energies

MDPI

Article

Securing Metering Infrastructure of Smart Grid: A Machine Learning and Localization Based Key Management Approach

Imtiaz Parvez, Arif I. Sarwat *, Longfei Wei and Aditya Sundararajan

Department of Electrical and Computer Engineering, Florida International University, Miami, FL 33174, USA; iparv001@fiu.edu (I.P.); lwei004@fiu.edu (L.W.); asund005@fiu.edu (A.S.)

* Correspondence: asarwat@fiu.edu; Tel.: +1-305-348-4941

Academic Editor: Chunhua Liu

Received: 20 April 2016; Accepted: 22 August 2016; Published: 29 August 2016

Abstract: In smart cities, advanced metering infrastructure (AMI) of the smart grid facilitates automated metering, control and monitoring of power distribution by employing a wireless network. Due to this wireless nature of communication, there exist potential threats to the data privacy in AMI. Decoding the energy consumption reading, injecting false data/command signals and jamming the networks are some hazardous measures against this technology. Since a smart meter possesses limited memory and computational capability, AMI demands a light, but robust security scheme. In this paper, we propose a localization-based key management system for meter data encryption. Data are encrypted by the key associated with the coordinate of the meter and a random key index. The encryption keys are managed and distributed by a trusted third party (TTP). Localization of the meter is proposed by a method based on received signal strength (RSS) using the maximum likelihood estimator (MLE). The received packets are decrypted at the control center with the key mapped with the key index and the meter's coordinates. Additionally, we propose the k-nearest neighbors (kNN) algorithm for node/meter authentication, capitalizing further on data transmission security. Finally, we evaluate the security strength of a data packet numerically for our method.

Keywords: advanced metering infrastructure (AMI); data security; key management system; k-nearest neighbors (kNN); received signal strength (RSS); smart city; smart meter; smart grid

1. Introduction

The smart grid is the modern electric power system utilizing an innovative communication and distribution system to deliver electricity to end users with improved monitoring, control and efficiency. A touted feature of the smart grid is the interaction among its entities using bidirectional communication. advanced metering infrastructure (AMI) is the distribution-level building block of the smart grid, consisting of millions of meters. The data of energy consumption are collected and reported by smart meters to the control center of the service provider (SP) periodically (typically with a resolution of 15 min) different from conventional meters, which record the entire monthly consumption data. It also allows consumers to engage in the electricity trade, more formally called "net metering", by selling surplus power back to the grid. AMI caters to the SP, the control and monitoring for outage management, demand response, disaster prevention and disaster recovery. Consequently, the communication in AMI is bidirectional [1–6]. The assessment of current methodologies employed by smart meters and their mesh/hierarchical connected wired/wireless network constituting the AMI are highly recommended considering the fact that it is the most imperative aspect of the smart grid from the perspective of utility companies, as well as the consumers.

Cyber physical systems (CPS) and the reliability of the smart grid have been the key points of interest, where a system needs to be designed to detect and prevent an unauthorized access by integrating both cyber and physical components of the grid [7]. The attacker might want to decode data packets, gain control and command over various components of the smart grid, inject false commands, jam the network and take over the control of the system. Smart meters are the primary basis for the collection of consumer usage data through access points (APs). Observing the power consumption data and usage patterns of electricity, a thief/attacker can learn the presence or absence of consumers at home and, thus, poses a greater threat for the community.

In 2013, the U.S. electric utilities had 51,924,502 AMI smart meter installations of which about 89% were residential customer installations [8]. These meters mainly consist of in-built full-duplex communication mode with periodical/on-demand reception and transmission of data. Different solutions for various attacks are proposed based on the usage of electricity in residential areas and security protocols involving various wireless local area networks (WLANs) [9–11]. An experimental setup was performed to analyze the routine usage of electricity corresponding to the time of the day. It was observed that easily-identifiable loads, such as boilers, directly corresponded to the time when laundry, meals and showers were taken. These data can be utilized by potential hackers to break into vacant residences [12].

The main hindrance of implementing security schemes in AMI is the limited memory and computational capability of smart meters. Additionally, AMI is a huge network comprised of thousands of meters. This requires AMI to have a light, but robust security scheme. In the geographical coordinate-based encryption scheme [13–15], for localization, the global positioning system (GPS) was proposed. However, GPS does not work well in some places, such as inside a multi-storied building, hilly places, as well as coordinates derived by GPS will expose the exact location of the consumer house. Furthermore, it exposes the exact position of the meter/consumer.

In this paper, we propose a key management-based security scheme utilizing the location of meters, derived from received signal strength (RSS) of the radio signal. The localization of meters by the RSS-based method will create a local positioning map different from the geographic coordinate system, in which every meter has its own coordinate. For data encryption, secret keys mapped with the coordinate points of the meters and a random index are proposed in our technique. The keys are distributed among the meters periodically by a trusted third party (TTP) of the key management system. Furthermore, we introduce the k-nearest neighbors (kNN) algorithm for meter authentication during the transport of data packets. The kNN algorithm is a technique used to predict class labels of unknown data [16–18]. The kNN classifier is simple, efficient and easy to implement. It is one of the most widely-used algorithms in pattern evaluation, text characterization, diagnosis of cancer and many more. In a real-world scenario, there are many datasets with little or no prior knowledge about their distribution. kNN is amongst the best choice for the classification with a dataset with little or no prior knowledge. For these reasons, the combination of data encryption by secret key and node authentication using the kNN algorithm provides a potential solution for AMI.

The rest of this paper is organized as follows: Section 2 provides a brief insight into the different challenges faced by security systems. Section 3 describes the literature review corresponding to our model. Section 4 gives the AMI architecture. Section 5 elicits the algorithm for the localization and kNN for the secure transmission of packets, in detail. Section 6 describes the encryption and data flow process. The simulation for localizing meters in residential areas and the kNN algorithm for node authentication are illustrated in Section 7. The security strength of a data packet is analyzed in Section 8. Finally, a brief conclusion is given in Section 9.

2. Challenges Faced by Advanced Metering Infrastructure Meters

Like any other systems, the AMI needs to fulfill four primary requirements of security viz. confidentiality, integrity, availability and accountability (non-repudiation) [19]. Confidentiality implies that data must be accessible only to the authorized users, and all unauthorized attempts must be denied. Since fine-grained consumption data of a smart meter convey consumers' lifestyle patterns, habits and energy usage, they must be concealed. Integrity requires reflecting authentic data correctly without any modification, addition or deletion. Since the hackers, as well as the consumers might want to alter the consumption data, integrity is a vital issue in the AMI data.

Availability means that the data must be available on demand at all times for authorized users of the system. Availability follows the concept of authorization, which in turn implies that the data in the system can be used only by users who are allowed to have access. This involves the concept of access controls, wherein not all users have the same degree of freedom and control over the dataset of the system. There are restrictions to using specific aspects of data, which ensures that not everything can be accessed by everyone. Availability takes this one step further by ensuring that the accessible data must not be denied to the user by the system at any point of time. Since the adversaries might want to jam the network, thereby preventing the system from making the data available, or much worse, incapacitating the system's feature to make the data available, the AMI must comply with this requirement. Accountability (non-repudiation) means that an entity doing a specific job must not deny it from doing that. In AMI, accountability ensures timely responses to the command and control, the integrity of the billing profile, etc.

End-user privacy is another challenge of AMI data security. Smart meters are essentially small banks of customer usage snapshots; when aggregated together over a period of time, they provide an immense wealth of information that if put to the wrong use might compromise the privacy of customers. Smart meters provide data that is usually granular or fine-grained and the high-frequency type of energy measurements whose illegitimate analysis results in or may result in the invasion of privacy, near real-time surveillance and behavioral profiling. When the analysis is coupled with an even more threatening hazard, such as manipulation of the analyzed data, the attackers get to open a window to observe how many people are at home and at what times, to determine people's sleeping and eating routines, appliance usage patterns and home vacancy patterns.

Taking it one step further, hackers become capable of wirelessly updating smart meter firmware and remotely disconnect a user or a large section of users. Attackers, armed with different consumer patterns, can stage efficient electricity thefts and frauds, running up bogus charges or cause an electrical appliance to malfunction, shutdown or surge, causing physical damage to life and property.

The AMI meters are inherently susceptible to buffer overflows and the seven state machine flaws, as illustrated in [20]. Attacks that exploit its hardware vulnerabilities, such as bus sniffing, clock speed and power glitches, are also prevalent [21]. An attacker can create abnormal operating conditions by varying the time and voltage levels crucial to the meter performance, consequently gaining access to previously inaccessible parts of the system. Exposing the chip's surface to lasers, micro-probing to inject false signals, capturing or intercepting data and manipulating registers are some of the more advanced methods employed to compromise the meter's integrity in a physical, as well as cyber fashion. In recent times, differential power analysis and other similar techniques have been successfully used to extract the secret keys and circumvent the embedded IC security mechanisms altogether, as shown in [22]. Therefore, all of these issues need to be addressed in the data security of AMI. In the scope of this study, we look forward to providing a security scheme that will endure all of the challenges.

