A Novel Authentication Management for the Data Security of Smart Grid

Imtiaz Parvez, Maryamossadat Aghili, Hugo Riggs, Aditya Sundararajan, Arif I. Sarwat, and Anurag K. Srivastava

***Abstract*—Bidirectional wireless communication is employed in various smart grid components such as smart meters and control and monitoring applications where security is vital. The Trusted Third Party (TTP) and wireless connectivity between the smart meter and the third party in the key management-based encryption techniques for the smart grid are expected to be totally trustworthy and dependable. In a wired/wireless medium, however, a man-in-the-middle may seek to disrupt, monitor and manipulate the network, or simply execute a replay attack, revealing its vulnerability. Recognizing this, this study presents a novel authentication management (model) comprised of two layer security schema. The first layer implements an efficient novel encryption method for secure data exchange between meters and control center with the help of two partially trusted simple servers (constitutes the TTP). In this setting, one server handles the data encryption between the meter and control center/central database, and the other server administers the random sequence of data transmission. The second layer monitors and verifies exchanged data packets among smart meters. It detects abnormal packets from suspicious sources. To implement this node-to-node authentication, One class support vector machine algorithm is proposed which takes advantages of the location information as well as the data transmission history (node identification, packet size, and data transmission frequency). This schema secures data communication, and imposes a comprehensive privacy throughout the system without considerably extending the complexity of the conventional key management scheme.**

I. Introduction

The term ”smart grid” equates to the advanced power system that incorporates bidirectional communication, ubiquitous computer capabilities, and intelligent technology to enhance reliability, control, efficiency, and safety within the distribution system. The Advanced Metering Infrastructure (AMI) serves as a fundamental component at the distribution level. It comprises a vast number of interconnected meters organized in a hierarchical or mesh or hybrid networking configuration. Wireless technology is used by the meters to connect with the control center. The commonly employed communication protocols for AMI encompass ZigBee, WiFi, and LTE [1], [2]. Among these diverse communication standards, ZigBee has attracted considerable attention [3], [4]. However, the short transmission range of ZigBee hinders the feasibility of transferring data directly from smart meters to the control center leading to cooperative transmission. Within the cooperative framework, each smart meter is responsible for gathering and retaining real-time energy consumption data. This information is subsequently transmitted at regular intervals to the control center by relaying it through adjacent nodes. The data packet continues to be forwarded until it reaches the data collector, at which point the aggregated packet data is dispatched to the control center [5]–[7].

Due to the utilization of wireless connection and hop-to-hop data aggregation and forwarding, security concerns emerge as a significant challenge for AMI. By analyzing use patterns, an adversary or thief may be able to anticipate the whereabouts of targeted customers within their residences, so posing a potential risk to their personal safety and privacy. Additionally, via the analysis of detailed energy consumption data, home appliance firms are able to get insights into consumers' lifestyles and habits and the energy usage of their household equipment. Hence, rival firms companies might want to eavesdrop in the hop-to-hop communication and can use this valuable information in their businesses. Consumers would want to tamper with consumption data to reduce their electricity bill. The most crucial thing is that the opponent/hacker might jam or take over the AMI network by sending a false signal to meters on an unsecured system, which may cause a wide area power outage along with an imbalance in the demand generation model.

The main obstacle for implementing AMI security scheme is the tightly bounded memory and low computation capability of the smart meters which calls for a lightweight and resilient security scheme. The key management-based encryption approach has been recognized in the literature as a notable security system for the smart grid, which incorporates a Trusted Third Party (TTP) [8]–[13]. Almost all TTP management solutions make the premise that the TPP can be entirely trusted. However, the TTP, the meters, and the communication linkages between the TTP and the meters might all be breached.

