for m days). After analyzing various factors that affect consumers' energy usage, we identified the input variables critical to the outcome of the prediction model as [i] power usage history in each time interval ($a^{(\tau_i)}$), [ii] outdoor temperature [4] in each interval (OT_{τ_i}), and [iii] day and time of the week (DTW_{τ_i}). Each day of the week is considered differently so as to improve prediction based on weekly routines [235]. All interactions present between these three variables is represented in Figure 4.2.

Classical time series forecasting techniques [25] use a statistical model for predicting future values based on previously observed values. However, such basic time series forecasting does not capture the complex interactions between different input variables, thus resulting in inferior forecasting. Due to the highly complex interactions and some dependencies between input variables, multi-class classification and regression analysis will also result in non-optimal prediction. To achieve better prediction results, we employ structured prediction using supervised machine learning techniques. Among candidate machine learning

[4]Many older SMs are not equipped with temperature sensors. In such cases, appropriate outdoor temperature values can be provided by the CH or EC.

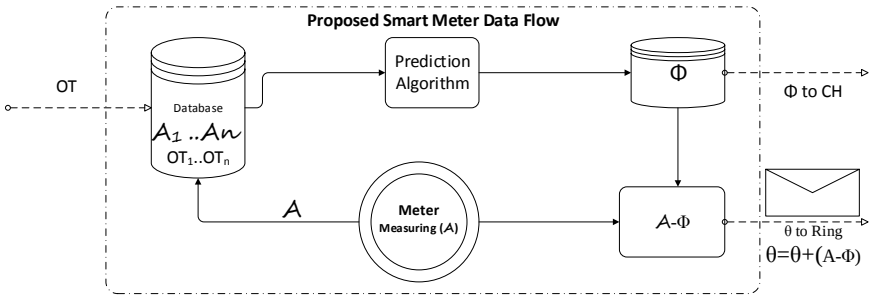


Figure 4.4: Proposed SM data flow.

techniques for structured prediction, we decided to use multi-layered perceptron (MLP) because it is specifically designed to discover the complex interactions among input variables. MLP is a feed-forward artificial neural network (ANN) model that uses a nonlinear activation function to map sets of input data onto a set of appropriate outputs. MLPs consisting of three or more layers (input, output, and one or more hidden layers) is called a deep neural network, where each node in one layer connects with a certain weight w_{pq} to every node in the following layer. The error in output of a node q in the nth training data point is represented as $e_q(n) = d_q(n) - y_q(n)$, where d is the target value and y is the value produced by the perceptron. The calculated error for each training data point is used to make corrections to the weights of the node as $\mathcal{E}(n) = \frac{1}{2}\sum_q e_q^2(n)$, which in turn minimizes the error in the entire output of the ANN. Change in each weight during an epoch is calculated as $\Delta w_{qp}(n) = -\eta \frac{\partial \mathcal{E}(n)}{\partial v_q(n)} y_p(n)$, where y_p is the output of the previous neuron and η is the learning rate.

In the learning phase of our MLP execution, for each epoch we input power usage history of past three weeks recorded in 5 minute intervals. Outdoor temperature for the corresponding interval and day of the week is also fed in each epoch (Figure 4.3). The output of the ANN is a structured object (Y) containing multiple possible $\Phi_{day_j}^{h_k}$ for next day. Given that the next day's temperature forecast and day of the week is known, the structure object is parsed for the matching $\Phi_{day_j}^{h_k}$. More details about the MLP specifications used in our

simulation experiments can be found in Section 4.5. Additionally, Connor et al. [49] demonstrated that neural networks trained on filtered data can perform better predictions than neural networks trained on unfiltered time series. Therefore, we use a low-pass filter over SM training data which shaves any load pattern for energy consumption above 4kW. The value 4kW was empirically determined based on the observation that in our training data, more than 94% of data samples are lower than 4kW.

A distributed model of SMs is used in our proposed SGN model, where the first level prediction is performed independently on all SMs belonging to the SGN. The prediction algorithm running inside the SM of a household h_k locally stores a small database (Figure 4.4), containing actual consumption patterns $\mathcal{A}^{h_k}$ and outdoor temperature measurements OT_i from last m days. Each day, the $\mathcal{A}^{h_k}$ and OT_i values are used to train the MLP and predict the $\Phi^{h_k}_{day_j}$ for next (j-th) day. Also, at the end of a day, that day's actual consumption pattern $\mathcal{A}^{h_k}_{day_{j-1}}$ is inserted into the queue of the database and the oldest actual consumption pattern $\mathcal{A}^{h_k}_{day_{j-m+1}}$ is removed. As mentioned before, $\Phi^{h_k}_{day_j}$ is computed and reported only once (before beginning of) each day. All communications between SM and CH for reporting $\Phi^{h_k}_{day_j}$ are assumed to be symmetrically encrypted, for example, by using AES [53].

4.4.2 Prediction at CH Level

The purpose of using prediction at SM is to remove specific load signatures (such as spikes and plateaus) from $\mathcal{A}^{h_k}_{day_j}$. Although the missing spikes and plateaus from the SM of one household represent a minuscule amount of energy for the grid, spikes and plateaus from multiple households in a cluster can add up to a significant amount of unpredicted energy. This accumulated error in prediction can affect processes that would use the predicted data, for example, intelligent electricity distribution, demand-response, etc. Therefore, we introduce another level of statistical prediction at the CH based on historical load profile of the cluster, while also factoring in individual predictions from all SMs in the cluster $\{\Phi^{h_1}_{day_j}, \Phi^{h_2}_{day_j}, \Phi^{h_3}_{day_j}, ...\}$.

As mentioned earlier, the CH is assumed to have load measurement capability to measure total energy consumption in it's neighborhood. The meter measures the energy injected into the entire cluster, without having access to actual individual SM readings at any point. As a result, CH can easily calculate the difference between the aggregated predicted values which are gathered from SMs and the measure of actual total energy consumption in the cluster. The proposed algorithm (Algorithm 1) uses average of difference between past load predictions and actual loads of the entire cluster ($\Lambda_{day_d} = \{\lambda^{(\tau_1)}, \lambda^{(\tau_2)}, \cdots, \lambda^{(\tau_n)}\}$), in order to complement missing loads. The output of the algorithm $\Psi_{day_j} = \{\psi^{(\tau_1)}, \psi^{(\tau_2)}, \cdots, \psi^{(\tau_n)}\}$ is the prediction for the whole cluster reported to CH, where $\psi^{(\tau_i)} = \delta^{(\tau_i)} + \sum_k p^{(\tau_i)}$ and $\delta^{(\tau_i)} = \frac{\sum_{d=j-m}^{d=j-1} \{\lambda_{day_d}^{(\tau_i)} - \sum_k p_{day_d}^{(\tau_i)}\}}{m}$.

Algorithm 5 Prediction algorithm executed by CH.

Prediction Function (for day j)

Define new $\Psi_{day_j} = \{\psi^{(\tau_1)}, \psi^{(\tau_2)}, \cdots, \psi^{(\tau_{288})}\}$

for $k = 1$ to K (K households in the cluster) **do**
  $\sum p_{day_j}^{(\tau_i)}$
end for

for $i = 1$ to 288 (5 minutes time intervals for 24 hours) **do**
  for $d = 1$ to m (m days to historical data) **do**
    $\delta^{(\tau_i)} = \lambda_{day_d}^{(\tau_i)} - \sum_k p_{day_d}^{(\tau_i)}$
  end for
  $\delta^{(\tau_i)} = \frac{\delta^{(\tau_i)}}{m}$
  $\psi^{(\tau_i)} = \delta^{(\tau_i)} + \sum_k p^{(\tau_i)}$
end for

Report Ψ_{day_j} to CH

For the second level prediction, the CH accumulates all $\Phi_{day_j}^{h_k}$ in the cluster, adds the calculated $\delta^{(\tau_i)}$ to $\sum_k p^{(\tau)}$ for each time interval (τ), and reports the resulted pattern Ψ_{day_j} to the EC. CH also stores a database of past Λ and $\sum_k p^{(\tau)}$ values from last m days, which is updated at the end of each day (Figure 4.5).

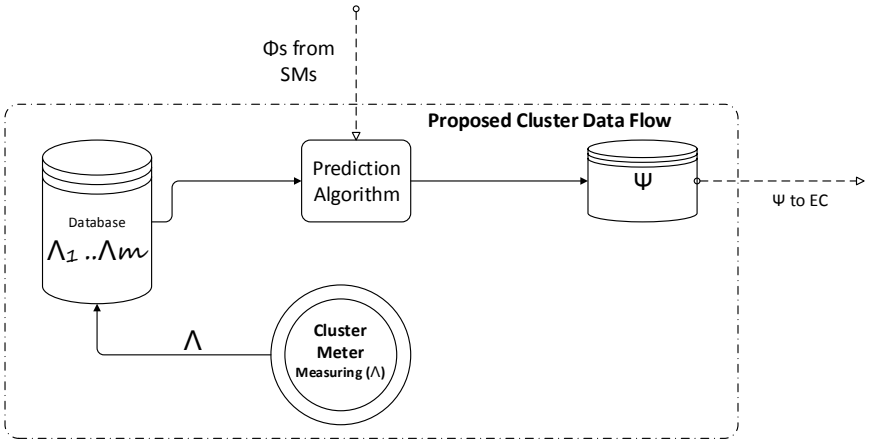


Figure 4.5: Proposed CH data flow.

4.4.3 Privacy Preserving Real-time Monitoring

The predicted pattern Ψ_{day_j} is a refined estimate of next day's energy consumption at the cluster level, as compared to individual SM predictions $\Phi_{day_j}^{h_k}$. However, unexpected events may occur which are not captured by the input variables of our prediction system, for example severe weather conditions, natural disasters, etc. To ensure proper functioning of SGN in case of an unexpected power demand, we incorporate a real-time reporting system in our architecture to measure the difference in actual and predicted energy consumption of all households. But, directly reporting difference in actual and predicted energy consumption pattern to CH defeats our goal of privacy, because CH can add back the difference to predicted pattern to obtain the actual pattern of individual SMs. So, the real-time reporting system uses a "token chain" mechanism to aggregate the difference in actual and predicted energy consumption pattern for all SMs in the cluster. The token chain design can be based on existing energy-efficient token passing mechanisms designed for ad-hoc wireless sensors networks [154] [233] and smart grid networks [190]. In the logical chain of all SMs belonging to a cluster, a token is circulated across all SMs (as shown in Figure 4.1) for aggregation of difference in actual and predicted energy consumption of the cluster. The difference in

aggregated actual and predicted energy consumption $\theta^{(\tau)} = \sum_k (p^{(\tau)} - a^{(\tau)})$ in each time interval τ, can be used to handle unexpected demand events in real-time. Due to aggregation of the difference in actual and predicted energy consumption, individual household privacy is not compromised. Figure 4.4 illustrates how each SM adds their difference in actual and predicted energy consumption to the token. The final token value containing the aggregated difference in actual and predicted energy consumption of the cluster is reported to EC (via CH) for regulating generation and distribution, if necessary. To protect the token chain against eavesdropping inference attacks, all SMs symmetrically encrypt (and decrypt for addition) the token using a shared secret, obtained by using a suitable key exchange protocol, such as the Diffie-Hellman protocol [57].

4.5 Evaluation

In order to validate the benefits of our proposed Seer Grid architecture, we conduct extensive experimental simulations using real smart meter data. In this section we present our experimental setup followed by results.

Table 4.1: Neural network training parameters.