3. Literature Review

Data security is the prime challenge faced by AMI, as explained in Section 2. In order to be accountable for securing data, many models/techniques have been proposed in the literature. Some of these models are verifiable computation models, anonymization, perturbation models, data obfuscate techniques, trusted aggregators, etc. [23].

The work in [24] proposed a tested and established method to provide security to the meter data integrity by either using digital signatures that a TTP might sign with a time stamp to enforce not just integrity, but also authentication. Additionally, data hashing using secure hash algorithm-256 (SHA-256) before performing the signature provides an added layer of security, so that the third party need not store the keys in plain text, but can just store the hash values of the keys for each smart meter. Though the computation of hashes might add a minimal overhead, the meters need to compute hashes every time before transmitting the data packets. In [11], a 128-bit advanced encryption standard-galois counter mode (AES-GCM) cryptographic system-based secure integrated circuit (IC) was proposed with the in-depth comparison of performance between the hardware and software-based crypto-engines. An integrated authentication and confidentiality (IAC) is proposed in [25] to mutually authenticate the back office with the smart meters in order to obtain the correct cryptographic keys to be used for performing secure data communications. However, this allows the back offices to exploit the encryption and message authentication engines that are custom-made for the particular security necessities and system-imposed restrictions.

Anonymization of meter data was done in [26], describing the mechanism of how a third party escrow authenticates anonymous meter readings without being aware of a particular smart meter identification (ID) or location or its corresponding customer. In [27], physical layer-based security has been proposed, where noise is added from a known distribution before transmitting and is reconstructed as an approximation of the original data. However, there exists a trade-off between the level of privacy achieved and the loss of information.

The method shown in [28] depicts a privacy-aware architecture for demand response analysis that does not require the centrally-collected AMI data, thus reducing the privacy issues associated with behavioral profiling and other threats and vulnerabilities discussed earlier. However, neither of these works take into account the efficiency and scalability issues pertaining to the authentication and privacy protocols presented between the AMI meters and the back office.

Although [29] proposes a homomorphic encryption to solve the pressing security concerns, when a large network is considered, the data retrieval at the control center becomes tedious and cumbersome. Game theory is another fruitful modeling algorithmic paradigm that has been exploited by [12,30] wherein they put it to use viewing an attacker and a defender scenario accompanied by an attack level and a severity level.

In the case of trusted aggregators, many studies can be found that include a TTP, key management, node-to-node authentication, etc. A node-to-node encryption by different secret keys has been proposed in [31]. However, again, for a huge network scenario, the packet overhead increases, since authentication needs to happen at every single node in the message packet's path. In [32,33], a public key management infrastructure (PKI) is proposed both to distribute the key and to manage the network. As with any PKI method, the distribution of the public keys becomes a point of vulnerability that entails various complications. Longitude, altitude and time form an encryption key in [13–15] for the purpose of data encryption. This technique ensures that the data cannot be decrypted outside a particular facility, such as a local utility company control center, different government agencies or corporations. For determining longitude and altitude, GPS has been proposed. However, GPS does not work well in some places, such as multi-storied buildings, hilly places and forests. Moreover, in today's signal processing system, any coordinate point can be generated at any place.

Here, in our paper, we propose a data encryption technique associated with latitude, longitude and a random key index. The longitude and altitude are derived by the RSS-based technique. Since the RSS-based technique involves error in determining the coordinates of meters, AMI will create a private coordinate system with constant error (due to constant position of the meter). These private coordinates will not expose the exact coordinates of meters/consumers. Moreover, if we consider each cluster of meters served by a TTP, our technique becomes scalable. Additionally, we use the kNN algorithm to ensure that the message is received from neighbor meters not from any unauthorized entity. This node authentication will also help to intercept malicious packets.

4. Architecture of Advanced Metering Infrastructure

Millions of smart meters are engineered to communicate with the local utility SP/control center using the AMI network. The bidirectional networks can be a mesh, or a hierarchical, or a hybrid. Periodic collection, storing and transmission of enormous volumes of data packets via the regional

APs form the primary workhorses of AMI. The data packets are transmitted through a gateway or data concentrator, as shown in Figure 1, which then relays the packets eventually to the control center. Each component of the AMI performs its own application. From a broader perspective, the AMI encompasses everything from home appliances to the control center, forming a comprehensive network. The following are different sub-components of the AMI architecture.

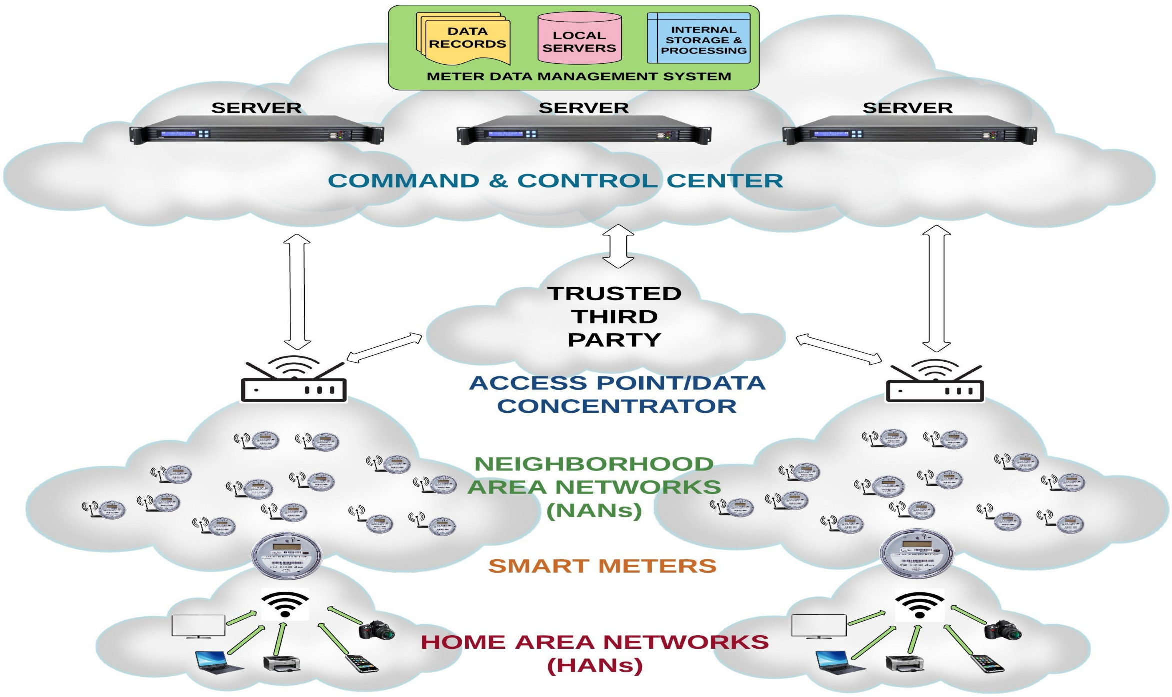

DATA RECORDS
LOCAL SERVERS
INTERNAL STORAGE & PROCESSING
METER DATA MANAGEMENT SYSTEM
SERVER
SERVER
SERVER
COMMAND & CONTROL CENTER
TRUSTED THIRD PARTY
ACCESS POINT/DATA CONCENTRATOR
NEIGHBORHOOD AREA NETWORKS (NANs)
SMART METERS
HOME AREA NETWORKS (HANs)

Figure 1. The proposed advanced metering infrastructure (AMI) architecture consists of the home area network (HAN), the neighborhood area network (NAN), the trusted third party (TTP), the data concentrator and the control center.

- Home appliance: These are the machines that are employed by consumers for performing every-day chores and activities, and those that consume electric power come under this category. Some examples are washing machines, driers, microwave ovens and air conditioners. The energy consumed by these machines is calculated per unit of the smart grid system. The unit consumption data are then relayed to the smart meter, which measures, calibrates and collects the information reported by these appliances.
- Smart meter: They form the backbone of the AMI, being responsible for collecting the consumption unit data from consumers before dispatching the data to the SP. The meters are designed to use the channel in periodic intervals of time by sending short bursts of information. The small network formed between the home appliances and the meter corresponding to that household is termed the home area network (HAN). The smart meters measure, collect and store data before sending the same in the form of encrypted packets.
- Neighborhood area network (NAN): Beyond HAN, there is a broader network that is made up of various smart meters within a locality and their corresponding APs. These meters communicate among themselves through a mesh/hierarchical/hybrid connected wired (PLC)/wireless (WiFi, ZigBee, GPRS, etc.) network termed the NAN.
- TTP: The entity administrating the security scheme is known as TTP. In our scheme, TTP authenticates the meter and conveys the random key index to the control center. Additionally, TTP updates the codebook containing encryption keys mapped with meters' coordinates periodically.
- Control center/back office/command and control center (CCC): The CCC is connected to the NAN by a wired or wireless connection, such as a fiber optic or a cellular network. A bill is

issued for the consumer based on the data received by the data center over the period of one month. Each day, the utility receives the data in 15–60-min intervals. This data are also used in the optimization of the electric power generation and distribution. Additionally, it also helps in the control and monitoring of the load from a remote location.