Taking into account semi-trusted servers and untrustworthy/unreliable communication channels, this paper presents an authentication management consists of two-layer security scheme. The first layer boosts the security of the data transmissions between the Smart Meter (SM) node and the control center/Metering Data Management Service (MDMS) by data encryption as well as randomized packet transmission. As mentioned in our early work [14], the scheme consists of two seperate servers. Secure communication using public-private key management between every smart meter and MDMS is handled by the master server. On the other hand, the auxiliary server manages the transmitted sequence of the data packet (using a public key received from the master server). The private key associated with the public key and generated random sequence are used to retrieve the data at the MDMS. This paper extends [14] by using One Class Support Vector Machine (OCSVM) and Received Signal Strength (RSS) techniques for authentication in node-to-node links. OCSVM is used to detect malicious packets from unknown sources considering data transmission history like transmission frequency, data packet size, and distance between sender and receiver. RSS algorithm is applied to localize meters via the RSS from its neighbor meters. In the prevision version of work [14], only RSSI based distance was used on node authentication algorithm. On the other hand, on OCSVM based node-to-node authentication, we use three features which makes the node-to-node authentication more robust. And the main motivation for selecting OCSVM for node authentication is because of its memory efficiency, effectiveness in small and medium-sized datasets, robustness to over fitting, various Kernel options, and global optimization advantage.

This comprehensive approach has, to the best of our knowl-

Imtiaz Parvez is with the Department of Computer Science, Utah Valley University, Orem, Utah, USA e-mail:imtiaz.parvez@uvu.edu

Maryamossadat Aghili, Hugo Riggs,and Arif Sarwat are with the Department of Electrical and Computer Engineering, Florida International University, Miami, Florida, USA e-mail:maghi001@fiu.edu, hrigg002@fiu.edu, asarwat@fiu.edu

Aditya Sundararajan is with Oak Ridge National Laboratory, Tennessee, USA e-mail:sundararajaa@ornl.gov

Anurag K. Srivastava is with the Department of Computer Science and Electrical Engineering, West Virginia University: Morgantown, West Virginia, USA. e-mail:anurag.srivastava@mail.wvu.edu.

edge, never been introduced in the literature previously. The utilization of two separate servers for key management, together with the implementation of randomly sequenced packet transmission, enhances the level of security resilience in an untrusted communication links and servers. Furthermore, node to node authentication based on OCSVM secures inter-node communication without the considerable overhead, and limited resource requirement makes it a suitable technique for a trusted node to node communication. Additionally. The usual data traffic flow between the meters and control center is unaffected by the communication between the meters and servers, which happens once for each session of data transfer. Therefore, our approach provides significant improvement in conventional key management based system by two level security- data security and node authentication. Additionally, by prudent design of cluster of meters served by a TTP (a master and an auxiliary serve) we can make this approach scalable.

The subsequent sections of the paper are structured in the following manner. Section II provides an overview of the existing literature pertaining to the security techniques that have been suggested for the AMI. In Section III, the architecture of the suggested model is presented. In Section IV, we provide a discussion on the theoretical foundations of RSS-based localization, OCSVM, and entropy of data packet in relation to our suggested system. Section V of this paper provides a detailed analysis of the implementation of the security strategy, as well as the communication flow between smart meters, servers, and MDMS. The analysis of simulation results and the theoretical security strength of a data packet is conducted in Section VI. In conclusion, this paper presents the final thoughts in Section VII.

II. Literature Review

Security challenges in AMI have received a substantial attention in recent years from a variety of communities, including electrical engineers, computer scientists, and IT specialists [15]–[17]. We can divide the works from the literature into major two categories: (1) non key management system based schemes and (2) key management system based schemes.

In non key management system based security schemes, few diverse approaches have been proposed for reinforcing the security in the AMI. In [18], randomization of the AMI configuration is proposed to make its behavior unpredictable to the hacker, whereas the behavior is predictable to the control center. In [19], [20], authors introduced anonymization of data by randomizing node identity using a TTP. But, communication overhead may be increased due to the need for the TTP to communicate with all nodes simultaneously. In [21], [22], homomorphic encryption has been introduced. Though it requires minimal calculation at data retrieval, but it may be complicated for a large network. In [23], the authors introduce a blockchain-based lightweight solution that utilizes a received signal strength indicator (RSSI) for localization and provenance through blockchain. The adversarial nodes can be identified with the variation of RSSI. The paper [24] presents an approach that aims to ensure privacy preservation during authentication and data aggregation in a smart grid system utilizing fog computing technology. The authors propose to employ the techniques of short randomizable signature and blind signature to establish a system of anonymous authentication. In addition, the smart meter readings are consolidated using the homomorphic Paillier cryptosystem.