Parameter	**Value**
Number of SMs in cluster (assumed neighborhood)	5
Training period	3 weeks (21 days)
Testing period	3 day
Number of predicted data points a day	288
Number of ANN Inputs	9
ANN Proto	50
Number of ANN hidden layers	3
Number of nodes in each hidden layer	10
Number of ANN output	1
ANN Learn Rule	Ext DBD
ANN transfer mode	Sigmoid
Epoch	288*21=6048
Number of iterations	10^6

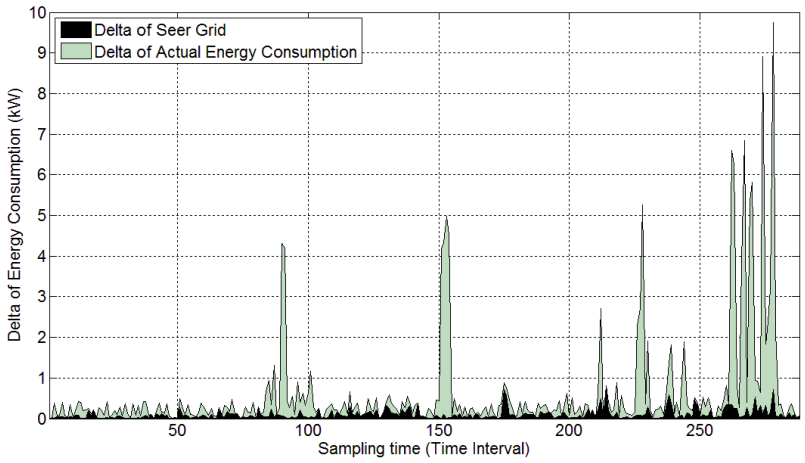
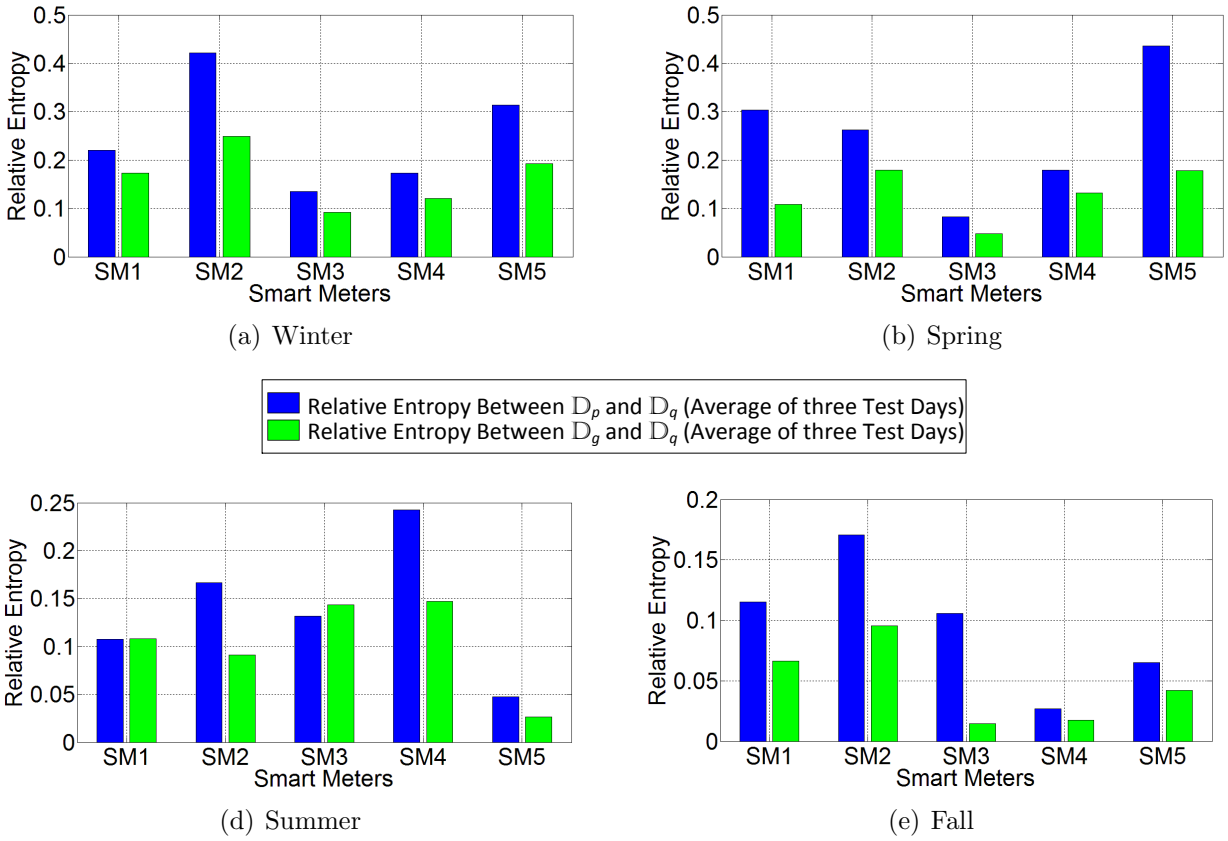


Figure 4.6: Comparison of $\mathbb{D}p$ and $\mathbb{D}q$ over a test day. The lower values of $\mathbb{D}p$ means $\mathbb{P}$ is relatively "smoother".

4.5.1 Experimental Setup

We use real SM data collected from residences equipped with $BS\ EN62053-21002003$ smart meters. The data was recorded in East Midlands, UK in the year 2008 [199]. The fabricated cluster we consider for evaluation consists of 5 households, each having one smart meter. Envisioning the limited memory that SMs may have, we limit the use of historical data in our experiments to three weeks, i.e, $m = 21$. Longer training period not only takes more storage space, but also makes less significant contribution in the prediction because of changing conditions (such as temperature) throughout the year. The ANN prediction algorithm is trained with data from past 21 days to predict the energy consumption for a test day. Every day, the last day's energy consumption information is added to the training set, and the oldest (22^{nd}) day's energy consumption information is removed from the training set. This helps account for changing seasons, and at the same time, limits memory requirements. The training data itself consists of eight variables: interval number and target date as indexing variable, 3 power usage measurements in the interval from last three weeks, and 3 outdoor temperature measurements in the interval from last three weeks. More specific



(a) Winter (b) Spring

(d) Summer (e) Fall

Figure 4.7: Relative entropy between $\mathbb{D}p$ and $\mathbb{D}q$ compared with relative entropy between $\mathbb{D}g$ and $\mathbb{D}q$, where $\mathbb{D}g$ is the series of differences between successive power measurements in GRN induced energy consumption data.

details of the parameters that we use to train our ANN prediction mechanism can be found in Table 4.1.

4.5.2 Privacy Implications

To evaluate the privacy implications of Seer Grid, we conduct two different experiments at the SM level. Both the experiments are designed to observe and compare the amount of information that can be inferred from Seer Grid's predicted energy consumption time series data, versus time series data of actual energy consumption. We define $\mathbb{P} = \{p_1, p_2, \ldots, p_n\}$ as the Seer Grid's predicted energy consumption time series and $\mathbb{Q} = \{q_1, q_2, \ldots, q_n\}$ as the actual energy consumption time series, where $p_t, q_t : 1 \leqslant t \leqslant n$ are the energy consumptions

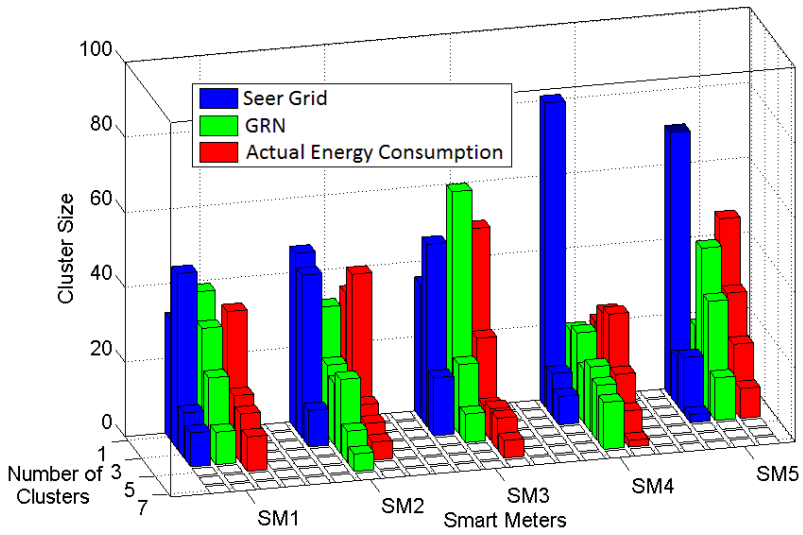


Figure 4.8: Number of clusters in $\mathbb{D}q$, $\mathbb{D}p$, and $\mathbb{D}g$, and percentage of distance in each cluster.

for each time interval. We also calculate difference between successive power measurements in $\mathbb{P}$ and $\mathbb{Q}$, symbolized as $\mathbb{D}p = \{dp(1), \ldots, dp(n)\}$ and $\mathbb{D}q = \{dq(1), \ldots, dq(n)\}$, where $dp(i) = p_i - p_{i-1}$ and $dq(j) = q_j - q_{j-1}$. $\mathbb{D}p$ indicates the changes in energy consumption load, which is important to understand privacy leaked through load signatures. We plot $\mathbb{D}p$ and $\mathbb{D}q$ in Figure 4.6 to visualize how much Seer Grid indeed suppresses load differences. It can be observed that Seer Grid has consistently less changes in energy consumption load throughout the test day, indicating consistent privacy protection. This motivates us to further analyze $\mathbb{D}p$ with respect to $\mathbb{D}q$, and compare privacy leakage of Seer Grid with another well-known protection mechanism in the following two experiments.

$$D(A\|B) = \sum_i A(i) \log \frac{A(i)}{B(i)} \tag{4.1}$$

Relative Entropy: In our first experiment, we try to quantitatively compare $\mathbb{D}p$ and $\mathbb{D}q$ over four seasons (each with 21 training and 3 test days): winter (January 1 to 24), spring (April 1 to 24), summer (July 1 to 24), and fall (October 1 to 24). Let A and B be the probability distributions of $\mathbb{D}p$ and $\mathbb{D}q$, respectively. We use the well-established relative entropy metric (Equation 4.1) as a non-symmetric measure of difference between the two

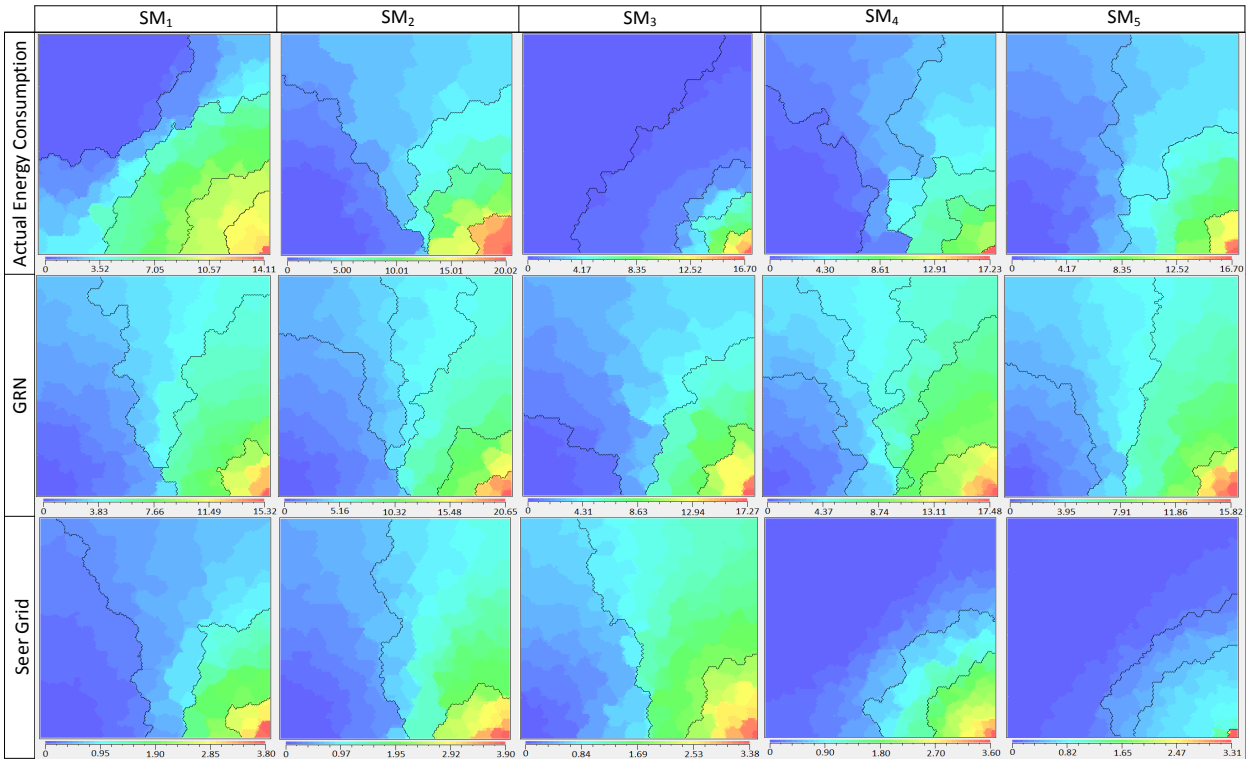


Figure 4.9: Cluster forms on $\mathbb{D}q$, $\mathbb{D}p$, and $\mathbb{D}g$. The experiment is performed with 21 training (January 1 to 21) and 3 test days (January 22 to 24). The results are averaged over 3 consecutive test days.

probability distributions A and B. Due to the premetric property of relative entropy, the larger the relative entropy $D(A\|B)$, the higher the level of protection offered.

In order to understand the level of privacy protection offered by Seer Grid, we compare it's relative entropy with a widely accepted perturbation technique - introduction of Gaussian random noise (GRN)[207]. Introduction of GRN adds or subtracts random values in each reported energy consumption interval, in order to mislead an adversary. However, completely random noise will reduce data utility adversely. Therefore, we use past energy consumption data features to set a level of noise which balances between privacy and utility, as proposed by [207]. Figure 4.7 shows the relative entropy values for Seer Grid and GRN, for the five smart meters under evaluation. From the results, we observe that the relative entropy of

Seer Grid is generally higher compared to relative entropy of GRN, which indicates that Seer Grid may offer better privacy protection.

Clustering: As a second metric to evaluate privacy of Seer Grid, we apply a clustering technique to recover activity information of consumers. Clustering creates groups of similar levels of energy consuming intervals. More number of the clusters can leak more granular information (spikes, switching on/off, and consumption pattern) in each household. In other words, higher number of clusters inferred by an adversary reveals more private information about appliances and activity within the household. We use the Self Organizing Map (SOM) algorithm [130] to create clusters on successive energy difference time series. The interesting aspect of SOM is that it learns to cluster data without supervision. In our application, SOM groups input values into n clusters such that the difference between power consumption values across clusters is minimized. We use the Viscovery tool [244] to apply SOM and calculate the optimum number of clusters for successive energy differences in actual, Seer Grid predicted and GRN perturbed time series (Figure 4.8). As defined before, $\mathbb{D}q$, $\mathbb{D}p$, and $\mathbb{D}g$ denote the successive differences in actual, Seer Grid and GRN energy time series, respectively. Figure 4.9 visualizes the clustering done by the SOM algorithm on the three series, for each of the five SMs. Each sub-figure in Figure 4.9 is clustered into specific distances between the cluster members, where the distance is varied from zero to the maximum in the time series. As evident from Figure 4.8 and 4.9, Seer Grid generally has the lowest number of clusters, and thus, reveals least information compared to actual and GRN induced energy time series. Figure 4.8 also illustrates clusters and distribution of cluster population within each cluster. Because Seer Grid prediction results in a "smoother" pattern, we observe a high population in the low distance clusters.