Since smart meters have limited memory and computational capability, AMI requires a light, but robust security scheme. Besides that, AMI is a huge network consisting of thousands of meters. In the case of data encryption by key management using private/public keys, we need to generate a huge number of unique keys, which is very complicated. Moreover, node-to-node identification by private/public keys has the vulnerability of exposing keys. Therefore, we propose a modified key management, which uses a key book mapped with the coordinates of the meters. This will provide the flexibility needed to use the same key book for different locations.

5. Localization Algorithm and Node Authentication Technique

In our scheme, we use the RSS-based technique for the localization of meters and the kNN algorithm for meter (node) authentication. In the rest of this section, the two algorithms are explained in detail.

5.1. Localization of the Smart Meter by the Received Signal Strength-Based Method

Let us assume, there are n partially-dispersed known position meters at positions (x_i, y_i) where $i = 1, 2, ..., n$, and a new meter is at position (x, y) as shown in Figure 2. If the RSS of new meter at (x_i, y_i) is $\mho_i$, then it follows the model [11]:

$$\mho_i = c - 10\gamma \log_{10}(d_i) + w_i \tag{1}$$

where c denotes a constant dependent on the transmitted signal power, frequency, etc.

$\gamma > 0$ is the path loss constant. The generic value of γ is 4–6 from which a value of 2.93 has been used here considering the residential area [34].

d_i defined as the Euclidean distance among the new meter and other meters is given by:

$$d_i = \sqrt{(x - x_i)^2 + (y - y_i)^2} \tag{2}$$

w_i is the zero mean random Gaussian noise with standard deviation σ_i. The typical value of σ is 6–12 dB.

We define the θ and $\mho$ as $\theta = [x, y, z]^T$ and $\mho = [\mho_1, \mho_2, ...\mho_n]^T$, where z is the reference transmission power.

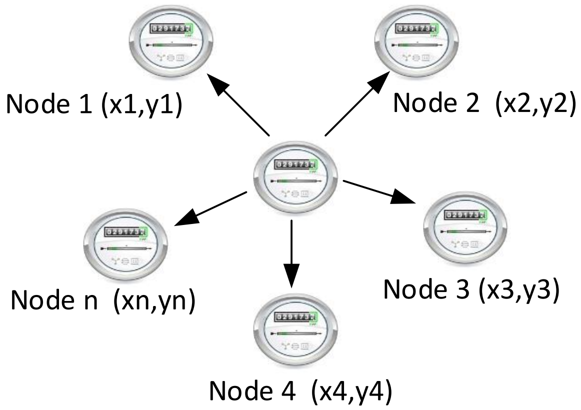


Figure 2. Localization of the new meter (unknown positioned meter).

The likelihood function of θ given an RSS measurement $\mho_i$, $f(\theta/\mho)$ can be written as:

$$f(\theta/\mho) = c_1 \exp(-\sum_{i=1}^{n} \frac{\mho_i - c + 10\gamma \log_{10}(d_i)}{2\sigma_i^2}) \tag{3}$$

where c_1 is a constant.

The maximum likelihood estimation of θ denoted by $\hat{\theta}$ is:

$$\begin{aligned} \hat{\theta} &= \arg\max_{\theta} f(\theta/\mho) \\ &= \arg\min\{\sum_{l=1}^{n} \frac{\mho_i - c + 10\gamma \log(d_i)}{2{\sigma_i}^2}\} \end{aligned} \tag{4}$$

The ML estimator returns the estimated position $(x_{\mathrm{r}}, y_{\mathrm{r}})$ and reference power (p_{r}) of the smart meter, i.e.:

$$(x_{\mathrm{r}}, y_{\mathrm{r}}, p_{\mathrm{r}}) = \{\hat{\theta}(1), \hat{\theta}(2), \hat{\theta}(3)\} \tag{5}$$

For no prior knowledge of the position and the reference power, the above optimization problem (5) must be solved in three dimensions, but most of the practical problems contain some prior information about the vector parameter. In this case, the Bayesian philosophy can be employed to find an estimation of the unknown vector parameter.

Since ML estimation in our localization problem is asymptotically optimal, we used the Deterministic Particle Swarm Optimization (D-PSO) [35,36] technique for finding the swarm (global) minimum. Since our problem is on three-dimensional space and this technique does not utilize the gradient of the problem to be optimized, D-PSO provides an effective solution for our problem. DPSO has some additional advantages over conventional PSO, such as: (1) a consistence result is achieved with a small number particles; (2) only one parameter (i.e., inertia weight) needs to be tuned; (3) the optimization structure is simple; and (4) the algorithm is easily implementable in frequently changeable environmental conditions [35]. In general, PSO/DPSO is initialized with a group (population) of random solutions of particles. Each particle has two states: its current position x and velocity v. Each particle has the ability to memorize its own best position *pbest*, and the best position that the swarm experienced, *gbest*. At each iteration, the position and velocity are updated according to Equations (1)–(4) of [36]). The particles will fly in the swarm in an N-dimensional space to have the best coordinate positions backed by the best current (personal) value of the evaluation function.

The position determination by the above RSS-based technique will create a local coordinate system for HAN, different from the geographical coordinate system.

5.2. Meter Authentication by the k-Nearest Neighbors Algorithm

The kNN algorithm was proposed by Cover and Hart, where k denotes the number of nearest neighbors that are helpful to predict the class of the test sample [18].

Let us consider a set of meters $M = \{m_i\}, i = \{1, 2, ...N\}$ with attribute $a_l^t(x_{m_i})$, $l = 1, 2, ...L, m_i \in M$ at instance t. We define $C_{m_i}^t$ and $c_{m_i}^t$ as the class variable and class value, respectively. The standard Euclidean distance between instance t and $t+1$ is:

$$d(x_{m_i}^t, x_{m_i}^{t+1}) = \sqrt{\sum_{l=1}^{L} (a_l^t(x_{m_i}) - a_l^{t+1}(x_{m_i}))^2} \tag{6}$$

When the value of attributes are nominal, the variation of the standard Euclidean distance can be written as:

$$\begin{aligned} d(x^t_{m_i}, x^{t+1}_{m_i}) &= \sum_{l=1}^{L} \delta(a^t_l(x_{m_i}) - a^{t+1}_l(x_{m_i})) \\ \text{subject to:} \quad \delta(a^t_l(x_{m_i}), a^{t+1}_l(x_{m_i})) &= \begin{cases} 0, & \text{if } \; a^t_l(x_{m_i}) = a^{t+1}_l(x_{m_i}) \\ 1, & \text{otherwise} \end{cases} \end{aligned} \tag{7}$$

The most common class value of $x^t_{m_i}$ at instant t:

$$c^t_{m_i}(x_{m_i}) = \arg\max \sum_{c_{m_i} \in C_{m_i}} \delta\left(c^t_{m_i}, c(x^{t+1}_{m_i})\right) \tag{8}$$

where $x^{t+1}_{m_i}$ is the kNN neighbors of $x^t_{m_i}$ and:

$$\delta(c^t_{m_i}, c(x^{t+1}_{m_i})) = \begin{cases} 1, & \text{if } \; c^t_{m_i} = c(x^{t+1}_{m_i}) \\ 0, & \text{otherwise} \end{cases} \tag{9}$$

For each meter m_i, the parameters distance between source meter m_j and meter m_i (d), packet size (s) and data transmission frequency (f) are used as data for the classification. We use Euclidean distance for the measure of similarity among the classes. To describe the steps of the algorithm, we use the terminology "tuple" as the dataset consists of d, s, f. The algorithm is described in pseudocode in Algorithm 1.

Algorithm 1 kNN algorithm.

```
For each data tuple $\psi_i \in \Psi, i = \{1,2,3,......,n\}$, 2 dimensional vector space $\psi_i =< W_{i1}, W_{i2}, ......, W_{in} >$
Training:
Set "ungrouped" tag to all data tuples
Calculate Euclidean distance $Ed_i = \sum\sum(x_l - x_k)$ among elements of $\psi_i$ where $x_l, x_k \in \psi_i$
Build a representative model $\Omega =< cls(\psi_i), sim(\psi_i), num(\psi_i) >$ where $cls$ represents class level, $sim$ represents similarity and $num$ of elements
Classification:
For new data tuple $\psi_k$, calculate representative parameter
    $\Omega_k =< cls(\psi_k), sim(\psi_k), num(\psi_k) >$
if representative parameter $\Omega_k$ belong to Model $\Omega$ then
    New data belongs to same group
else
    New data belongs to different group
end if
End
```

kNN is a lazy learning approach. However, we use the inductive learning classification model [17] for learning to improve the efficiency of the kNN technique. The training takes place only at the beginning and once. An authenticate set of data is used for training. After that, the induction model is used for classification.