On the other hand, in case of key management system based schemes, vast works can be found in the literature. The key management system based security schemes can be categorized into four major categories [16]: (1) Key graph based scheme, (2) Authentication based scheme, (3) PUF based scheme, and (4) Hybrid scheme. In addition to introducing Information-Centric Networking (ICN) in AMI systems, the authors in [11] introduce a key graph-based key management system for numerous smart meters. In the proposed ICN-AMI structure, a key graph consists of the user key, the group key, and the root key for unicast, multicast, and broadcast purposes. This approach addresses the confidentiality and integrity of data, but it does not address authentication. In [25], the authors propose Multi-Group Key Graph (individual and batch rekeying) based Versatile and Scalable key management scheme for AMI (VerSAMI) and VerSAMI+ which supports unicast, multicast, and broadcast communications. It addresses the packet overhead and does not address the node/packet authentication. In order to secure the smart metering network, the paper [26] introduces a sophisticated hybrid encryption system which combines public and symmetric key encryptions. The foundational components for the suggested scheme are the Elliptic Curve Integrated Encryption Scheme (ECIES) and Advanced Encryption Scheme (AES). A precomputation approach that offers quicker encryption and decryption is given in order to reduce the computational cost of ECIES. In [27], the author introduces a simple anonymous authentication and key agreement approach for the smart grid, which enables the service provider and the smart meter to establish a shared session key and authenticate one another. In comparison to current smart grid authentication systems, it ensures the anonymity and untraceability of the smart meter while achieving quick mutual authentication between the service provider and the smart meter.

In [28], event driven asset centric key management is proposed where key management (i.e. key generation, refreshment, revocation, etc.) is orchestrated automatically based on events from assets or nodes. In [29], public key management has been proposed for smart grid based on elliptic curve public key cryptography and Needham Schrouder symmetric key authentication. Even though, scalability and simplicity are two advantages of this approach, it does not come with experimental proof. In [30], symmetric key establishment mechanism is proposed based on X.1035 standard which reduces data delivery time up to 75%. In [31], group key management with three-tier network model is proposed which requires moderate key storage. The paper [32] introduces a novel authentication protocol that incorporates a key establishment mechanism. The system enables service providers to securely commence communication with multiple smart meters, facilitating the dynamic update of power consumption data. The protocol under consideration has been formally verified for correctness using GNY logic. In the paper [33], the authors proposes a novel quantum-defended lattice-based anonymous mutual authentication and key-exchange (MAKE) protocol for secure group (SG) systems. The suggested technique has the capability to achieve resilient conditional identity anonymity and key management through the utilization of small integer solutions and inhomogeneous small integer solutions lattice hard assumptions. This eliminates the need for additional complex cryptographic primitives.

In [34], a security scheme is proposed for smart meters applying digital signatures that a trusted third party can sign and timestamp. Additionally, data hashing using SHA-256 ahead of applying the signature adds a layer of security. For an end-to-end communication solution, [35] proposes Identity-Based Signcryption (IBS) with zero configuration encryption and authentication. In [36], encryption of node-to-node links using a secret key has been proposed. However, in large networks, packet overhead might be increased for both IBS and node-to-node authentication. In [37], Diffie–Hellman key protocol based message authentication is proposed in addition to Hash based message authentication. This approach provides higher scalability, lower memory utilization and less delay for decryption. Four broad countermeasures to thwart attacks on smart meters have been proposed in [38]: (1) authentication and strong encryption of communication that deals with HAN and NAN and buses within smart meters, (2) secure key

management which form the critical backbone to a secure AMI, (3) securing the firmware to avoid being manipulated by the attackers or mistakenly by authorized personnel, and (4) security-driven firmware development cycle that conducts frequent walkthroughs and security assessments.