Comparison with SARMA [224]: Singh et al. proposed the use of Seasonal Auto Regressive Moving Average (SARMA) for household load prediction. We compare our household level prediction with SARMA by recreating their experiments for our earlier defined test days across four seasons. Figure 4.10 shows the root mean square error and normalized

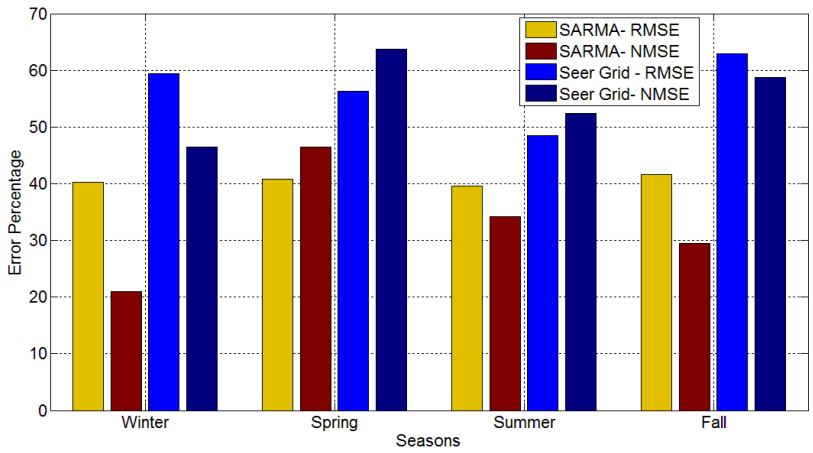


Figure 4.10: Root mean square error (RMSE) and normalized mean square error (NMSE) percentage in predicted household loads by SARMA and Seer Grid. Results are averages of the five test SMs.

mean square error percentage in the predicted loads. Higher error percentage means that the prediction is further off from the actual load values, implying more privacy against inference attacks. It is observed that Seer Grid's household level prediction has an overall higher error percentage compared to SARMA, which means Seer Grid offers more privacy than SARMA. However, it should be noted that SARMA was designed to have low error percentage in household level prediction, so as to improve utility. On the other hand, our primary goal for the household level prediction is to improve privacy. However, the error percentage should not be very high, otherwise the cluster level prediction may suffer loss of data utility. We evaluate the utility of the cluster level prediction below, to validate that data utility is in fact not significantly compromised in Seer Grid, even when there exists relatively higher error percentage in household level prediction.

4.5.3 Data Utility

In this section we empirically evaluate the utility implications of Seer Grid using the well-accepted squared correlation metric [207, 120]. We conduct experiments over four seasons: winter (January 1 to 24), spring (April 1 to 24), summer (July 1 to 24), and fall (October 1

to 24). The results, averaged over the 3 test days, are presented in Table 4.2. The squared correlation between actual and predicted energy consumption patterns of SM vary between 51.07% and 80.09%, and averages at 62.10% across all 5 SMs. As an example, Figure 4.11(a) shows the actual and predicted energy consumption pattern for a SM on 22nd January, and Figure 4.11(b) shows the squared correlation between them. The squared correlation between actual and predicted energy consumption pattern for CH vary between 89.95% and 91.15%, and averages at 90.60%. Figure 4.11(c) shows the actual and predicted energy consumption pattern for CH on 22nd January, and Figure 4.11(d) shows the squared correlation between them. Evidently, SM prediction is less correlated than CH prediction by a clear margin, as seen in Table 4.2. We also check the standard deviation of the test days to verify there does not exists any bias. Standard deviation values appear random, without any visible connection with the squared correlation results, leading us to believe that our results are unbiased.

Table 4.2: Squared correlation coefficient (R^2) between predicted and actual energy consumption patterns for each SM and CH, and the standard deviations of the 3 test days.

Season	SM1	SM2	SM3	SM4	SM5	CH
Winter, R^2: Actual vs Predicted	0.5715	0.5529	0.7793	0.6421	0.5772	0.9098
Winter, 3 Test Days Std Deviation	0.1240	0.0618	0.1470	0.0901	0.0187	0.0118
Spring, R^2: Actual vs Predicted	0.5107	0.5627	0.8009	0.6687	0.5799	0.9115
Spring, 3 Test Days Std Deviation	0.0728	0.1236	0.1588	0.0459	0.1868	0.0095
Summer, R^2: Actual vs Predicted	0.5888	0.5341	0.6322	0.6439	0.6528	0.8985
Summer, 3 Test Days Std Deviation	0.1775	0.1366	0.0922	0.0479	0.1855	0.01725
Fall, R^2: Actual vs Predicted	0.6195	0.6025	0.6477	0.6450	0.6072	0.9041
Fall, 3 Test Days Std Deviation	0.0572	0.1412	0.0284	0.0808	0.1074	0.0102

Comparison with Jain and Satish [111]: Jain and Satish proposed the novel use of support vector machines (SVM) to perform short-term load forecasting at cluster level [111]. To better understand how our cluster level prediction would perform, we do a comparative analysis with [111] by recreating their experiments, in the same period of our earlier defined test days across four seasons. Figure 4.10 shows the maximum percent error and average

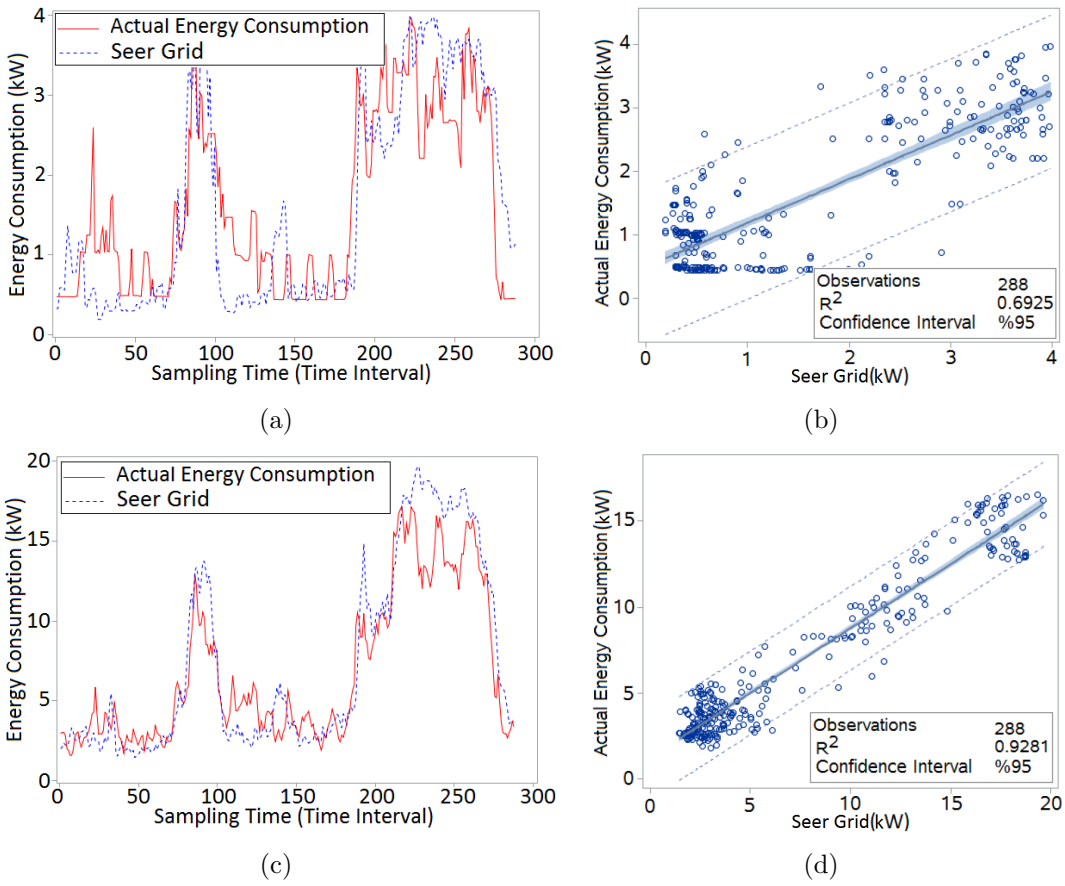


Figure 4.11: Exemplary results from 22nd January 2008, showing the correlation between actual and predicted energy consumption patterns at different levels of Seer Grid. (a) Actual and predicted energy consumption patterns for one of the SMs, (b) Correlation between actual and predicted energy consumption patterns for the same SM, (c) Actual and predicted energy consumption patterns for CH, (d) Correlation between actual and predicted energy consumption patterns for CH.

percent error in cluster level load prediction for Jain and Satish, and Seer Grid. In this case, lower percent error in prediction implies better utility for EC. Seer Grid's cluster level prediction has marginally lower percent error in winter and fall, marginally higher percent error in spring, and a significantly lower percent error in summer. Overall, we can conclude that Seer Grid's utility is similar to [111], if not better. Readers should note that [111]'s SVM based load forecasting at cluster level is only a single level prediction, where the prediction model is trained with actual data from past. Whereas, in case of Seer Grid, the cluster level prediction is primarily based on the household level prediction, which provides better privacy

as seen earlier. Therefore, Seer Grid having similar utility as other cluster level prediction schemes is very promising.

4.6 Discussions

4.6.1 Smart Meter Performance Analysis

Although Seer Grid uses complex prediction schemes, it does not suffer from significant computational and communicational bottlenecks. As the prediction is once a day event, SMs have an entire day to compute for next day, which should be sufficient even for less powerful computing systems. Reporting the predicted data is also an once-a-day event, and SMs can avoid network congestion if they report using a multiplexed (time, frequency, or code sharing) channel with other SMs. Among all SM privacy preserving frameworks, the most closely resembling (in terms of resource requirements, architecture, and assumptions made) frameworks are based on homomorphic cryptography [17]. So, we compare the computational complexity of Seer Grid versus frameworks based on homomorphic cryptography. Seer Grid's household level prediction through multilayer perceptron has a time complexity of $O(x^2)$ [125], while the one based on Paillier cryptographic protocol has time complexity of $O(y^3)$ [17]. Therefore, time complexity of homomorphic cryptography based SM privacy frameworks is $O(y^3t)$ while Seer Grid's time complexity is $O(x^2t)$, where x and y are the size of input in Seer Grid and homomorphic (Paillier) cryptosystem [17], respectively, with $x \ll y$, and t is the number of daily samples in both schemes. In other words, Seer Grid's asymptotic time complexity is lower than similar aggregation frameworks based on homomorphic cryptography.

4.6.2 Implications

The Importance of Two Level Prediction: One may think that only a single level of prediction may achieve the same results as two-levels, but a single level of prediction has some inherent drawbacks. If the prediction is done only at the CH level (where households

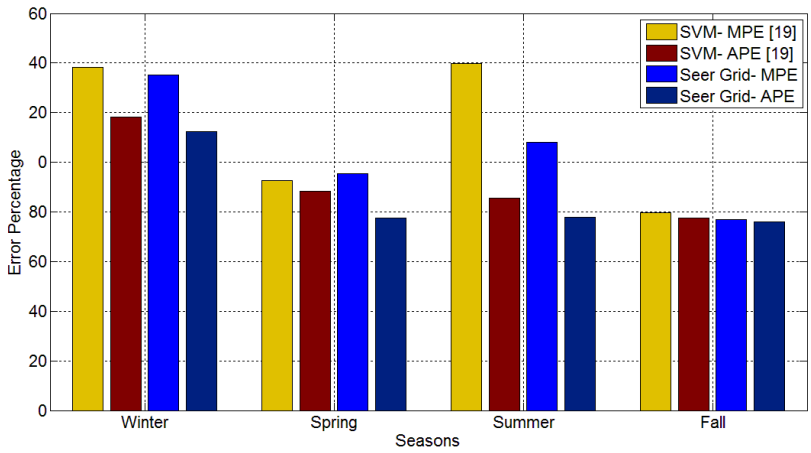


Figure 4.12: Maximum percent error (MPE) and average percent error (APE) in cluster level load prediction for Jain and Satish, and Seer Grid. Results are averages of the five test SMs.

report their actual consumption to CH), we lose privacy at the SM level. Whereas, if prediction is done only at the SM level, the cluster-wide difference between actual and predicted consumption data will be larger, resulting in data utility loss.

Training Parameters: In our experiments, we took a heuristic approach for determining the training parameters (epochs, iterations, learning rule, etc.) for the ANN used by SMs. The parameters were chosen in such a way that it satisfies our goal of optimizing both privacy and utility of SM data. From the experimental results we observe that the correlation between actual and predicted energy consumption pattern varies moderately across households and seasons. This is primarily because of different characteristics of the training data (actual energy consumption for last 21 days) leading to differently converged prediction model in each SM. In future, we plan to develop a unified prediction framework for the SMs which will analyze characteristics of the training data, and accordingly govern learning rate such that prediction accuracy remains below a privacy preserving threshold with high likelihood. Unlike this work, where all SMs use the same prediction parameters, the unified framework will adapt to the characteristics of local training data of individual SMs. As a

result, the convergence in learning will be more uniform across SMs and seasons, thus offer a more stable privacy guarantee.