In the real world, every meter will be accompanied by a few neighboring meters, except in multi-storied building. Besides that, we used only three parameters, distance, packet size and data transmission frequency. Therefore, the classification will be fast. More advanced approaches, such as neural network, bioinspired algorithms, ant colony optimization algorithms, genetic algorithms, etc., have better accuracy. However, these approaches are complex and need more memory/computation ability to implement. On the other hand, kNN is simple and easy to implement.

The use of the kNN algorithm along with the RSS-based technique will allow data/packets to be received from authenticated neighbor meters. Therefore, this will ensure node-to-node authentication, intercepting malicious packets. Therefore, kNN added an extra layer of security, whereas the main scurrility is provided by encryption by keys.

6. Encryption Process

A detailed description of the entire encryption and data flow process in the AMI is provided in this section. As shown in Figure 3, there is an involvement of a TTP, which will perform the authentication of the different smart meters using their node IDs. Once the TTP authenticates a particular smart meter (m_i), it sends the key index to the CCC. At the same time, the encrypted data are sent to the CCC via intermediary nodes (other smart meters). Finally, when the encrypted packet reaches the CCC, the destination will decrypt the message using the key associated with the random key index and the meter's (m_i) coordinates: latitude and longitude. Before we proceed to the steps of encryption and data flow, the assumptions are outlined below.

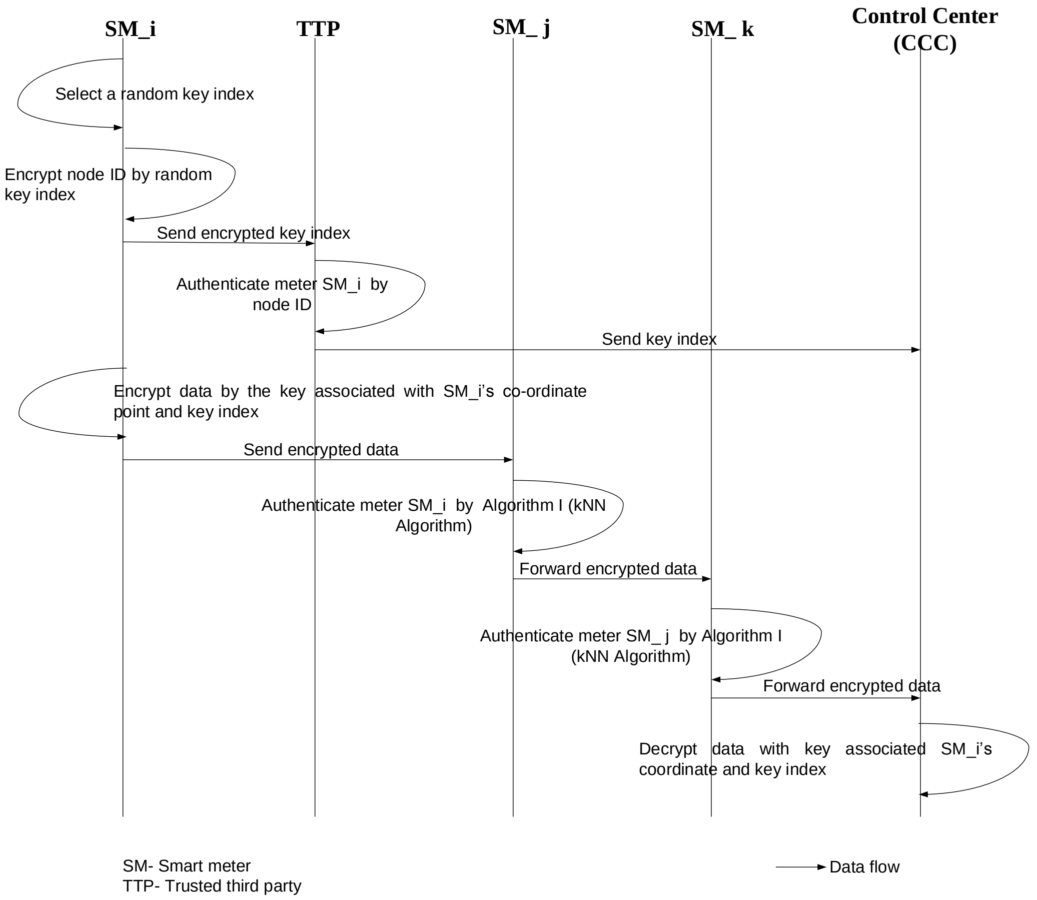
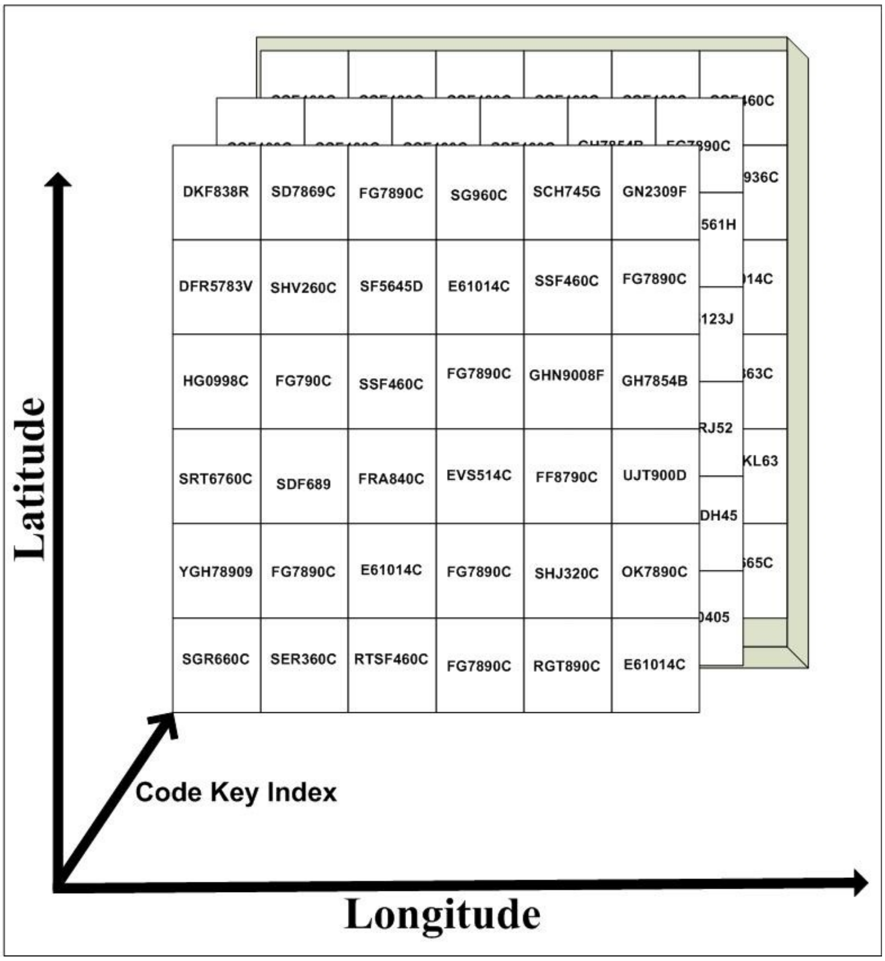

SM_i
TTP
SM_ j
SM_ k
Control Center (CCC)
Select a random key index
Encrypt node ID by random key index
Send encrypted key index
Authenticate meter SM_i by node ID
Send key index
Encrypt data by the key associated with SM_i's co-ordinate point and key index
Send encrypted data
Authenticate meter SM_i by Algorithm I (kNN Algorithm)
Forward encrypted data
Authenticate meter SM_ j by Algorithm I (kNN Algorithm)
Forward encrypted data
Decrypt data with key associated SM_i's coordinate and key index
SM- Smart meter
TTP- Trusted third party
Data flow

Figure 3. Encryption and data transmission process among various components of AMI. CCC: command and control center.

Assumptions:

- The meters are divided into multiple clusters of a size of 100–300 m. Each cluster is served by a TTP and an AP.

- The meter has a limited memory and computational capability.
- The control center has sufficient computational capability.
- The communication protocol is ideal with the available bandwidth.
- Every meter holds records of the location of its neighboring meters.
- Every meter transmits data at a constant power.
- There is a codebook that has an encryption key associated with each coordinate point of the geolocation, as shown Figure 4.
- The TTP updates the codebook associated with the geolocation/coordinate point periodically.