The protocol proposed in the paper [39] utilizes a physically unclonable function (PUF) and incorporates a one-time pad mechanism. This approach offers a notable advantage by eliminating the need for the Diffie-Hellman key setup protocol. The cryptographic key utilized in this scenario is derived from a genuine random source known as a Physical Unclonable Function (PUF), which is included within the secure module's (SM) hardware. In [40], authentication between smart meter and utility server along with low overhead key management has been proposed. The mutual authentication consists of four steps whereas the key management is founded on ID based public/private key pair model with lower transmission overhead for key refreshment. The paper [41] presents an authentication mechanism called Anonymous Secure Authentication mechanism for the SG environment (ASAP-SG). ASAP-SG is capable of achieving authentication and key agreement between smart meters and service providers through the utilization of elliptic curve cryptography and physical unclonable function. The proposed approach minimizes the computational cost for both smart meters and service providers, resulting in reduced communication overhead. The author in [42] presents a key establishment protocol for secure group communication that possesses several desirable characteristics such as a high level of anonymity, resilience against well-known attacks with perfect forward secrecy, and efficiency in terms of computational and communication costs. It eliminates the need for a Public Key Infrastructure (PKI) and the number of necessary messages for mutual authentication is reduced to merely two.

To distribute the keys and manage the network, a wireless sensor network based Public Key Management Infrastructure (PKI) has been proposed in [12], [43]. However, it requires to generate a large numbers of unique keys for a large networks. In [44], a Key Management System (KMS) has been introduced based on DLMS/COSEM standard providing two main information security features: data access security and data transport security. Since DLMS/COSEM is an open standard and allows a number of variations in the protocol implementation, it might increase the complexity in the client side. In the study [45], a security architecture consisting of two layers was presented to ensure the security of communication between the meter and the Data Concentrator (DC), as well as between the DC and the control center. The recommended approach for encrypting the meter-to-DC communication is based on the IEC 62056 standard. Similarly, for the DC to control center communication, the use of a public Key Management System (KMS) has been suggested. However, it is necessary to conduct encryption and decryption twice in each time step.

In contrast to previous research in the literature, our proposed approach offers enhancements in the AMI systems. In our approach, we use randomization of data transmission and node-to-node authentication methods to address the challenges posed by unreliable communication scenarios. Additionally, we leverage a machine learning technique to enhance the effectiveness of our proposed method. The use of data packet randomization and the utilization of machine learning for node authentication enhances the robustness of the schema without incurring additional costs.

III. Architecture of The Proposed Secured AMI

The proposed AMI is similar to a typical AMI which is a web-like network with millions of meters except it has two extra servers as shown in Fig. 1. Two mechanisms are proposed for the encryption of each data packet and authenticate the communication among meters.

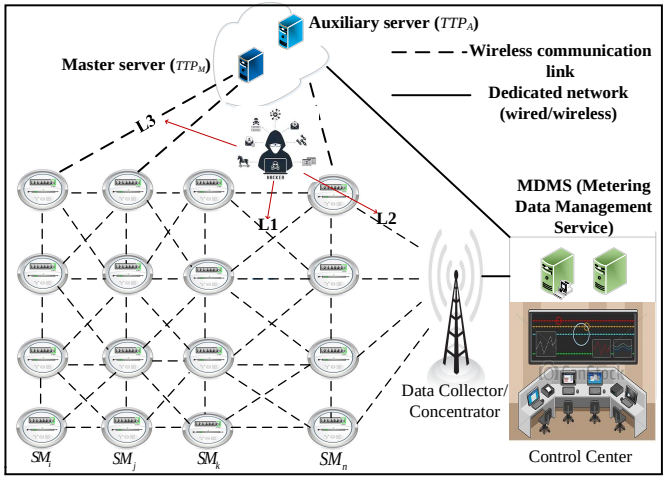

Fig. 1: An AMI architecture comprising a control center, data concentrator, and cluster of mesh- or hybrid-connected meters.

A detailed description of the components of proposed AMI is provided here.