Privacy due to Uncertainty: Uncertainty in next day's energy consumption provides user privacy in Seer Grid, which is similar to how uncertainty in location data provides spatio-temporal privacy [164]. The naturally occurring irregularities in consumers' day-to-day schedule results in smoother household prediction patterns (that hides load signatures), which also means that the predicted energy pattern cannot be used to determine temporary house unoccupancies with complete confidence.

Larger Cluster: We consider a very small scale cluster in our experiments, and yet achieve considerably high prediction accuracy at the cluster level. As evident from previous cluster level prediction schemes [255], accuracy tends to dramatically improve with increasing cluster size. Thus, we think our results are highly encouraging for large scale implementation.

4.6.3 Dishonest and Malfunctioning Smart Meters

SMs are often the target of bad data injection attacks, primarily due to monetary incentives [163][88]. However, it is critical for ECs to prevent such attacks, not only to avoid financial losses, but also to ensure proper distribution of electricity. Previous efforts in this direction suggested the use of embedded sensors for 'Trusted Metering' [188], having a centralized or dedicated detection system, or a hybrid system of embedded sensors and centralized detection[88]. In Seer Grid, as the CH collects predicted energy consumption data of individual smart meters in advance, existing anomaly detection mechanisms can be effectively applied on the predicted energy consumption data reported by individual smart meters.

In real deployment, SMs and/or communication links can also experience failures, due to which they may be unable to report the predicted energy consumption information. Similar challenge is also faced by existing SMs (and many proposed aggregation schemes), and can be a non-trivial issue to address. If a negligible number of SMs (belonging to the same

energy company) are unresponsive, the effects most likely will be unnoticeable. However, if a large number of SMs are unresponsive, the effects can be significant. In Seer Grid, such cases of malfunctioning SMs can perhaps be handled more efficiently than other aggregation schemes, due to the readily available past prediction data. For example, if the next day's predicted energy consumption data is not reported by a SM, the cluster head can simply substitute it with the same week-day's prediction of that SM from last week. The intuition behind this exemplary approach is that households generally have similar usage pattern for each day of the week [215].

4.6.4 Deployment Barriers

Smart meter deployment presents EC with many logistical, technical and commercial challenges. The primary incentive for ECs to deploy SMs is efficiency and thus savings over time. Conventional SMs, already deployed in many places, perfectly serve this commercial benefit. However, these SMs were not designed to provide privacy for consumers. As a result, any new framework designed to enable consumer privacy will require modification or re-deployment by the EC, which will require additional investment from ECs. Because this new investment does not add any additional efficiency improvements, ECs might be reluctant in deploying any privacy preserving add-ons to existing SMs. This is a major limitation faced by many novel privacy preserving frameworks proposed for smart grids [33]. Cavoukian and Dix [33] pointed out that privacy by design is the best approach. Therefore, deployment of Seer Grid can be easier in new localities (without existing smart metering infrastructure), than to implement in localities where smart metering is already in place. Given that Seer Grid will require additional hardware and software to function, below are the few directions we think can aid deployment:

- *Add-On Service:* ECs can offer SMs with Seer Grid's prediction framework as an add-on service. That is, privacy-aware consumers can opt in for the privacy preserving frame-

work, by paying an one time fee, which would cover the cost of additional hardware and software installation.

- *Off-Loading Computation:* Instead of adding a computing unit (for performing the prediction operations) built inside the SMs, it may be beneficial to off-load the operations to a household computer. For example, the prediction operations can be undertaken by a paired (using low energy communication protocols, such as Bluetooth) smartphone or PC, once per day. The prediction results can be communicated back to the SM for reporting to EC. Also, future upgrades may be easier for consumers, as smartphones and PCs are more frequently upgraded [243].

Alternatively, privacy issues can result in poor SM adoption in privacy-aware communities [156]. By addressing privacy issues in a way that does not hamper utility too much, ECs can increase SM adoption. This can be an incentive for ECs to participate in implementing frameworks like Seer Grid[5].

[5]The content of this chapter has appeared in [23].

CHAPTER 5
CONCLUDING REMARKS AND FUTURE RECOMMENDATIONS

In this dissertation, security, privacy, and seamless network availability are investigated in CI-CPS. CI-CPS are smart systems which collaborate to do computations for critical infrastructures. Critical infrastructures can be a part of wireless mobile networks, monitoring, controlling structures, decision makers and actuators. Different proposed methods in each chapter of this dissertation presents different solutions to overcome existing drawbacks and vulnerabilities in CI-CPS.

In this direction, Chapter 2 addressed the problem of securing location discovery of wireless and mobile components (of a CI-CPS) by proposing a novel spread-spectrum-based approach to eliminate incorrect localization data injected by malicious location anchors. Chapter 3 presented a framework to increase the capacity (and consequently availability) of existing wireless networks, by utilizing a secondary cognitive radio network based approach. Chapter 4 presented a novel framework to enable privacy-preserving smart meter data reporting in a smart grid CI-CPS, with a minimal impact on data utility.

Chapter 2 presented a new approach for securing localization in anchor-based localization protocols. The proposed approach implements a request confusion strategy in order to anonymize localization requests, and a reactive jamming strategy on the CDMA response channel to actively disable malicious or cheating anchors. Simulation results showed that if appropriate parameters are chosen, the proposed technique is effective in eliminating cheating anchors in localization protocols, with significant accuracy. An extensive literature review shows that our technique is one of the first such techniques that deploys jamming on a DSSS or CDMA communication channel for actively securing location discovery in wireless networks.

Chapter 3 introduced an optimal resource allocation for improving NSU supported in an underlay CSGN, based on the requirements to transmit huge traffic on the future smart grid network. Our approach is, to the best of our knowledge, the first scheme utilizing sub-OCSs to improve number of SUs in CSGNs. We have investigated, through extensive simulations and analysis, the maximum possible error (due to the co-existence of SUs and PUs in our CDMA-based CSGN) that can be tolerated by primary users of the network. Our scheme dedicates a substring of OCSs used by Smart Appliances, referred as sub-OCS, to SUs in higher hierarchies in the proposed CSGN which enables all SUs and PUs to transmit simultaneously. This will result in low, and tolerable, error in the received data for PN (TV network) and negligible error for SN. As a result, PUs and SUs (using OCSs and sub-OCSs) will be able to transmit at the same time. Simulation results prove that our proposed method has insignificantly impact on both PN and SN, with a considerable increase in additional SUs supported.

In Chapter 4, we proposed Seer Grid, an alternate SGN architecture aimed to reduce the privacy-utility trade-off faced by SMs. As a result of two-level energy load prediction in Seer Grid, there exists high correlation between predicted and actual energy consumption patterns at cluster level, which indicates excellent utility preservation. However, the correlation between predicted and actual energy consumption patterns of individual SM is weak, which indicates good privacy preservation for households. Evaluation results strongly support our proposition of Seer Grid.

5.1 Future Research Challenges

Thanks to the numerous benefits of deploying Smart Meters (SM) in Smart Grid Networks (SGN), such as real-time demand-response, efficient power distribution, and dynamic pricing, SMs are gaining rapid popularity [114]. However, one of the most criticized aspects of SMs is the lack of privacy for households equipped with SM, due to which many communities have even declined to adopt SMs [114]. Privacy of household energy consumption data is

essential for preventing inference attacks based on load signatures [207]. As a result, various privacy-preserving protocols for SM reporting were proposed by researchers [45, 237, 46, 85]. These protocols enable reporting of SM data to the Energy Company (EC) in an anonymous fashion, thereby preventing inference attacks on the energy consumption data of individual households. If such anonymous SM reporting protocols are implemented by ECs, it can encourage the adoption of SMs and enable the benefits of SGNs.

However, anonymous SM reporting protocols introduce a new problem of false data reporting. As privacy-preserving protocols remove the linkability of the reported energy consumption data (through aggregation, encryption, blind signatures, etc.) with the actual energy consumption, residents or remote adversaries can now tamper with the SM reporting system without facing any consequences. One of the primary motivation for reporting false energy consumption data can be to trick a dynamic pricing system for lower energy costs [66, 231]. Another motivation can be to sabotage the normal operations of the electricity grid, by an remote adversary using distribute false data reporting attacks [105]. Therefore, false data reporting attacks can jeopardize the quality of service expected from a SGN. Furthermore, any losses incurred by the EC due to undetected 'cheating' will indirectly lead to higher energy prices for all customers of EC. This may in turn encourage honest customers to cheat as well, causing severe loss of SM data utility.

As newer SMs are becoming capable of directly communicating with EC over the Internet [108, 223], several recent works proposed the use of blind signatures to facilitate anonymous SM reporting over the omnipresent Internet [38, 45, 237, 46, 85]. With our view that anonymous SM reporting over the Internet will eventually become more widely adopted, a naive approach to detect cheating by maintaining a history of reported energy consumption data for each SM (and checking if certain SMs fall outside an usual threshold) will not work, due to the security properties of blind signatures. Blind signatures based SM reporting ensures that reported data is from an authenticated SM, but does not allow EC to link it back to a SM. Moreover, multiple reports from the same SM over a period of time are not linkable,

making it harder for the EC to even maintain pseudonymous history. To be best of our knowledge, the problem of detecting and mitigating cheating in anonymous SM reporting has not been addressed until now.

As privacy-preserving SM reporting can ensue in false data injection attacks, the EC would prefer to have a protocol which preserves SM privacy and simultaneously prevents cheating. To mitigate the problem of cheating in blind signatures based SM reporting, we require a framework to gracefully de-anonymize cheating SMs, while maintaining anonymity of all other honest customers. Repetitive cheaters should be identified after a certain number of cheatings, and the EC may penalize the cheating SMs so as to discourage cheating. While such a framework should be able to mitigate cheaters, it also leads us to several open questions, such as how much of cheating goes undetected, how anonymity is affected at each level of the iterative scheme, or what should be the appropriate penalty for cheating such that the reparations of cheating outweighs the incentives? As part of our future works, we will model a two-player non-cooperative game between the EC (trying to maximize revenue) and cheating customers (trying to trick the EC for lower electricity bills). Game-theoretic results obtained by solving the game can lead us to the optimal answers to these questions, which should be used by EC while implementing such a framework.

BIBLIOGRAPHY

BIBLIOGRAPHY

[1] US department of energy, residential energy consumption survey (recs) - data - u.s. energy information administration (eia), https://www.eia.gov/consumption/residential/data/, Augest 2009.

[2] *The Department of Homeland Security*, chapter Chemical, Biological, Radiological, and Nuclear Countermeasures. The White House, April 2003.

[3] I. F. Akyildiz, W. Lee, and K. R. Chowdhury. Spectrum Management in Cognitive Radio Ad-hoc Networks. *IEEE Commun. Magazine*, 2009.

[4] B. Alomair, A. Clark, J. Cuellar, and R. Poovendran. Toward a statistical framework for source anonymity in sensor networks. *IEEE Transactions on Mobile Computing*, 12(2):248–260, 2013.

[5] T. Alpcan and T. Basar. A game theoretic approach to decision and analysis in network intrusion detection. In *IEEE CDC '03*, volume 3, pages 2595–2600 Vol.3, Dec 2003.

[6] R. Amin, J. Martin, and X. Zhou. Smart Grid Communication using Next Generation Heterogeneous Wireless Networks. In *IEEE SmartGridComm'12*, Taiwan, Nov 2012.

[7] R. Amin, J. Martin, and X. Zhou. Smart grid communication using next generation heterogeneous wireless networks. In *IEEE SmartGridComm '12*, pages 229–234, Nov 2012.

[8] H. M. Ammari. *The Art of Wireless Sensor Networks*. Springer, 2014.

[9] R. Anderson and S. Fuloria. On the security economics of electricity metering. In *IEEE WEIS '10*. Citeseer, 2010.

[10] B. Arazi, I. Elhanany, O. Arazi, and H. Qi. Revisiting Public-Key Cryptography for Wireless Sensor Networks. *IEEE Computer*, 38(11):103–105, 2005.

[11] B. Arazi, I. Elhanany, O. Arazi, and H. Qi. Revisiting Public-key Cryptography for Wireless Sensor Networks. *Computer*, 38(11):103–105, 2005.

[12] A. Attar, M. R. Nakhai, and A. H. Aghvami. Cognitive radio transmission based on direct sequence mc-cdma. *Wireless Communications, IEEE Transactions on*, 7(4):1157–1162, Apr 2008.

[13] J. Bachrach and C. Taylor. *Handbook of Sensor Networks*, chapter Localization in Sensor Networks, pages 277–310. John Wiley & Sons, Inc., 2005.

[14] P. Bahl and V. N. Padmanabhan. RADAR: an in-building RF-based User Location and Tracking System. In *IEEE INFOCOM '00*, pages 775–784, Tel-Aviv, Israel, Mar 2000.

[15] J. Beaver, M. A. Sharaf, A. Labrinidis, and P. K. Chrysanthis. Location-aware routing for data aggregation for sensor networks. 2003.