DKF838R	SD7869C	FG7890C	SG960C	SCH745G	GN2309F
DFR5783V	SHV260C	SF5645D	E61014C	SSF460C	FG7890C
HG0998C	FG790C	SSF460C	FG7890C	GHN9008F	GH7854B
SRT6760C	SDF689	FRA840C	EVS514C	FF8790C	UJT900D
YGH78909	FG7890C	E61014C	FG7890C	SHJ320C	OK7890C
SGR660C	SER360C	RTSF460C	FG7890C	RGT890C	E61014C

Figure 4. Mapping of the encryption key based on the coordinate point.

The following are the different steps during the message encoding, transmission and decoding processes:

Initialization: The meter, which is ready to send information, identified as the source meter m_i, performs an initiation process for every session (typically once in every 15/30 min). This process involves the selection of a random key index. This key index is then encrypted by node ID and sent to the TTP. The TTP has a codebook that also contains information of all smart meters. Hence, it uses the codebook to identify the node by its ID, decrypt the key index and send it to the CCC.

Data encryption: In the next step, the source meter encrypts the consumption data, which are plain text, with an encryption key associated with its own coordinate points (latitude and longitude) and the key index.

Encryption:

$$k_i \oplus m_i \longrightarrow C_i \tag{10}$$

Data forwarding and authentication: This encrypted message is forwarded to its peers, since relaying is the only way the packets can reach the CCC eventually, considering that each meter has very low transmission power. Now, any neighboring meter (m_j) receiving the encrypted packet from a source meter (m_i) determines its authentication by Algorithm 2. Once authenticated, the corresponding packets are forwarded to the next neighboring meters.

Decryption: The CCC receives the encrypted data and decrypts the same with the help of the key associated with the source meter's location and the key index it received from the TTP.

Decryption:

$$C_i \xrightarrow{k_i} m_i \tag{11}$$

Algorithm 2 Transmitting algorithm.

```
Derive the position of neighbor meter $m_i$ by the RSS-based method
Calculate the distance between source meter $m_j$ and neighboring meter $m_i$, $R_{i,j}$, where $m_i, m_j \in M$
Get $s$ and $f$ from packet header
Run the kNN algorithm (Algorithm I)
if New data belongs to same group then
    Forward data to the next meter $m_k$, where $m_k \in M$
else
    Discard the data and report to CCC
end if
End
```

7. Simulation Results

In this section, we evaluate the performance of the RSS-based technique and the kNN algorithm in our architecture. Rectangular, hexagonal and octagonal shapes are considered for the position map of the meter location. These shapes comprise what is known as the area of interest (AOI) with an approximate dimension of 10 m × 10 m, as depicted in Figure 5. Therefore, the approximate distance between two meters is 10 m [37]. Each edge represents a known position meter (node), and the emitter (i.e., unknown positioned meter/node) is the center of the AOI. The estimation of the position of the unknown positioned meter through Equation (1) is the optimization problem. In our simulation, we use the deterministic-particle swarm optimization (D-PSO) for that optimization. Due to absence of the random values, the particles follow a deterministic behavior, and the solution is consistent with the small number of particles in each independent iteration.

In Figure 6, we observe that with an increase of the number of meters, the mean square error from the exact position will decrease. This means that for more neighbor meters, the error in the localization of the meter is decreased. At the same time, with an increase of the path loss exponent/constant, the mean square error will decrease as illustrated in Figure 7. The path loss constants for free space, urban area and suburban area are 2, 2.7–3.5 and 3–5, respectively [34]. Therefore, the results of Figure 7 demonstrate the prediction of localization error in different areas, such as urban, suburban and free space. In the path loss model of radio signals, random noise is added, which varies by the standard deviation. In Figure 8, the surface diagram is drawn as a function of the variance of noise and the number of nodes. As the variance of the noise increases, the error also increases correspondingly. Therefore, for the presence of a building, wall, trees, etc., the error for determining the location of meters will be high. However, free space between meters gives the best performance of localization.

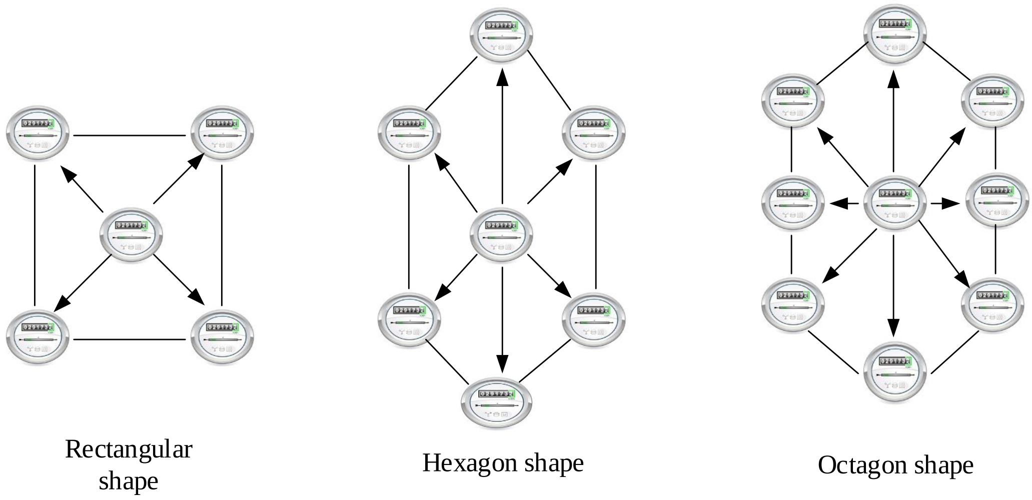


Figure 5. Different shapes of the area of interest (AOI).

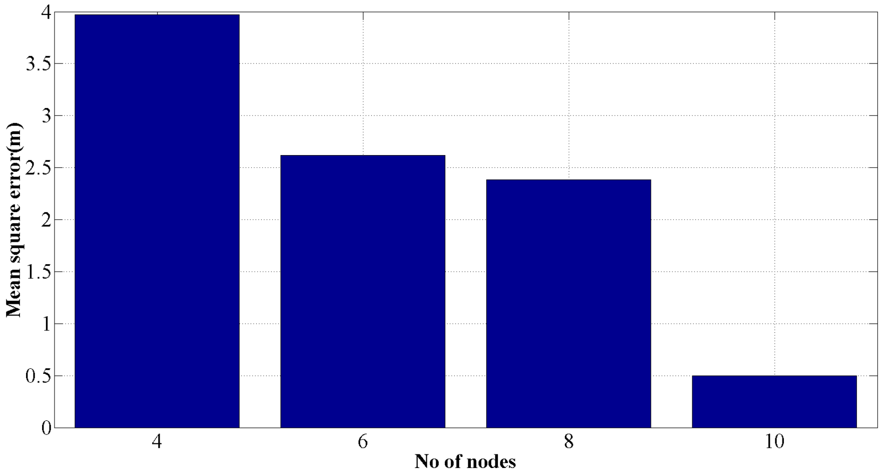


Figure 6. Number of nodes vs. mean square error (m).

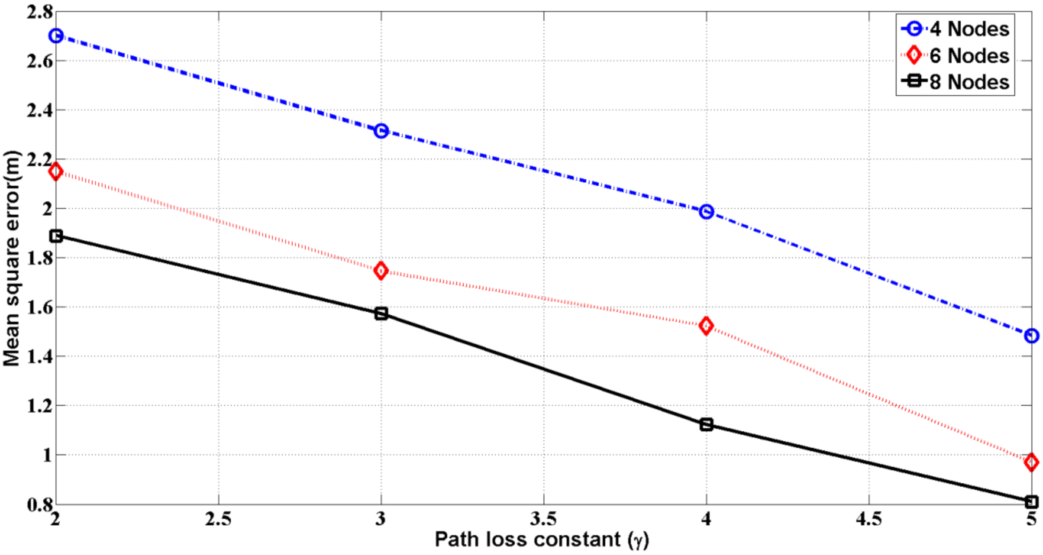


Figure 7. Path loss vs. mean square error (m).