- *Smart meter:* It is a solid state device responsible for collecting, storing and sending data to MDMS using wireless communication in a fixed interval time less than 1 hour.
- *Home Area Network (HAN):* All home appliances are connected to a smart meter by a network, forming HAN.
- *Neighborhood Area Network (NAN):* The meters are connected to each other through a mesh or hierarchical or hybrid wireless network termed as NAN. In our architecture, we assume that NAN utilizes ZigBee protocol.
- *Data Collector/concentrator:* The head end of the NAN is the data concentrator or gateway which collects data from NAN and forwards those to MDMS by a dedicated wired or wireless connection (e.g., optical fiber, a cellular network, etc.).
- *Metering Data Management Service (MDMS) /control center:* The MDMS receives the consumption data from the AMI network, and calculates bills based on them. Having fine-grained collected data, MDMS also monitors, manages, and optimizes power generation and electricity distribution in the grid.
- *Master server:* Master server generates a pair of private and public keys for a SM ahead of each session. The public key is unicasted by the master server to be used in encryption. On the other hand, the unicasted private key is received at the Auxiliary server and MDMS for decryption. In the proposed architecture, the connection between the Master server and NAN is via untrusted wireless communication whereas the Master and Auxiliary server are connected to MDMS by reliable communication (such as optical fiber, 5G).
- *Auxiliary server:* Ahead of data encryption using the public key received from the Master server, the smart meter creates a random sequence. This sequence is encrypted by public key and sent to the Auxiliary server. The Auxiliary receives the random sequence then authenticates it using the private key of the smart meter before forwarding it to MDMS for final decryption.

Since in our model, a cluster of meters is supported by a Master and an Auxiliary server, our scheme can be scalable by prudent design of clusters.

A. Basic Notifications and Definitions

The notifications and definitions used in the presented algorithm and data flow scheme are stated in Table I.

TABLE I: Symbol notations and definitions.

Notation	Description
SM_i	Smart meter node i
TTP_M	Master server
TTP_A	Auxiliary server
$\mathcal{AE}_i$	Asymmetric encryption scheme for meter i
$\mathcal{K}_i$	Randomized key generation algorithm for meter i
pk_i	Public key for meter i
sk_i	Secret/private key for meter i
M_i	Cleartext message/data of meter i
C_i	Ciphertext message/data of meter i
$\mathcal{E}(\{u,v\},w)$	Encrypt the cleartext u and v with the key w
$\mathcal{D}(\{y\},q)$	Deterministically decrypt ciphertext y with the key q
t	Timestamp instance
z_i^t	Message packet size for node i at time t
n	Number of packet segments for a given meter i at time t
$(c_1, c_2, \cdots, c_n) \in C_i$	Segmented packets of cipher-text for meter i
$(r_1, r_2, \cdots, r_n) \in R^t$	Random sequences at time instant t
$(p_1, p_2, \cdots, p_n) \in P$	Probability of jth packet transmission
AP	Data concentrator/Access Point
γ	Path loss component
δ_l	Variance of random noise
SM	Set of all smart meters
L_x	Set of ZigBee connections, where $x \in [1,3]$
Req_i	Key request message sent by SM_i to TTP_M
N	Number of smart meter nodes, i.e. $i \in [1, N]$
PSO	Particle Swarm Optimization
OCSVM	One class support Vector Machine

Let, $\mathbf{SM} = \{SM_i\}_{i=1}^{N}$ denotes the set of participating smart meters connected as a network in our system. Also, let graph $G = (\mathbf{SM} \cup \{\mathcal{AP}\}, L_1 \cup L_2 \cup L_3)$ represents the network topology of smart meters where:

- AP represents the data concentrator,
- L_1 denotes the set of ZigBee connections connecting neighboring smart meters together,
- L_2 represents the set of ZigBee connections connecting the data collector AP to a few nearby smart meters (see Fig. 1 for illustration)
- L_3 is untrusted communication links (such as ZigBee, WiMax etc.) between the Master sever TTP_M/ Auxiliary sever TTP_A and every smart meter in **SM**,
- Both the TTP_M and TTP_A are connected to MDMS via a dedicated network. Additionally, they are connected by a trusted connection.

It is worth mentioning that each SM has a unique ID which is stored in the data packet, TTP_M and TTP_A. TTP_M and MDMS exchange the corresponding private key based on this ID. They both stored the corresponding private keys of the SM in their internal databases to retrieve at the time of packet arrival for decryption purpose.

For a node-to-node authentication of data packets, a machine learning algorithm is proposed to run in each smart meter. The proposed algorithm explores all the incoming data packets in real time and identifies whether it is reliable or spiteful based on the three features of the packet including 1) the distance of the packet sender which is estimated by RSS-based localization algorithm, 2) time intervals that a packet is received in destination, and 3) packet size. Based on the decision of the algorithm, the packet is accepted and forwarded to the next SM in the grid, otherwise the packet will be discarded. Fig. 2 illustrates the mechanism.