[16] R. Bellovin. Cryptography: Authentication, blind signature, and digital cash, https://math.berkeley.edu/ rmb/writings/chaum.pdf, university of california, berkeley math department, 2009.

[17] A. Bessani and S. Bouchenak, editors. *Distributed Applications and Interoperable Systems*, volume 9038 of *Lecture Notes in Computer Science*. Springer, 2015.

[18] A. O. Bicen, O. B. Akan, and V. C. Gungor. Spectrum-aware and cognitive sensor networks for smart grid applications. *Communications Magazine, IEEE*, 50(5):158–165, May 2012.

[19] A. Boukerche, H. A. B Oliveira, E. F. Nakamura, and A. A. F. Loureiro. A voronoi approach for scalable and robust dv-hop localization system for sensor networks. In *ICCCN '07*, pages 497 –502, 2007.

[20] A. Boustani, N. Alamatsaz, M. Jadliwala, and V. Namboodiri. Locjam: A novel jamming-based approach to secure localization in wireless networks. In *CCNC '14*, Jan 2014.

[21] A. Boustani, M. Jadliwala, H. M. Kwon, and N. Alamatsaz. Optimal resource allocation in cognitive smart grid networks. In *IEEE CCNC '15*, pages 499–506, Jan 2015.

[22] A. Boustani, S. Khorsandi, R. Danesfahani, and N. MirMotahhary. An Efficient Frequency Reuse Scheme by Cell Sectorization in OFDMA Based Wireless Networks. In *IEEE ICCIT '09*, Korea, Nov 2009.

[23] A. Boustani, A. Maiti, S. Yousefian Jazi, M. Jadliwala, and V. Namboodiri. Seer grid: Privacy and utility implications of two-level load prediction in smart grids. *IEEE Transactions on Parallel and Distributed Systems*, PP(99):1–1, 2016.

[24] A. Boustani, J. Sabet, M. Azizi, N. Mirmotahhary, and S. Khorsandi. Persian code: A new orthogonal spreading code generation algorithm for spread spectrum cdma systems. In *IEEE WiAD '10*, pages 1–5, Jun 2010.

[25] G. E. P. Box and G. M. Jenkins. Time series analysis: Forecasting and control. In *Holden-Day series in time series analysis*. Holden-Day, 1976.

[26] J. Bruck, J. Gao, and A. Jiang. Localization and Routing in Sensor Networks by Local Angle Information. In *MobiHoc '05*, pages 181–192, Urbana-Champaign, IL, USA, 2005.

[27] S. Bu and F. R. Yu. Green cognitive mobile networks with small cells for multimedia communications in the smart grid environment. *Vehicular Technology, IEEE Transactions on*, 63(5):2115–2126, Jun 2014.

[28] N. Bulusu, J. Heidemann, and D. Estrin. GPS-less Low Cost Outdoor Localization for Very Small Devices. *IEEE Personal Communications Magazine*, pages 28–34, Oct 2000.

[29] N. Bulusu, J. Heidemann, and D. Estrin. Density adaptive algorithms for beacon placement. In *IEEE ICDCS '01*, page 489, Phoenix, AZ, USA, Apr 2001.

[30] W. Cao, Q. Zhang, and A. Nallanathan. A UWB Localization Scheme for LOS and NLOS Environments using Orthogonal Codes. In *ICUWB '11*, Italy, 2011.

[31] R. Cardell-Oliver, K. Smettem, M. Kranz, and K. Mayer. A Reactive Soil Moisture Sensor Network: Design and Field Evaluation. *International Journal of Distributed Sensor Networks*, 1(2):149–162, 2005.

[32] M. W. Carter, H. H. Jin, M. A. Saunders, and Y. Ye. Spaseloc: An Adaptive Subproblem Algorithm for Scalable Wireless Sensor Network Localization. *SIAM J. on Optimization*, 17(4):1102–1128, 2006.

[33] A. Cavoukian and A. Dix. *Smart Meters in Europe: Privacy by Design at its Best.* Information and Privacy Commissioner of Ontario, Canada, 2012.

[34] V. Chakravarthy, L. Xue, Z. Ruolin, W. Zhiqiang, and M. Temple. Novel overlay/underlay cognitive radio waveforms using sd-smse framework to enhance spectrum efficiency-part ii: analysis in fading channels. *Communications, IEEE Transactions on*, 58(6):1868–1876, Jun 2010.

[35] C. Chang. An interference-avoidance code assignment strategy for the hierarchical two-dimensional-spread mc-ds-cdma system: A prototype of cognitive radio femtocell system. *Vehicular Technology, IEEE Transactions on*, 61(1):166–184, Jan 2012.

[36] R. Chang, C. Chu, and Y. Chiu. An efficient anonymous scheme for computer and communication privacy. In *IEEE CCST '04*, pages 199–203, Oct 2004.

[37] M. Chaouch. Clustering-based improvement of nonparametric functional time series forecasting: Application to intra-day household-level load curves. *IEEE Transactions on Smart Grid*, 5(1):411–419, 2014.

[38] D. Chaum. Blind signatures for untraceable payments. In *CRYPTO '82*, pages 199–203. Plenum Press, 1982.

[39] D. Chaum and E.V.Heyst. Group signatures. In *EUROCRYPT '91*, 1991.

[40] H. H. Chen. *The Next Generation CDMA Technologies.* John Wiley and Sons, 2007.

[41] J. Chen, Y. Niu, and F. Yao. Joint Design of Frequency and Power Adaptation in FHSS Systems Based on Cognitive Radio. In *ACM WiCOM '09*, China, Sep 2009.

[42] Y. Chen, W. S. Lin, F. Han, Y. H. Yang, Z. Safar, and K. J. R. Liu. A cheat-proof game theoretic demand response scheme for smart grids. In *IEEE ICC '12*, pages 3362–3366. IEEE, 2012.

[43] R. Cheng and S. Prabhakar. Using uncertainty to provide privacy-preserving and high-quality location-based services. In *ACM MobileHCI '04*, volume 4, 2004.

[44] S. Cheng and Z. Yang. Energy-efficient power control game for cognitive radio systems. In *ACM SNPD '07*, volume 1, pages 526–530, Jul 2007.

[45] J. C. L. Cheung, T. W. Chim, S. M. Yiu, V. O. K. Li, and L. C. K. Hui. Credential-based privacy-preserving power request scheme for smart grid network. In *IEEE GLOBECOM '11*, pages 1–5, Dec 2011.

[46] T. W. Chim, S. M. Yiu, L. C. K. Hui, and V. O. K. Li. Privacy-preserving advance power reservation. *Communications Magazine, IEEE*, 50(8):18–23, Aug 2012.

[47] C. C. Chou, D. S. L. Wei, C. C. J.Kuo, and K. Naik. An efficient anonymous communication protocol for peer-to-peer applications over mobile ad-hoc networks. *Selected Areas in Communications, IEEE Journal on*, 25(1):192–203, Jan 2007.

[48] J. Claessens, B. Preneel, and J. Vandewalle. Solutions for anonymous communication on the internet. In *IEEE CCST '99*, pages 298–303, 1999.

[49] J. T. Connor, R. D. Martin, and L. E. Atlas. Recurrent neural networks and robust time series prediction. *IEEE Transactions on Neural Networks*, 5(2):240–254, 1994.

[50] S. E. Coull, M. Green, and S. Hohenberger. Access controls for oblivious and anonymous systems. *ACM TISSEC '11*, 14(1):10, 2011.

[51] Crossbow. *Imote2 IPR2400 High Performance Wireless Sensor Network Node Datasheet, , Device Manual*. Key Number 6020-0117-01 Rev A, Date Accessed Nov 2016.

[52] Crossbow. *MICA2 Wireless Measurement System Datasheet, Device Manual, 2007*. Key Number 6020-0042-08 Rev A, Date Accessed Nov 2016.

[53] J. Daemen and V. Rijmen. *The Design of Rijndael: AES-The Advanced Encryption Standard*. Springer Science & Business Media, 2002.

[54] N. B. Dale, C. Weems, and M. R. Headington. *Programming and Problem Solving With C++*. Jones & Bartlett Publishers, 1998.

[55] M. Dashti, P. Azmi, and K. Navaie. Resource allocation for underlay cdma cognitive radio networks. In *IEEE WCNC '12*, pages 2792–2796, Apr 2012.

[56] B. Defend and K. Kursawe. Implementation of privacy-friendly aggregation for the smart grid. In *ACM SEGS '13*, pages 65–74. ACM, 2013.

[57] W. Diffie and M. Hellman. New directions in cryptography. *IEEE Transactions on Information Theory*, 22(6):644–654, Nov 1976.

[58] L. Doherty, L. E. Ghaoui, and K. S. J. Pister. Convex Position Estimation in Wireless Sensor Networks. In *IEEE INFOCOM '01*, Anchorage, AK, USA, Apr 2001.

[59] R. Dong, A. A. Cárdenas, L. J. Ratliff, and S. S. Sastry. Quantifying the utility-privacy tradeoff in the smart grid. *IEEE CoRR '14*, abs/1406.2568, 2014.

[60] J. Dricot, F. Horlin, and P. D. Doncker. On the Co-Existence of Dual-Polarized CDMA Networks. In *IEEE CrownCOM '09*, Germany, Jun 2009.

[61] C. Efthymiou and G. Kalogridis. Smart grid privacy via anonymization of smart metering data. In *IEEE SmartGridComm '10*, pages 238–243. IEEE, 2010.

[62] A. Elezabi, M. Kashef, M. Abdallah, and M. M. Khairy. Cognitive interference-minimizing code assignment for underlay cdma networks in asynchronous multipath fading channels. In *ACM IWCMC '09*, New York, NY, USA, 2009.

[63] T. Eren, D. Goldenberg, W. Whiteley, Y. R. Yang, A. S. Morse, B. D. O. Anderson, and P. N. Belhumeur. Rigidity, Computation and Randomization of Network Localization. In *IEEE INFOCOM '04*, Hong Kong, China, Apr 2004.

[64] Z. Erkin, J. R. Troncoso-Pastoriza, R. L. Lagendijk, and F. Perez-Gonzalez. Privacy-preserving data aggregation in smart metering systems: an overview. *Signal Processing Magazine, IEEE*, 30(2):75–86, 2013.

[65] M. Erol-Kantarci and H. Mouftah. Wireless multimedia sensor and actor networks for the next generation power grid. *Ad Hoc Networks*, 9(4):542–551, June 2011.

[66] M. Esmalifalak, G. Shi, Z. Han, and L. Song. Bad data injection attack and defense in electricity market using game theory study. *Smart Grid, IEEE Transactions on*, 4(1):160–169, Mar 2013.

[67] I. Esnaola, S. M. Perlaza, and H. V. Poor. Equilibria in data injection attacks. In *IEEE GlobalSIP '14*, pages 779–783, Dec 2014.

[68] M. Faisal, A. A. Cardenas, and D. Mashima. How the quantity and quality of training data impacts re-identification of smart meter users? In *IEEE SmartGridComm '15*, pages 31–36, Nov 2015.

[69] L. Fang, W. Du, and P. Ning. A Beacon-Less Location Discovery Scheme for Wireless Sensor Networks. In *IEEE INFOCOM '05*, Miami, FL, USA, Mar 2005.

[70] S. Fang, C. Chuang, and C. Wang. Attack-resistant wireless localization using an inclusive disjunction model. *Communications, IEEE Transactions on*, 60(5):1209 – 1214, May 2012.

[71] H. Fathi, S. SeongHan, K. Kobara, and H. Imai. Protocols for authenticated anonymous communications. In *IEEE PIMRC '07*, pages 1–5, Sep 2007.

[72] FCC. Interference limits policy the use of harm claim thresholds to improve the interference tolerance of wireless systems, white paper, receivers and spectrum working group, fcc technological advisory council, 2013.

[73] F.C.C. Facilitation Opportunities for Flexible, Efficient, and Reliable Spectrum Use Employing Cognitive Radio Technologies. Technical Report ET Docket No. 03-108, FCC 03-322, NPRM & Order, Dec 2003.

[74] R. Feng, X. Guo, N. Yu, and J. Wan. Robust multihop localization for wireless sensor networks with unreliable beacons. *International Journal of Distributed Sensor Networks*, 2012, 2012.

[75] Simulator for Wireless Sensor. J-sim: A simulation and emulation environmentl. citeseer.ist.psu.edu/740832.html, Date Accessed Nov 2016.

[76] E. R. Frykberg and J. J. Tepas. Terrorist Bombings: Lessons Learned from Belfast to Beirut. *Annals of Surgery*, 208(5):567–576, 1988.

[77] S. Gao, L. Zhang, S. Fan, J. Wang, and Yu Deng. Dynamic Uplink Power Allocation with Hierarchical Interference Bound for Multi-cell Multi-user Cognitive Radio System. In *IEEE ITAP '10*, China, 2010.

[78] F. D. Garcia and B. Jacobs. Privacy-friendly energy-metering via homomorphic encryption. In *Security and Trust Management*, pages 226–238. Springer, 2011.

[79] G. Gaubatz, J. Kaps, and B. Sunar. *Security in Ad-hoc and Sensor Networks, Lecture Notes in Computer Science*, volume 3313/2005, chapter Public Key Cryptography in Sensor Networks-Revisited, pages 2–18. Springer Berlin / Heidelberg, 2005.

[80] D. Gerakoulis and E. Geraniotis. *CDMA Access and Switching for Terrestrial and Sattelite Networks*. John Wiley and Sons, 2001.