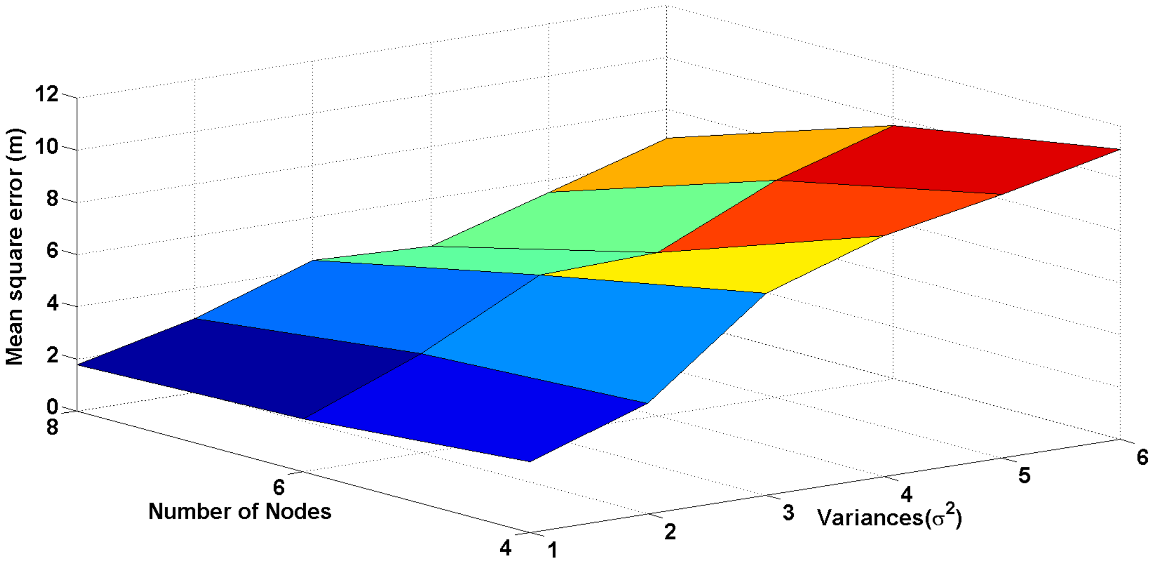


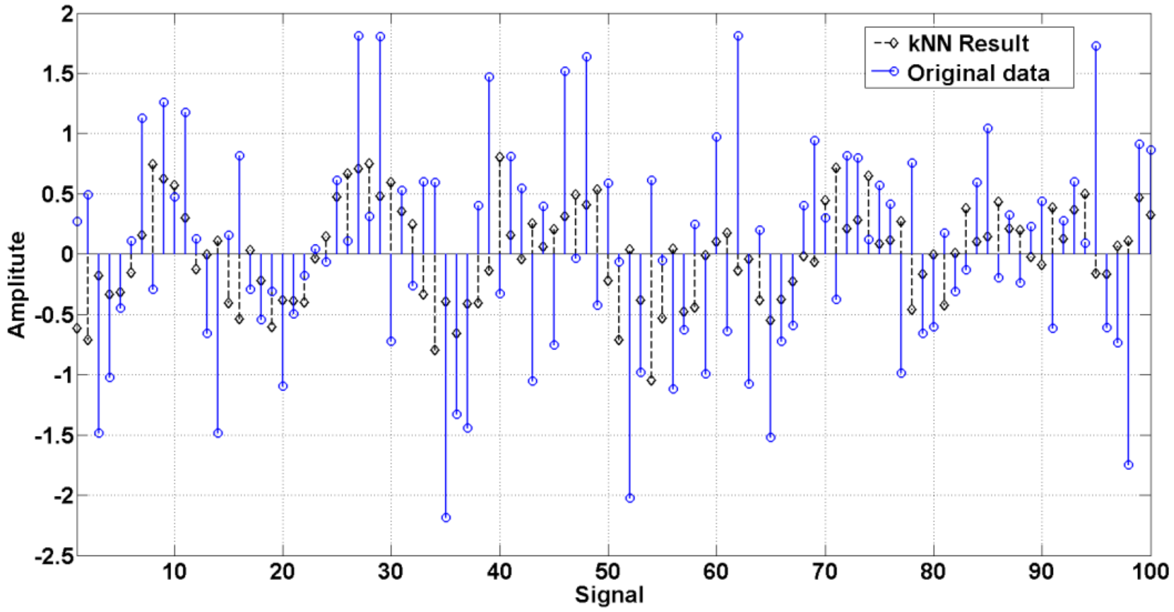
Figure 8. Variance vs. mean square error (m).

Since the meters are mounted on a stationary wall/pole and the environment surrounding meters is stable, the calculated error in the localization method of an unknown/new positioned meter at neighbor meters by the RSS-based method will be almost constant. Moreover, the meters are connected by a mesh network, so the failure of a meter will cause data transport by alternative paths. Another advantage of the mesh network is the ad hoc nature of the connection. If any meter discontinues sending data to the CCC for a specific period, that meter will be the subject of inspection/maintenance.

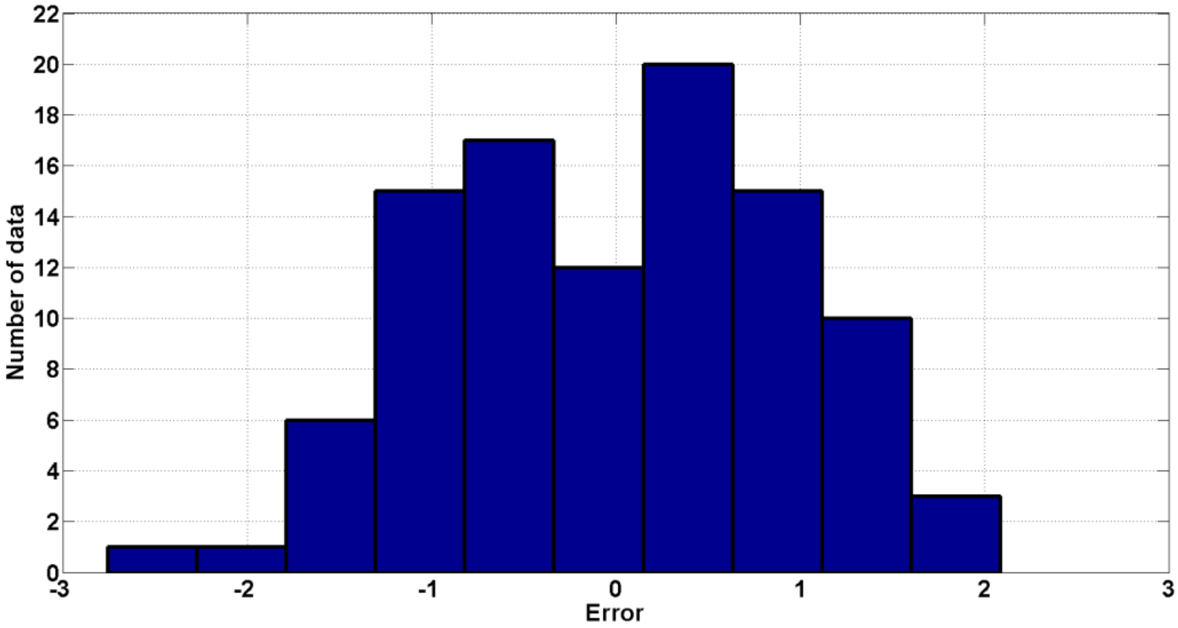
Now, we evaluate the performance of the kNN algorithm in the AMI environment. Let the number of neighbor meters $k = 4$; Figure 9a presents a dataset of a size of 100 and the predicted result of the dataset derived by the kNN algorithm. Compared with the original dataset and the predicted result, Figure 9b shows the error histogram between these two datasets. It is noted that the error is well distributed on both sides of the histogram. This error distribution is used to authenticate the test data/real data from meters. If we train a meter with an original dataset, we can find a prediction model with an error distribution. In the next step, for the test data, we compare whether this test belongs to the same group of training data. Similarly, Figure 10a presents a dataset of a size of 200 and the predicted result derived by the kNN algorithm. Figure 10b shows the statistics of the error distribution between the original and predicted dataset. It is noted that the error is better distributed for 200 data than that of 100 data.