IV. RSS Algorithm, OCSVM and Entropy of a Data Packet

As mentioned earlier, packet data is circulated among the meters in the path reaching the destination AP. To point out the malicious packet data from an unauthorized source, the content of the data packet should be screened carefully before delivery and dispatch to the next node in the grid. One class classification, or concept learning in the absence of counter examples, has the potential to tackle these kinds of problems. Among different implementations of the Support Vector Machine (SVM), one class classification algorithm (OCSVM) is selected in this work and the performance is compared with another anomaly detection algorithm: Isolation Forest [46], [47]. Appropriate features of the packet data should be extracted and fed into the OCSVM for training and later on for testing the new packet. Location (i.e. distance) of the sender, which is an informative feature for detecting the unauthorized source, is not directly defined in the data packet. To extract this feature, RSS algorithm is utilized. RSS can pinpoint the location of neighboring meters based on the received electromagnetic signals. Other features of the data packet such as packet size, and transmission frequency are relatively simple to capture or infer. In the following part, the RSS and OCSVM algorithm are discussed in details.

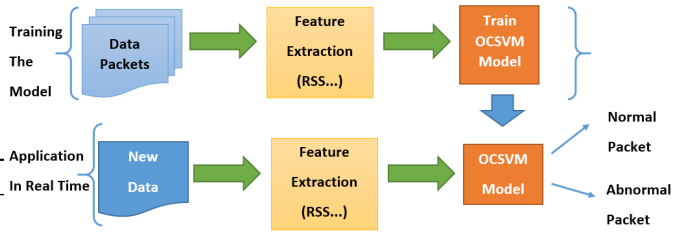

Fig. 2: Training and application of OCSVM model in smart meters

A. Received Signal Strength (RSS) based Localization

Let's assume an unknown positioned meter at a location (x, y) accompanied by somewhat dispersed meters of known position at locations (x_l, y_l), where $1 \leq l \leq n$. The received signal strength at location (x_l, y_l) from the unknown position meter can be denoted by ψ_l [48]–[50]

$$\psi_l = c - 10\gamma \log(d_l) + w_l, \tag{1}$$

Such that c is an unknown constant that depends on transmission power, frequency, etc., and γ is the path loss exponent. Path loss exponent defines the decay rate of electromagnetic signal. In our model, $\gamma = 2.93$ is used regarding a residential area. The parameter d_l represents euclidean distance between the known and unknown position meter defined as follows:

$$d_l = \sqrt{(x - x_l)^2 + (y - y_l)^2}, \tag{2}$$

and w_l is the zero mean random Gaussian noise with standard deviation σ_l. The value of σ_l ranges from 6 to 12 dBm.

Let us define, the $\boldsymbol{\theta}$ and $\boldsymbol{\psi}$ as $\boldsymbol{\theta} = [x, y, z]^T$ and $\boldsymbol{\psi} = [\psi_1, \psi_2, ..., \psi_n]^T$, where z is the reference transmission power.

The likelihood function of $\boldsymbol{\theta}$ for a given RSS measurement $\boldsymbol{\psi}$, $f(\boldsymbol{\theta}|\boldsymbol{\psi})$ is given by

$$f(\boldsymbol{\theta}|\boldsymbol{\psi}) = c_1 \exp\left\{ -\sum_{l=1}^{n} \frac{\{\psi_l - c + 10\gamma \log(d_l)\}^2}{2\sigma_l^2} \right\}, \tag{3}$$

where c_1 is a constant.

[1]Unless or otherwise specified, node and meter are same thing in the rest of the study.

The Maximum Likelihood (ML) estimate of $\boldsymbol{\theta}$, denoted by $\hat{\boldsymbol{\theta}}$, can be found from the following equation

$$\begin{aligned}\hat{\boldsymbol{\theta}} &= \arg\max f(\boldsymbol{\theta}|\boldsymbol{\psi}) \\ &= \arg\min \left\{ -\sum_{l=1}^{n} \frac{\{\psi_l - c + 10\gamma \log(d_l)\}^2}{2{\sigma_l}^2} \right\},\end{aligned} \tag{4}$$

The equation presented above is an optimization problem. A range of optimization approaches, including differential evolution, dynamic relaxation, and Particle Swarm Optimization (PSO), can be employed for the purpose of solving (4). In the current problem, PSO is used to solve the nonlinear optimization problem. Finally, the ML estimator yields the location (x, y) and reference power z of the unknown positioned meter

$$(x, y, z) = \{\hat{\theta}(1), \hat{\theta}(2), \hat{\theta}(3)\}. \tag{5}$$

Now the distance between any two meters SM_i and SM_j is

$$d_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}. \tag{6}$$

where (x_i, y_i) and (x_j, y_j) are derived positions of meter SM_i and SM_j respectively. The distance d_{ij} is used as a feature in OCSVM algorithm.