[81] A. Ghassemi, S. Bavarian, and L. Lampe. Cognitive radio for smart grid communications. In *IEEE SmartGridComm '10*, pages 297–302, Oct 2010.

[82] M. Ghofrani, M. Hassanzadeh, M. Etezadi-Amoli, and M. S. Fadali. Smart meter based short-term load forecasting for residential customers. In *IEEE NAPS '11*, pages 1–5. IEEE, 2011.

[83] L. Girod and D. Estrin. Robust Range Estimation using Acoustic and Multimodal Sensing. In *IEEE/RSJ '01*, pages 1312–1320, Maui, HI, USA, 2001.

[84] D. Goldenberg, A. Krishnamurthy, W. C. Maness, Y. R. Yang, A. Young, A. S. Morse, A. Savvides, and B. D. O. Anderson. Network Localization in Partially Localizable Networks. In *IEEE INFOCOM '05*, Miami, FL, USA, Mar 2005.

[85] Y. Gong, Y. Cai, Y. Guo, and Y. Fang. A privacy-preserving scheme for incentive-based demand response in the smart grid. *Smart Grid, IEEE Transactions on*, PP(99):1–1, 2015.

[86] greentechgrid. Arcadian's smart grid: Licensed spectrum network to own or rent, http://www.greentechmedia.com/articles/read/arcadians-utilityoffering-licensed-spectrum-to-own-or-rent, Sep 2009.

[87] M. Greis. *Tutorial for the Network Simulator "ns"*. VINT group, 2005. http://www.isi.edu/nsnam/ns/, Date Accessed Nov 2016.

[88] D. Grochocki, J. H. Huh, R. Berthier, R. Bobba, W. H. Sanders, A. A. Cárdenas, and J. G. Jetcheva. Ami threats, intrusion detection requirements and deployment recommendations. In *IEEE SmartGridComm '12*, pages 395–400. IEEE, 2012.

[89] GSGF. Global Smart Grid Federation, m2m: 800 million electric smart meters to be installed globally by 2020, 2014. http://www.globalsmartgridfederation.org/2014/02/28/m2m-800-million-electric-smart-meters-to-be-installed-globally-by-2020/, Date Accessed Nov 2016.

[90] S. Han, E. Chang, L. Gao, and T. Dillon. *Proceedings of the* 1^{st} *European Conference on Computer Network Defence School of Computing, University of Glamorgan, Wales, UK*, chapter Taxonomy of Attacks on Wireless Sensor Networks, pages 97–105. Springer London, 2006.

[91] J. Hao, E. Kang, D. Jackson, and J. Sun. Adaptive defending strategy for smart grid attacks. In *ACM SEGS '14*, pages 23–30, New York, NY, USA, 2014. ACM.

[92] L. Hao, S. Yang, S. Lu, and G. Chen. A dynamic anonymous p2p reputation system based on trusted computing technology. In *IEEE GLOBECOM '07*, pages 332–337, Nov 2007.

[93] C. Hartung, R. Han, C. Seielstad, and S. Holbrook. FireWxNet: A Multi-tiered Portable Wireless System for Monitoring Weather Conditions in Wildland Fire Environments. In *MobiSys '06*, pages 28–41, Uppsala, Sweden, 2006.

[94] S. He, G. Xuan, and L. Wu. Sequence Design for Cognitive FH-CDMA Systems. In *IEEE ICIEA '07*, China, May 2007.

[95] T. He, C. Huang, B. M. Blum, J. A. Stankovic, and T. Abdelzaher. Range-free Localization Schemes for Large Scale Sensor Networks. In *MOBICOM '03*, pages 81–95, San Diego, CA, USA, 2003.

[96] W. He, X. Liu, H. Nguyen, K. Nahrstedt, and T. Abdelzaher. Pda: Privacy-preserving data aggregation in wireless sensor networks. In *IEEE INFOCOM '07*, pages 2045–2053. IEEE, 2007.

[97] W. He, H. Nguyen, X. Liu, K. Nahrstedt, and T. Abdelzaher. ipda: An integrity-protecting private data aggregation scheme for wireless sensor networks. In *IEEE MILCOM '08*, pages 1–7. IEEE, 2008.

[98] X. He, X. Zhang, and C. Kuo. A distortion-based approach to privacy-preserving metering in smart grids. *Access, IEEE*, 1:67–78, 2013.

[99] J. Hightower and G. Borriello. Location Systems for Ubiquitous Computing. *Computer*, 34(8):57–66, Aug 2001.

[100] B. Hofmann-Wellenhof, H. Lichtenegger, and J. Collins. *Global Positioning System: Theory and Practice*. Springer Verlag, 1997.

[101] H. Holma and A. Toskala. *WCDMA for UMTS- HSPA Evoloution and LTE*. John Wiley and Sons, 2010.

[102] S. Hong, Y. Luming, W. Weiping, and D. Guihua. A delay demand-based anonymous communication mechanism. In *ChinaCom '06*, pages 1–5, Oct 2006.

[103] W. Hu, D. Willkomm, M. Abusubaih, J. Gross, G. Vlantis, M. Gerla, and A. Wolisz. Dynamic Frequency Hopping Communities for Efficient IEEE 802.22 Operation. *IEEE Commun. Magazine*, May 2007.

[104] J. Huang, H. Wang, Y. Qian, and C. Wang. Priority-based traffic scheduling and utility optimization for cognitive radio communication infrastructure-based smart grid. *Smart Grid, IEEE Transactions on*, 4(1):78–86, March 2013.

[105] Y. Huang, M. Esmalifalak, H. Nguyen, R. Zheng, Z. Han, H. Li, and L. Song. Bad data injection in smart grid: attack and defense mechanisms. *Communications Magazine, IEEE*, 51(1):27–33, Jan 2013.

[106] S. Imran, R. V. Karthick, and P. Visu. Dd-sarp: Dynamic data secure anonymous routing protocol for manets in attacking environments. In *ACM ICSTM '15*, pages 39–46, May 2015.

[107] S. Inshi and A. Youssef. Design and implementation of an online anonymous feedback system. In *IEEE QBSC '08*, pages 58–61, Jun 2008.

[108] Itron. Itron modular smart meter, em420i. https://www.itron.com/eu/technology/product-services-catalog/products/f/4/f/em420i, Date Accessed Nov 2016.

[109] M. Jadliwala, S. Upadhyaya, H. R. Rao, and R. Sharman. Security and Dependability Issues in Location Estimation for Emergency Sensor Networks. In *ACM WeB '05*, Venetian, Las Vegas, NV, USA, Dec 2005.

[110] M. Jadliwala, S. Zhong, S. J. Upadhyaya, C. Qiao, and J. P. Hubaux. Secure distance-based localization in the presence of cheating beacon nodes. *IEEE Transactions on Mobile Computing*, 9(6):810–823, 2010.

[111] A. Jain and B. Satish. Clustering based short term load forecasting using support vector machines. In *PowerTech, IEEE Bucharest '09*, pages 1–8. IEEE, 2009.

[112] A. Jain and B. Satish. Short term load forecasting by clustering technique based on daily average and peak loads. In *IEEE PES '09*, pages 1–7. IEEE, 2009.

[113] J. Jan, Y. Chen, and Y. Lin. The design of protocol for e-voting on the internet. In *IEEE ICCST '01*, pages 180–189, Oct 2001.

[114] M. Jawurek. *Privacy in Smart Grids*. PhD thesis, PhD thesis, Friedrich-Alexander-Universität Erlangen-Nürnberg, 2013.

[115] M. Jawurek, M. Johns, and K. Rieck. Smart metering de-pseudonymization. In *ACM ACSAC '27*, pages 227–236. ACM, 2011.

[116] X. Ji and H. Zha. Sensor Positioning in Wireless Ad-hoc Sensor. Networks using Multidimensional Scaling. In *IEEE INFOCOM '04*, Honk Kong, China, Mar 2004.

[117] T. Jiang, H. Wang, and Y. Zhang. Modeling channel allocation for multimedia transmission over infrastructure based cognitive radio networks. *Systems Journal, IEEE*, 5(3):417–426, Sep 2011.

[118] P. Juang, H. Oki, Y. Wang, M. Martonosi, L. S. Peh, and D. Rubenstein. Energy-efficient Ccomputing for Wildlife Tracking: Design Tradeoffs and Early Experiences with ZebraNet. *ACM SIGOPS Operating Systems Review*, 36(5):96–107, 2002.

[119] V. k. Garg. *Wireless Communication and Networking*. Morgan Kaufmann Publishers, 2007.

[120] G. Kalogridis, C. Efthymiou, S. Z. Denic, T. Lewis, and R. Cepeda. Privacy for smart meters: Towards undetectable appliance load signatures. In *IEEE SmartGridComm '10*, pages 232–237. IEEE, 2010.

[121] R. Kannan, L. Ray, and S. S. Iyengar. Randomized message forwarding with equalized incoming/outgoing traffic rate: A mechanism for ensuring anonymous communication. In *IEEE ICISIP '05*, pages 183–188, Dec 2005.

[122] H. Karl and A. Willig. *Protocols and Architectures for Wireless Sensor Networks*. John Wiley and Sons, 2005.

[123] B. Karp and H. T. Kung. GPSR: Greedy Perimeter Stateless Routing for Wireless Networks. In *ACM MOBICOM '00*, Boston, MA, USA, Aug 2000. ACM SIGMOBILE.

[124] M. Kashef, M. Abdallah, A. Elezabi, and M. Khairy. System parameter selection for asymmetric underlay cdma networks with interference-minimizing code asssignment. In *IEEE SPAWC '09*, pages 722–726, Jun 2009.

[125] M. J. Kearns. *The computational complexity of machine learning*. ACM distinguished dissertations. Cambridge, Mass. MIT Press, 1990. Revision of the author's thesis (Ph. D.–Harvard University, 1989).

[126] D. Kelly, R. Raines, R. Baldwin, B. Mullins, and M. Grimaila. Towards a taxonomy of wired and wireless anonymous networks. In *IEEE ICC '09*, pages 1–8, Jun 2009.

[127] M. Kennedy. Leveraging investment in fiber optic communications, IEEE smart grid newsletter, june 2011. http://smartgrid.ieee.org/june-2011/105-leveraging-investment-in-fiber-optic-communications, Date Accessed Nov 2016.

[128] V. Khoury, S. Vassilaras, and C. B. Papadias. CR-DMAC: A MAC Protocol for Cognitive Radio Networks with Directional Antennas. In *ACM CogART '11*, Spain, Oct 2011.

[129] W. Kleiminger, C. Beckel, T. Staake, and S. Santini. Occupancy detection from electricity consumption data. In *ACM BuildSys '13*, pages 1–8. ACM, 2013.

[130] Teuvo Kohonen. The self-organizing map. *Neurocomputing*, 21(13):1 – 6, 1998.

[131] J. Z. Kolter and M. J. Johnson. Redd: A public data set for energy disaggregation research. In *ACM KDD '11*, volume 25, pages 59–62. Citeseer, 2011.

[132] T. Kucukdeniz. Long term electricity demand forcesting: An alternative approach with support vector machines. *İÜ Mühendislik Bilimleri Dergisi*, 1(1), 2010.

[133] T. Kwon, H. Song, J. Lee, Y. Kim, J. Lee, and D. Hong. A Power Division Reuse Partitioning Scheme with Half Frequency Reuse Factor for OFDMA Downlink Systems. In *IEEE ICC '08*, China, May 2008.

[134] L. Lazos and R. Poovendran. SeRLoc: Secure Range-Independent Localization for Wireless Sensor Networks. In *IEEE WiSe '04*, pages 21–30, Philadelphia, PA, USA, 2004.

[135] L. Lazos, R. Poovendran, and S. Čapkun. Rope: RObust Position Estimation in Wireless Sensor Networks. In *ACM IPSN '05*, page 43, Los Angeles, CA, USA, 2005.

[136] Y. Lee, S. Lee, J. Y. Hwang, B. H. Chung, and D. G. Lee. Anonymous access control framework based on group signature. In *ITCS '10*, pages 1–5, Aug 2010.

[137] F. Li, B. Luo, and P. Liu. Secure information aggregation for smart grids using homomorphic encryption. In *IEEE SmartGridComm '10*, pages 327–332. IEEE, 2010.

[138] M. Li, S. N. Batalama, D. A. Pados, T. Melodia, M. J. Medley, and J. D. Matyjas. Cognitive Code-Division Links with Blind Primary-System Identification. *IEEE Trans. on Wireless Communication*, Nov 2011.

[139] Z. Li, W. Trappe, Y. Zhang, and B. Nath. Robust Statistical Methods for Securing Wireless Localization in Sensor Networks. In *ACM IPSN '05*, page 12, Los Angeles, CA, USA, 2005.

[140] W. D. Lin and J. Jan. A wireless-based authentication and anonymous channels for large scale area. In *ACM ISCC '01*, pages 36–41, 2001.

[141] M. Lisovich, D. Mulligan, and S. Wicker. Inferring personal information from demand-response systems. *Security & Privacy, IEEE*, 8(1):11–20, 2010.

[142] D. Liu, P. Ning, and W. Du. Attack-Resistant Location Estimation in Sensor Networks. In *IPSN '05*, pages 99–106, Los Angeles, CA, USA, Apr 2005.