The performance of kNN algorithm for different data sizes and different numbers of neighbors (k) is illustrated in Figure 11. We found that with the increment of the size of the data and the decrement of k, the mean square error between the original data and the predicted data increases quite precisely. This means that for a greater number of data classes and a smaller dataset, the kNN algorithm performs well. In our study, we proposed only three data classes (i.e., distance, frequency and packet size), which will provide results with reasonable error.

Smart meters send the consumed data periodically at a specific interval defined by the utility company, and the data packet size is constant. For these reasons, a meter can authenticate the source meter by the kNN algorithm using the data of sending frequency, packet size and distance between two meters.

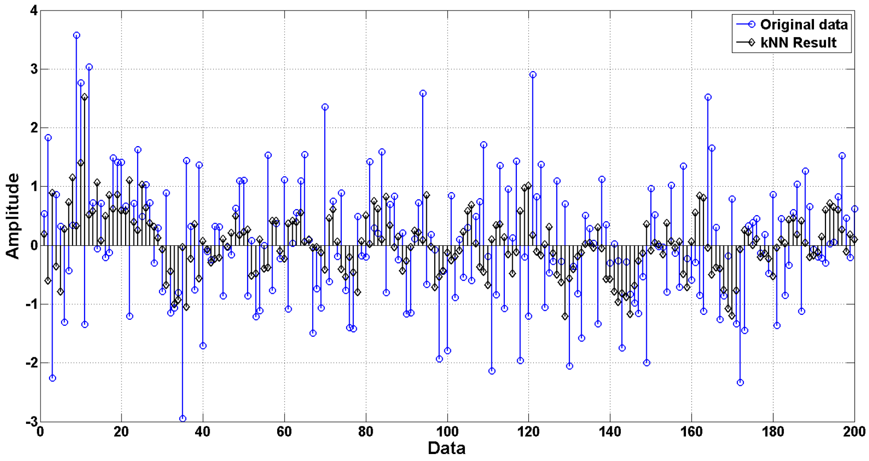


(**a**)

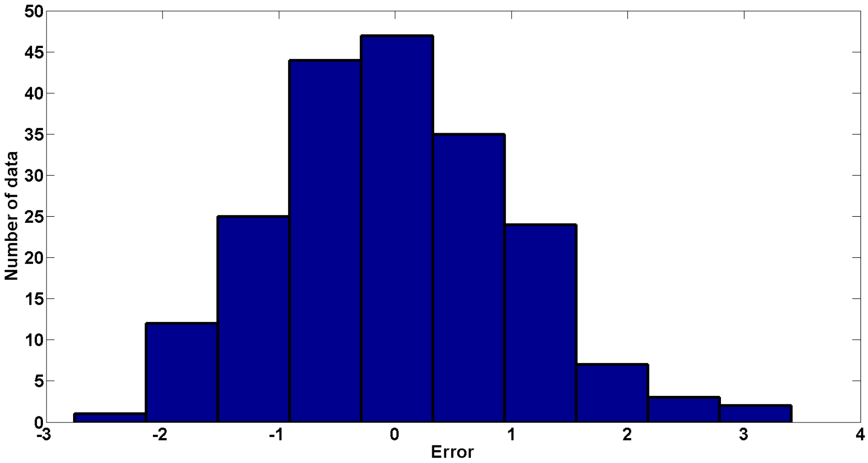


(**b**)

Figure 9. Performance of the kNN algorithm for a dataset of a size of 100. (**a**) Comparison of the data values and their prediction by the kNN algorithm; (**b**) histogram of the error distribution.



(a)



(b)

Figure 10. Performance of the kNN algorithm for a dataset of a size of 200. (**a**) Comparison of the data values and their prediction by the kNN algorithm; (**b**) histogram of the error distribution.

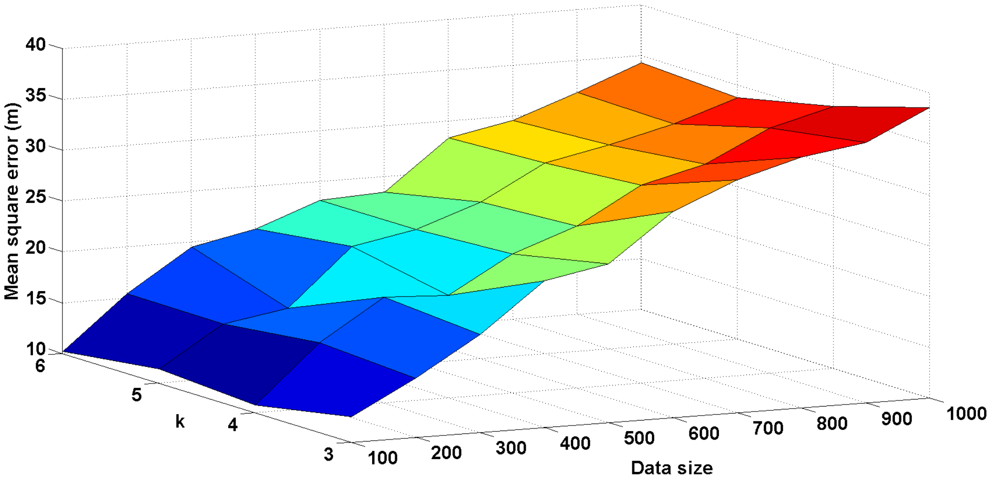


Figure 11. The relation of the error between the number of neighbors and the data size.

8. Security Strength Analysis

The security strength of a data packet can be measured by entropy. The value of entropy reflects the uncertainty of a random variable. The more certainty about a value there is, the smaller the entropy value.

The entropy for a sequence S [38]:

$$H(s) = \sum_{S} P(S = x) \log_2 P(S = x)$$

where $P(S = x)$ is the probability of taking the S value over x.

Let us consider that a smart meter sends a data packet of 128 bits encrypted by a 128-bit symmetric key to the control center. For an 8-bit random key index, the security strength of the random sequence is 2^8. On the other hand, for a 128-bit symmetric key algorithm, the security strength is 2^{128}. So, for an 8-bit random key index and a 128-bit symmetric key, the security strength of the packet is ($2^8 + 2^{128}$).

Therefore, if a hacker wants to decrypt a data packet of 128 bits, he or she needs ($2^8 + 2^{128}$) tries to decrypt the message unless he or she is lucky. This is impractical.

9. Conclusions

In this article, we propose a two-level security scheme consisting of data packet encryption and node authentication. For packet encryption, we use an encryption key corresponding to the meter's location and a random key index. The codebook relating the encryption code to its geographical location could be updated periodically in a secured way by the TTP. Since the same code book is used for all smart meters, this invalidates the need for using different keys for different meters/nodes each time. Furthermore, the inclusion of a random key index increases the uncertainty in packet encryption. In the case of node-to-node authentication, we use the kNN classifier, which is robust, and simple to implement due to the few data.

The positioning map of meters in HAN is a local map different from the geographical coordinate system. Therefore, exposing the coordinates will not reveal the exact location of a consumer's house. Besides, the position of smart meters determined by the RSS-based method is nearly constant due to the constant power transmission and stable positions of meters. This ensures the stability and security of the smart grid system without giving the opportunity to the potential hackers to manipulate the data.

Our scheme segregates the AMI network into small clusters of meters where each cluster is served by a TTP and AP. For this reason, our scheme is applicable in any AMI network that uses a wireless network.

An automated, secure and reliable metering paradigm is an essential component for the designing and building of smart cities with high standards of living and providing vibrant socioeconomic climates to the citizens. Since fine-grained data of meters contain important information about the consumers, revealing the data compromises citizen's privacy. Our paper proposes a technique to resolve a fraction of smart grid security threats and making the communication more secure and reliable.

Acknowledgments: This research was supported in part by the U.S. National Science Foundation under the grant RIPSType 2-1441223 and the Presidential fellowship under Florida International University.

Author Contributions: Imtiaz Parvez and Arif Sarwat conceived and designed the project. Imtiaz Parvez, Arif Sarwat and Longfei Wei completed the simulation and validated the results. Finally, Imtiaz Parvez and Aditya Sundararajan wrote the paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Akkaya, K.; Rabieh, K.; Mahmoud, M.; Tonyali, S. Customized certificate revocation lists for IEEE 802.11s-based smart grid AMI networks. *IEEE Trans. Smart Grid* **2015**, *6*, 2366–2374.