[143] D. Liu, P. Ning, and W. Du. Detecting Malicious Beacon Nodes for Secure Location Discovery in Wireless Sensor Networks. In *IEEE ICDCS '05*, pages 609–619, Columbus, OH, USA, Jun 2005.

[144] J. Liu, Y. Zhang, and F. Zhao. Robust Distributed Node Localization with Error Management. In *MobiHoc '06*, pages 250–261, Florence, Italy, 2006.

[145] S. Liu, L. Lazos, and M. Krunz. Cluster-based control channel allocation in opportunistic cognitive radio networks. *Mobile Computing, IEEE Transactions on*, 11(10):1436–1449, Oct 2012.

[146] Y. Liu, P. Ning, and M. K. Reiter. False data injection attacks against state estimation in electric power grids. *ACM Transaction Information System Security*, 14(1):13:1–13:33, June 2011.

[147] L. Lixin, L. Chaoling, and Z. Yanzhou. A remote anonymous attestation scheme with improved privacy ca. In *IEEE MINES '09*, volume 1, pages 153–157, Nov 2009.

[148] C. H. Lo and N. Ansari. Decentralized controls and communications for autonomous distribution networks in smart grid. *Smart Grid, IEEE Transactions on*, 4(1):66–77, Mar 2013.

[149] C. H. Lo and N. Ansari. Decentralized controls and communications for autonomous distribution networks in smart grid. *IEEE transactions on smart grid*, 4(1):66–77, 2013.

[150] J. Lopez. Unleashing Public-key Cryptography in Wireless Sensor Networks. *Journal of Computer Security*, 14(5):469–482, 2006.

[151] K. Lorincz, D. Malan, T. R. F. Fulford-Jones, A. Nawoj, A. Clavel, V. Shnayder, G. Mainland, S. Moulton, and M. Welsh. Sensor networks for emergency response: Challenges and opportunities. *IEEE Pervasive Computing, Special Issue on Pervasive Computing for First Response*, 3(4):16–23, October 2004.

[152] RF DataTech Ltd. RF DataTech - Advanced Wireless Data Solutions. http://www.rfdatatech.co.uk/, Date Accessed Nov 2016.

[153] R. Lu, X. Lin, H. Zhu, P. H. Ho, and X. Shen. A novel anonymous mutual authentication protocol with provable link-layer location privacy. *Vehicular Technology, IEEE Transactions on*, 58(3):1454–1466, Mar 2009.

[154] X. Lu, G. Fan, and R. Hao. A dynamic token passing mac protocol for mobile ad hoc networks. In *ACM IWCMC '06*, pages 743–748. ACM, 2006.

[155] Y. Luo, S. S. Cheung, and Y. Shuiming. Anonymous biometric access control based on homomorphic encryption. In *ICME '09*, pages 1046–1049, Jun 2009.

[156] D. N. Mah, J. Marinus, P. Hills, and J. Tao. Consumer perceptions of smart grid development: Results of a hong kong survey and policy implications. *Energy Policy '12*, 49:204–216, 2012.

[157] P. Maille, P. Reichl, and B. Tuffin. *of threats and costs: a game-theoretic approach to security risk management*. Springer, 2013.

[158] R. Majumder, G. Bag, and K. H. Kim. Power sharing and control in distributed generation with wireless sensor networks. *Smart Grid, IEEE Transactions on*, 3(2):618–634, Jun 2012.

[159] G. Mao, B. D. O. Anderson, and B. Fidan. Path Loss Exponent Estimation for Wireless Sensor Network Localization. *Computer Networks*, 51(10):2467–2483, 2007.

[160] D. Mashima. Authenticated down-sampling for privacy-preserving energy usage data sharing. In *IEEE SmartGridComm '15*. IEEE, 2015.

[161] D. Mashima and A. Roy. Privacy preserving disclosure of authenticated energy usage data. In *IEEE SmartGridComm '14*, pages 866–871, Nov 2014.

[162] S. McLaughlin, P. McDaniel, and W. Aiello. Protecting consumer privacy from electric load monitoring. In *ACM CCS '11*, pages 87–98. ACM, 2011.

[163] S. McLaughlin, D. Podkuiko, and P. McDaniel. Energy theft in the advanced metering infrastructure. In *Critical Information Infrastructures Security*, pages 176–187. Springer, 2009.

[164] S. Merrill, N. Basalp, J. Biskup, E. Buchmann, C. Clifton, B. Kuijpers, W. Othman, and E. Savas. Privacy through uncertainty in location-based services. In *IEEE MDM '13*, volume 2, pages 67–72. IEEE, 2013.

[165] L. Ming, S. N. Batalama, D. A. Pados, T. Melodia, M. J. Medley, and J. D. Matyjas. Cognitive code-division links with blind primary-system identification. *Wireless Communications, IEEE Transactions on*, 10(11):3743–3753, Nov 2011.

[166] S. Misra, G. Xue, and S. Bhardwaj. Secure and Robust Localization in a Wireless Ad Hoc Environment. *IEEE Transactions on Vehicular Technology*, 2008.

[167] M.S. Mohammadi and M. M. Taheri. Blind Source Separation and Tracking of Multiple Frequency Hopping Signals for Cognitive Radio Commuication. In *IEEE WD '08*, UAE, Nov 2008.

[168] A. H. Mohsenian-Rad and A. Leon-Garcia. Distributed internet-based load altering attacks against smart power grids. *IEEE Transactions on Smart Grid*, 2(4):667–674, 2011.

[169] D. Moore, J. Leonard, D. Rus, and S. Teller. Robust Distributed Network Localization with Noisy Range Measurements. In *SenSys '04*, pages 50–61, Baltimore, MD, USA, 2004.

[170] R. Moses, D. Krishnamurthy, and R. Patterson. A self-localization method for wireless sensor networks. *Eurasip Journal on Applied Signal Processing, Special Issue on Sensor Networks*, 2003(4):148–158, Mar 2003.

[171] J. Newsome, E. Shi, D. Song, and A. Perrig. The sybil attack in sensor networks: analysis & defenses. In *ACM IPSN '04*, pages 259–268. ACM, 2004.

[172] P. H. Nguyen, W. L. Kling, G. Georgiadis, M. Papatriantafilou, L. A. Tuan, and L. Bertling. Distributed routing algorithms to manage power flow in agent-based active distribution network. In *ISGT Europe, IEEE PES '10*, pages 1–7, Oct 2010.

[173] P. H. Nguyen, W. L. Kling, and P. F. Ribeiro. Agent-based power routing in active distribution networks. In *ISGT Europe, IEEE PES '10*, pages 1–6, Dec 2011.

[174] Y. Q. Ni, D. H. Nyang, and X. Wang. A-kad: an anonymous p2p protocol based on kad network. In *IEEE MASS '09*, pages 747–752, Oct 2009.

[175] D. Niculescu and B. Nath. DV based Positioning in Ad hoc Networks. *Journal of Telecommunication Systems*, 2003.

[176] J. R. Norris. *Markov Chains*. Number no. 2008 in Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 1998.

[177] Y. Ouyang, Z. Le, Y. Xu, N. Triandopoulos, S. Zhang, J. Ford, and F. Makedon. Providing anonymity in wireless sensor networks. In *ICPS '07*, pages 145–148, 2007.

[178] H. Pan, E. Hou, and N. Ansari. E-note: An e-voting system that ensures voter confidentiality and voting accuracy. In *IEEE ICC '12*, pages 825–829, Jun 2012.

[179] J. Pan and J. Li. Masr: An efficient strong anonymous routing protocol for mobile ad hoc networks. In *IEEE MASS '09*, pages 1–6, Sep 2009.

[180] P. Papadimitratos, M. Poturalski, P. Schaller, P. Lafourcade, D. Basin, S. Čapkun, and J. Hubaux. Secure Neighborhood Discovery: A Fundamental Element for Mobile Ad Hoc Networking. *IEEE Communications Magazine*, 46(2), 2008.

[181] S. Papadimitriou, F. Li, G. Kollios, and P. S. Yu. Time series compressibility and privacy. In *ACM VLDB '07*, pages 459–470. VLDB Endowment, 2007.

[182] R. L. Pickholtz, D. L. Schilling, and L. B. Milstein. Revisions to "theory of spread-spectrum communications - a tutorial". *Communications, IEEE Transactions on*, 32(2):211–212, Feb 1984.

[183] W. Pires, T. H. P. Figueiredo, H. C. Wong, and A. F. Loureiro. Malicious Node Detection in Wireless Sensor Networks. In *IEEE IPDPS '04*, page 24, Santa Fe, NM, USA, Apr 2004.

[184] H. V. Poor and O. Hadjiliadis. *Quickest detection*, volume 40. Cambridge University Press Cambridge, 2009.

[185] N. B. Priyantha, A. Chakraborty, and H. Balakrishnan. The Cricket Location-Support System. In *ACM MOBICOM '00*, pages 32–43, Boston, MA, USA, August 2000. ACM SIGMOBILE.

[186] R. C. Qiu, Z. Hu, Z. Chen, N. Guo, R. Ranganathan, S. Hou, and G. Zheng. Cognitive radio network for the smart grid: Experimental system architecture, control algorithms, security, and microgrid testbed. *Smart Grid, IEEE Transactions on*, 2(4):724–740, Dec 2011.

[187] Q. Qu, L. B. Milstein, and D. R. Vaman. Cognitive radio based multi-user resource allocation in mobile ad hoc networks using multi-carrier cdma modulation. *Selected Areas in Communications, IEEE Journal on*, 26(1):70–82, Jan 2008.

[188] M. Raciti and S. Nadjm-Tehrani. Embedded cyber-physical anomaly detection in smart meters. In *Critical Information Infrastructures Security*, pages 34–45. Springer, 2012.

[189] M. Rahman, M. Manshaei, E. Al-Shaer, and M. Shehab. Secure and private data aggregation for energy consumption scheduling in smart grids. *Dependable and Secure Computing, IEEE Transactions on*, PP(99):1–1, 2015.

[190] M. A. Rahman, M. H. Manshaei, E. S. Al-Shaer, and M. Shehab. Secure and private data aggregation for energy consumption scheduling in smart grids. *IEEE Transactions on Dependable and Secure Computing*, (99):1–1, 2015.

[191] S. Rajagopalan, L. Sankar, S. Mohajer, and V. H. Poor. Smart meter privacy: A utility-privacy framework. In *IEEE SmartGridComm '11*, pages 190–195. IEEE, 2011.

[192] R. Rajbanshi, Q. Chen, A. M. Wyglinski, G. J. Minden, and J. B. Evans. Quantitative Comarision of Agile Modulation Techniques for Cognitive Radio Transceivers. In *IEEE CCNC '07*, USA, Jan 2007.

[193] T. S. Rappaport. *Wireless Communications: Principles and Practice, 2^{nd} Edition*, chapter Mobile Radio Propagation: Large-Scale Path Loss. Pearson Education, Inc., 2003.

[194] V. Rastogi and S. Nath. Differentially private aggregation of distributed time-series with transformation and encryption. In *ACM SIGMOD '10*, pages 735–746. ACM, 2010.

[195] I. Ray and N. Narasimhamurthi. An anonymous electronic voting protocol for voting over the internet. In *ACM WECWIS '01*, pages 188–190, 2001.

[196] S. Ray, R. Ungrangsi, F. Pellegrini, A. Trachtenberg, and D. Starobinski. Robust Location Detection in Emergency Sensor Networks. In *IEEE INFOCOM'03*, pages 1044–1053, San Francisco, CA, USA, Mar 2003.

[197] M. G. Reed, P. F. Syverson, and D. M. Goldschlag. Proxies for anonymous routing. In *ACM CSAC '96*, pages 95–104, Dec 1996.

[198] A. Rial and G. Danezis. Privacy-preserving smart metering. In *ACM WPES' 11*, pages 49–60. ACM, 2011.

[199] I. Richardson and M. Thomson. One-minute resolution domestic electricity use data, 2008-2009. colchester, essex: Uk data archive [distributor], sn: 6583, october 2010.

[200] R. Rivest, A. Shamir, and Y. Tauman. How to leak a secret. In *ASIACrypt*, 2001.

[201] C. Robinson. Sensors bolster army prowess. *SIGNAL Magazine, AFCEA's International Journal*, 2004. http://www.afcea.org/signal/articles/anmviewer.asp?a=30, Date Accessed Nov 2016.

[202] I. Rouf, Mustafa H, M. Xu, W. Xu, R. Miller, and M. Gruteser. Neighborhood watch: security and privacy analysis of automatic meter reading systems. In *ACM CCS '12*, pages 462–473. ACM, 2012.

[203] S. D. Roy, S. Kundu, G. Ferrari, and R. Raheli. Performance evaluation of cognitive radio cdma networks with spectrum sensing. In *IEEE ICC '10*, pages 1–5, May 2010.

[204] N. Safa, S. Sarkar, R. Safavi-Naini, and M. Ghaderi. Secure localization using dynamic verifiers. In Vijay Atluri and Claudia Diaz, editors, *ESORICS '11*, volume 6879 of *Lecture Notes in Computer Science*, pages 1–20. Springer Berlin / Heidelberg, 2011.

[205] H. Salameh, M. Krunz, and O. Younis. Cooperative Adaptive Spectrum Sharing in Cognitive Radio Networks. *Tran. on Networking*, 2010.

[206] P. Samadi, H. Mohsenian-Rad, R. Schober, and V. W. S. Wong. Advanced demand side management for the future smart grid using mechanism design. *Smart Grid, IEEE Transactions on*, 3(3):1170–1180, Sep 2012.