2. Parvez, I.; Sriyananda, M.G.S.; Güvenc, I; Bennis, M; Sarwat, A.I. CBRS Spectrum Sharing between LTE-U and WiFi: A Multi-Armed Bandit Approach. 2016. Available online: downloads.hindawi.com/journals/misy/aip/5909801.pdf (accessed on 29 August 2016).
3. Yang, Z.; Shi, Z.; Jin, C. SACRB-MAC: A high-capacity MAC protocol for cognitive radio sensor networks in smart grid. *Sensors* **2016**, *16*, 464. doi:10.3390/s16040464.
4. Parvez, I.; Sundararajan, A.; Sarwat, A. Frequency band for HAN and NAN communication in smart grid. In Proceedings of the IEEE Symposium on Computational Intelligence Applications in Smart Grid (CIASG), Orlando, FL, USA, 9–12 December 2014; pp. 1–5.
5. Brown, J.; Khan, J. Performance analysis of an LTE TDD based smart grid communications network for uplink biased traffic. In Proceedings of the 2012 IEEE Third International Conference on Smart Grid Communications (SmartGridComm), Tainan, Taipei, 5–8 November 2012; pp. 1502–1507.
6. Parvez, I.; Islam, N.; Rupasinghe, N.; Sarwat, A.I.; Güvenc, I. LAA-based LTE and ZigBee coexistence for unlicensed-band smart grid communications. In Proceedings of SoutheastCon 2016, Norfolk, VA, USA, 30 March–3 April 2016.
7. Gharavi, H.; Chen, H.H.; Wietfeld, C. Guest editorial special section on cyber-physical systems and security for smart grid. *IEEE Trans. Smart Grid* **2015**, *6*, 2405–2408.
8. *Independent Statistics and Analysis*; U.S Energy Information Administration: Washington, DC, USA.
9. Amadeo, M.; Campolo, C.; Quevedo, J.; Corujo, D.; Molinaro, A.; Iera, A.; Aguiar, R.L.; Vasilakos, A.V. Information-centric networking for the internet of things: Challenges and opportunities. *IEEE Netw.* **2016**, *30*, 92–100.
10. Shi, X.; Li, Y.; Cao, Y.; Tan, Y. Cyber-physical electrical energy systems: Challenges and issues. *CSEE J. Power Energy Syst.* **2015**, *1*, 36–42.
11. Parvez, I.; Jamei, M.; Sundararajan, A.; Sarwat, A. RSS based loop-free compass routing protocol for data communication in advanced metering infrastructure (AMI) of Smart Grid. In Proceedings of the 2014 IEEE Symposium on Computational Intelligence Applications in Smart Grid (CIASG), Orlando, FL, USA, 9–12 December 2014; pp. 1–6.
12. Somkaew, W.; Thepphaeng, S.; Pirak, C. Data security implementation over ZigBee networks for AMI systems. In Proceedings of the 11th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), Nakhon Ratchasima, Thailand, 14–17 May 2014; pp. 1–5.
13. Scott, L.; Denning, D. A location based encryption technique and some of its applications. In Proceedings of the National Technical Meeting of The Institute of Navigation, Anaheim, CA, USA, 22–24 January 2003; Volume 4, pp. 734–740.
14. Parvez, I.; Abdul, F.; Sarwat, A.I. A Location Based Key Management System for Advanced Metering Infrastructure of Smart Grid. In Proceedings of 2016 IEEE Green Technologies Conference (GreenTech), Kansas City, MO, USA, 2016, pp. 62–67.
15. Firoozjaei, M.; Vahidi, J. Implementing geo-encryption in GSM cellular network. In Proceedings of the 9th International Conference on Communications (COMM), Bucharest, Romania, 21–23 June 2012; pp. 299–302.
16. Xiong, J.; Zhang, Q.; Sun, G.; Zhu, X.; Liu, M.; Li, Z. An information fusion fault diagnosis method based on dimensionless indicators with static discounting factor and KNN. *IEEE Sens. J.* **2016**, *16*, 2060–2069.
17. Guo, G.; Wang, H.; Bell, D.; Bi, Y.; Greer, K. KNN Model-Based Approach in Classification. In *On the Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE*; Meersman, R., Tari, Z., Schmidt, D., Eds.; Springer: Heidelberg, Germany, 2003; Volume 2888, pp. 986–996.
18. Cover, T.; Hart, P. Nearest neighbor pattern classification. *IEEE Trans. Inf. Theory* **1967**, *13*, 21–27.
19. Cleveland, F.M. Cyber security issues for advanced metering infrasttructure (AMI). In Proceedings of the 2008 IEEE Power and Energy Society General Meeting—Conversion and Delivery of Electrical Energy in the 21st Century, Pittsburgh, PA, USA, 20–24 July 2008; pp. 1–5.
20. Sustek, L. Understanding Smart Meters to Design Intelligent, Secure Systems. Available online: http://www.eetimes.com/document.asp?doc_id=1279025 (accessed on 20 August 2015).
21. *Advanced Metering Infrastructure Attack Methodology*. Guardians Inc.: Washington, DC, USA, 2009.
22. Meritt, K. Differential Power Analysis attacks on AES. 2012. Available online: https://people.rit.edu/kjm5923/DPA_attacks_on_AES.pdf (accessed on 25 December 2015).

23. Komninos, N.; Philippou, E.; Pitsillides, A. Survey in smart grid and smart home security: Issues, challenges and countermeasures. *IEEE Commun. Surv. Tutor.* **2014**, *16*, 1933–1954.
24. Nabeel, M.; Zage, J.; Kerr, S.; Bertino, E.; Athula Kulatunga, N.; Sudheera Navaratne, U.; Duren, M. Cryptographic key management for smart power grids. *CERIAS Tech. Rep.* **2012**, arXiv:1206.3880.
25. Yan, Y.; Hu, R.; Das, S.; Sharif, H.; Qian, Y. An efficient security protocol for advanced metering infrastructure in smart grid. *IEEE Netw.* **2013**, *27*, 64–71.
26. Efthymiou, C.; Kalogridis, G. Smart grid privacy via anonymization of smart metering data. In Proceedings of the First IEEE International Conference on Smart Grid Communications (SmartGridComm), Gaithersburg, MD, USA, 4–6 October 2010; pp. 238–243.
27. Lee, E.K.; Gerla, M.; Oh, S.Y. Physical layer security in wireless smart grid. *IEEE Commun. Mag.* **2012**, *50*, 46–52.
28. Wicker, S.; Thomas, R. A privacy-aware architecture for demand response systems. In Proceedings of the 44th Hawaii International Conference on System Sciences (HICSS), Kauai, HI, USA, 4–7 January 2011.
29. Niu, Y.; Tan, X.; Chen, S.; Wang, H.; Yu, K.; Bu, Z. A security privacy protection scheme for data collection of smart meters based on homomorphic encryption. In Proceedings of the 2013 IEEE EUROCON, Zagreb, Croatia, 1–4 July 2013; pp. 1401–1405.
30. Wei, L.; Moghadasi, A.H.; Sundararajan, A.; Sarwat, A.I. Defending mechanisms for protecting power systems against intelligent attacks. In Proceedings of the 10th System of Systems Engineering Conference (SoSE), San Antonio, TX, USA, 17–20 May 2015; pp. 12–17.
31. Yan, Y.; Qian, Y.; Sharif, H. A secure and reliable in-network collaborative communication scheme for advanced metering infrastructure in smart grid. In Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC), Cancun, Quintana Roo, Mexico, 28–31 March 2011; pp. 909–914.
32. Yu, K.; Arifuzzaman, M.; Wen, Z.; Zhang, D.; Sato, T. A Key Management Scheme for Secure Communications of Information Centric Advanced Metering Infrastructure in Smart Grid. *IEEE Trans. Instrum. Meas.* **2015**, *64*, 2072–2085.
33. Bhatia, R.; Bodade, V. Defining the framework for wireless-AMI security in smart grid. In Proceedings of the 2014 International Conference on Green Computing Communication and Electrical Engineering (ICGCCEE), Coimbatore, India, 6–8 March 2014; pp. 1–5.
34. Durgin, G.; Rappaport, T.S.; Xu, H. Measurements and models for radio path loss and penetration loss in and around homes and trees at 5.85 GHz. *IEEE Trans. Commun.* **1998**, *46*, 1484–1496.
35. Ishaque, K.; Salam, Z. A deterministic particle swarm optimization maximum power point tracker for photovoltaic system under partial shading condition. *IEEE Trans. Ind. Electron.* **2013**, *60*, 3195–3206.
36. Xu, B.; Ren, Y.; Zhu, P.; Lu, M. A PSO-based approach for multi-cell multi-parameter estimation. In Proceedings of the 2014 International Conference on Control, Automation and Information Sciences (ICCAIS), Gwangju, Korea, 2–5 December 2014; pp. 65–70.
37. Markkula, J.; Haapola, J. LTE and hybrid sensor-LTE network performances in smart grid demand response scenarios. In Proceedings of the 2013 IEEE International Conference on Smart Grid Communications (SmartGridComm), Vancouver, BC, Canada, 21–24 October 2013; pp. 187–192.
38. Parvez, I.; Islam, A.; Kaleem, F. A key management-based two-level encryption method for AMI. In Proceedings of the IEEE Power and Energy Society General Meeting, National Harbor, MD, USA, 27–31 July 2014; pp. 1–5.