[207] L. Sankar, S. R. Rajagopalan, S. Mohajer, and H. V. Poor. Smart meter privacy: A theoretical framework. *IEEE Transactions on Smart Grid*, 4(2):837–846, 2013.

[208] D. Sarath, K. E. Nolan, P. D. Sutton, and L. E. Doyle. Enabling Dynamic Spectrum Access using SS-MC-CDMA. In *IEEE CrownCom'07*, USA, Aug 2007.

[209] D. Sarath, K. E. Nolan, P. D. Sutton, and L. E. Doyle. Exploring the Reconfigurability Options of Multi-carrier CDMA in Cognitive Radio Systems. In *IEEE PIMRC '07*, Greece, Sep 2007.

[210] N. Sastry, U. Shankar, and D. Wagner. Secure Verification of Llocation Claims. In *WiSe '03*, pages 1–10, San Diego, CA, USA, 2003.

[211] M. Savi, C. Rottondi, and G. Verticale. Evaluation of the precision-privacy tradeoff of data perturbation for smart metering. *IEEE Transactions on Smart Grid*, 6(5):2409–2416, 2015.

[212] A. Savvides, W. Garber, S. Adlakha, R. Moses, and M. B. Srivastava. On the Error Characteristics of Multihop Node Localization in Ad-hoc Sensor Networks. In *IPSN '03*, pages 317–332, Palo Alto, CA, USA, Apr 2003.

[213] A. Savvides, C. Han, and M. B. Strivastava. Dynamic Fine-grained Localization in Ad-Hoc Networks of Sensors. In *MobiCom '01*, pages 166–179, Rome, Italy, 2001.

[214] U. Sawant. *Grid-based Coordinated Routing in Wireless Sensor Networks*. PhD thesis, University of North Texas, 2006.

[215] J. E. Seem. Pattern recognition algorithm for determining days of the week with similar energy consumption profiles. *Energy and Buildings '05*, 37(2), 2005.

[216] R. Sevlian and R. Rajagopal. Short term electricity load forecasting on varying levels of aggregation, arxiv preprint arxiv:1404.0058, ieee, 2014.

[217] Y. Shang, W. Ruml, Y. Zhang, and M. Fromherz. Localization from Connectivity in Sensor Networks. *IEEE Transactions on Parallel and Distributed Systems*, 15(11):961–974, 2004.

[218] M. Shao, Y. Yang, S. Zhu, and G. Cao. Towards statistically strong source anonymity for sensor networks. In *IEEE INFOCOM '08*. IEEE, 2008.

[219] E. Shi, R. Chow, T. H. Chan, D. Song, and E. Rieffel. Privacy-preserving aggregation of time-series data. In *NDSS '11*, 2011.

[220] K. Shigeki, M. Soshi, and A. Miyaji. An agent-based model of anonymous communication protocols. In *ACM WET ICE '01*, pages 177–182, 2001.

[221] V. Shnayder, B. Chen, K. Lorincz, T. R. F. Fulford-Jones, and M. Welsh. Sensor networks for medical care. Technical Report TR-08-05, Harvard University, April 2005.

[222] R. Shokri, G. Theodorakopoulos, J. Y. Le Boudec, and J. P. Hubaux. Quantifying location privacy. In *IEEE SP '11*, pages 247–262, 2011.

[223] Sierra Wireless. Sierra wireless selected as cellular modem provider for itron's innovative openway smart grid solution. http://www.sierrawireless.com, 2014.

[224] R. P. Singh, P. X. Gao, and D. Lizotte. On hourly home peak load prediction. In *IEEE SmartGridComm '12*, pages 163–168. IEEE, 2012.

[225] S. Slijepcevic, S. Megerian, and M. Potkonjak. Location Errors in Wireless Embedded Sensor Networks: Sources, Models, and Effects on Applications. *SIGMOBILE '02*, 6(3):67–78, 2002.

[226] M. R. Souryal and N. Golmie. Analysis of Advanced Metering over a Wide Area Cellular Network. In *IEEE SmartGridComm '11*, 2011.

[227] C. Stevenson, G. Chouinard, Z. Lei, W. Hu, S. J. Shellhammer, and W. Caldwell. Ieee 802.22: The first cognitive radio wireless regional area network standard. *Communications Magazine, IEEE*, 47(1):130–138, January 2009.

[228] R. Stoleru and J. A. Stankovic. Probability Grid: A Location Estimation Scheme for Wireless Sensor Networks. In *SECON '04*, pages 430–438, Santa Clara, CA, USA, October 2004.

[229] S. Tamura, H. A. Haddad, H. Tsurugi, and A. K. M. Rokibul. Mechanisms for anonymous memories. In *IEEE ICIT '09*, pages 1–6, Feb 2009.

[230] A. S. Tanenbaum, C. Gamage, and B. Crispo. Taking Sensor Networks from the Lab to the Jungle. *Computer*, 39(8):98–100, 2006.

[231] A. Tcixcira, S. Amin, H. Sandberg, K. H. Johansson, and S. S. Sastry. Cyber security analysis of state estimators in electric power systems. In *ACM CDC '10*, pages 5991–5998, Dec 2010.

[232] Telefnica Digital. The smart meter revolution- towards a smarter future. Jan 2014. https://iot.telefonica.com/multimedia-resources/the-smart-meter-revolution-towards-a-smarter-future, Date Accessed Nov 2016.

[233] S. Thaskani, K. V. Kumar, and G. R. Murthy. Energy efficient cross-layer design protocol by using token passing mechanism for wsn. In *IEEE Symposium on Computers & Informatics*, pages 572–575. IEEE, 2011.

[234] W. Tollefsen, M. Pepe, D. Myung, M. Gaynor, M. Welsh, and S. Moulton. iRevive, a Pre-hospital Mobile Database for Emergency Medical Services. *IJHTM '04*, May-August 2004.

[235] N. C. Truong, J. McInerney, L. Tran-Thanh, E. Costanza, and S. Ramchurn. Forecasting multi-appliance usage for smart home energy management. In *IEEE IJCAI '13*, pages 2908–2914, 2013.

[236] H. Tsai and A. Harwood. A scalable anonymous server overlay network. In *IEEE AINA '06*, volume 1, pages 973–978, Apr 2006.

[237] H. R. Tseng. A secure and privacy-preserving communication protocol for v2g networks. In *IEEE WCNC '12*, pages 2706–2711, Apr 2012.

[238] M. A. Tubaishat and S. Madria. Sensor Networks: An Overview. *IEEE Potentials*, 22(2):20–23, Apr 2003.

[239] R. Urgaonkar and M. J. Neely. Opportunistic scheduling with reliability guarantees in cognitive radio networks. In *IEEE INFOCOM '08*, pages –, Apr 2008.

[240] U.S.-Canada Power System Outage Task Force. Final report on the august 14, 2003 blackout in the united states and canada, https://reports.energy.gov/b-f-web-part1.pdf, 2004.

[241] S. Čapkun and J. P. Hubaux. Secure Positioning of Wireless Devices with Application to Sensor Networks. In *IEEE INFOCOM '05*, pages 1917–1928, Miami, FL, USA, Mar 2005.

[242] S. Čapkun and J. P. Hubaux. Secure Positioning in Wireless Networks. *IEEE Journal on Selected Areas in Communications (JSAC)*, 24(2):221–232, 2006.

[243] V. S. Venkitachalam, V. Namboodiri, S. Joseph, E. Dee, and C. A. Burdsal. What, why, and how: Surveying what consumers want in new mobile phones. *IEEE Consumer Electronics Magazine*, 4(2):54–59, Apr 2015.

[244] Viscovery SOMine Tool. Data mining suite, eudaptics software gmbh. https://www.viscovery.net, Date Accessed Nov 2016.

[245] R. Vogt, I. Nikolaidis, and P. Gburzynski. Divalia: a practical framework for anonymous peer-to-peer file exchange in wireless ad-hoc networks. In *IEEE CNSR '06*, pages 8 pp.–156, May 2006.

[246] S. Vural and E. Ekici. On multihop distances in wireless sensor networks with random node locations. *Mobile Computing, IEEE Transactions on*, 9(4):540–552, Apr 2010.

[247] B. Wang and D. Zhao. Performance analysis in cdma-based cognitive wireless networks with spectrum underlay. In *IEEE GLOBECOM '08*, pages 1–6, Nov 2008.

[248] H. Wang, Y. Qian, and H. Sharif. Multimedia communications over cognitive radio networks for smart grid applications. *Wireless Communications, IEEE*, 20(4):125–132, Aug 2013.

[249] W. Wang, G. Duan, J. Wang, and J. Chen. An anonymous communication mechanism without key infrastructure based on multi-paths network coding. In *IEEE GLOBECOM '09*, pages 1–6, Nov 2009.

[250] R. Want, A. Hopper, V. Falcao, and J. Gibbons. The Active Badge Location System. *ACM Transaction on Information Systems*, pages 91–102, Jan 1992.

[251] T. Wauters, F. D. Turck, and C. Develder. Overlay networks for smart grids. *IEEESTD '13*, page 223, 2013.

[252] Y. Wei, Z. Yu, and Y. Guan. Location Verification Algorithms for Wireless Sensor Networks. In *ICDCS '07*, page 70, Toronto, Canada, 2007.

[253] M. Weiss, A. Helfenstein, F. Mattern, and T. Staake. Leveraging smart meter data to recognize home appliances. In *IEEE PerCom '12*, pages 190–197. IEEE, 2012.

[254] T. Wigren. Adaptive Enhanced Cell-ID Fingerprinting Localization by Clustering of Precise Position Measurements. *IEEE Transactions on Vehicular Technology*, Sep 2008.

[255] T. K. Wijaya, S. Humeau, F. R. J. Samuel, M. Vasirani, and K. Aberer. Individual, Aggregate, and Cluster-based Aggregate Forecasting of Residential Demand. Technical report, Lausanne, Switzerland, 2014.

[256] X. Wu. Applying pseudonimity for anonymous data delivery in location-aware mobile ad hoc networks. *Vehicular Technology, IEEE Transactions on*, 55(3):1062–1073, May 2006.

[257] Y. Wu. A novel anonymous communication strategy based on structured peer-to-peer overlay networks. In *IEEE ISITAE '07*, pages 581–585, Nov 2007.

[258] J. Xiao, L. Ren, and J. Tan. Research of TDOA Based Self-localization Approach in Wireless Sensor Network. In *IEEE IROS '06*, pages 2035–2040, Beijing, China, Oct 2006.

[259] Y. Xiao. *Security and Privacy in Smart Grids*. CRC press Taylor and Francis Group, 2014.

[260] Y. Xu, G. Chen, J. Ford, and F. Makedon. *Critical Infrastructure Protection, IFIP International Federation for Information Processing*, volume 253/2007, chapter Detecting Wormhole Attacks in Wireless Sensor Networks, pages 267–279. Springer Boston, 2007.

[261] F. Ye, H. Luo, S. Lu, and L. Zhang. Statistical en-route filtering of injected false data in sensor networks. *IEEE Journal on Selected Areas in Communications*, 23(4):839–850, 2005.

[262] K. Yedavalli, B. Krishnamachari, S. Ravula, and B. Srinivasan. Ecolocation: A Sequence Based Technique for RF-only Localization in Wireless Sensor Networks. In *IPSN '05*, Los Angeles, CA, USA, Apr 2005.

[263] R. Yu, C. Zhang, X. Zhang, L. Zhou, and K. Yang. Hybrid spectrum access in cognitive-radio-based smart-grid communications systems. *Systems Journal, IEEE*, 8(2):577–587, Jun 2014.

[264] R. Yu, Y. Zhang, S. Gjessing, C. Yuen, S. Xie, and M. Guizani. Cognitive radio based hierarchical communications infrastructure for smart grid. *Network, IEEE*, 25(5):6–14, Sep 2011.

[265] J. Zhang and H. Chen. Efficient provable secure id-based anonymous signcryption scheme. In *IEEE PACCS '09*, pages 415–418, May 2009.

[266] J. Zhang, H. Duan, W. Liu, and J. Wu. Analysis of anonymity in p2p anonymous communication systems. In *IEEE WAINA '10*, pages 860–865, Apr 2010.

[267] X. Zhang and H. Su. Opportunistic spectrum sharing schemes for cdma-based uplink mac in cognitive radio networks. *Selected Areas in Communications, IEEE Journal on*, 29(4):716–730, Apr 2011.

[268] Y. Zhang, W. Liu, and W. Lou. Anonymous communications in mobile ad hoc networks. In *IEEE INFOCOM '05*, volume 3, pages 1940–1951 vol. 3, Mar 2005.

[269] S. Zhong, M. Jadliwala, S. Upadhyaya, and C. Qiao. Towards a Theory of Robust Localization Against Malicious Beacon Nodes. In *IEEE INFOCOM '08*, pages 1391–1399, Phoenix, AZ, USA, Apr 2008.

[270] R. Zhou, X. li, V. Chakravarthy, B. Wang, and Z. Wu. Inter-Carrier Interference Self-Cancellation in Synchronous Downlink MC-CDMA System. In *ACM IWCMC '09*, Germany, Jun 2009.

[271] H. Zhuang, Z. Luo, J. Zhang, and H. Yanikomeroglu. Hierarchical and adaptive spectrum sensing in cognitive radio based multi-hop cellular networks. In *VTC '10*, pages 1–6, Sep 2010.