Since GPS doesn't work in some places such as inside the multi-stored building, hilly areas, forests, etc., we used RSS based localization over GPS. Additionally, GPS reveals exact position of meters/consumers which we want to avoid. On the other hand, RSS technique will build a local map for meters.

B. OCSVM Algorithm

One class SVM classifier is motivated by the SVM classifier [51], [52]. The one class classification problem allows to find the hyperplane to separate the training distributions from the origin of the feature space.

OCSVM maps the input vector to feature dimension according to the kernel function, and separates it from the origin with maximum margin.

Let us consider, a set of training data $I = (i_1, i_2, ..., i_n) \in \mathcal{I}$, and Ω be the feature map $\mathcal{I} \longrightarrow \mathcal{H}$ such that the dot product of $\mathcal{H}$ is computed by kernel k

$$k(i, i') =< \Omega(i), \Omega(i') >_{\mathcal{H}} \tag{7}$$

The regular family for the data set

$$C^m_{w,\rho} = \{i | f_{w,\rho}(i) > 0\} \tag{8}$$

where $f_{w,\rho}(i) = \text{sgn}(< w, \Phi(i) > -\rho)$ and (w, ρ) is the vector to offset parameterizing a hyperplane in the feature space associated with kernel.

$f_{w,\rho}$ is estimated by minimizing regularization

$$R^{reg}[f^m_{w,\rho}(\cdot)] = R^{emp}[f^m_{w,\rho}(\cdot)] + \frac{1}{2}||f^m_{w,\rho}(\cdot)||_{\mathcal{H}} \tag{9}$$

Outliers are penalized via slack variables ξ operating in the objective function to control the trade-off from empirical risk and regularizes the penalty.

The quadratic programming minimization function

$$\min_{w,\xi_i,\rho} \frac{1}{2}\|w\|^2 + \frac{1}{vn}\sum_{i=1} n\xi_i - \rho \tag{10}$$

$$\begin{aligned}&\text{such that } (w \cdot \Phi(x_i)) \geq \rho - \xi_i, \\ &\text{and } \xi_i \geq 0, \qquad\qquad i = 1, 2, 3, ..., n.\end{aligned} \tag{11}$$

where Φ is the kernel function for mapping, ξ_i is the slack variables, $v \in (0, 1]$ is a fixed constant, and ρ contains decision value it determines if a given point is inside the estimated high density region. Thus the decision function $f^m_{w,p}(x)$ is of the form

$$f(x) = \text{sgn}(w^{*T}\Phi(x) - \rho^*) \tag{12}$$

where ρ^* and w^* are the values of w and ρ solving from the equation (10).

In OCSVM, C is the smoothness operation and v is characteristic of the solution [53]:

- Symbol v determines an upper bound on the outliers.
- It is the lower bound for amount of of training samples for use as support vectors.

C. Entropy of a Data Packet

Entropy is a metric for analyzing the robustness of an encryption methodology [14]. In other word, entropy demonstrates the feasibility degree of capturing the lock by chance. The more certain about a value, the more diminished the entropy.

The entropy for a sequence S

$H(S) = \sum_x P(S = x) \log_2 P(S = x)$

Such that, $P(S = x)$ is the probability of taking S a value x.

If the size of a random variable or packet generated by a meter is n bit, then the entropy and security strength of the data packet are n and 2^n, respectively. The higher the entropy, the harder the decryption process. For analyzing the performance of the proposed encryption schema, this metric is selected.

V. Privacy Scheme Implementation and Data Traffic Flow

In this section, the privacy scheme implementation and the data flow process are described in details to clarify how each layer of the schema affects the grid security. In the data flow architecture, the following assumptions exist:

- The Master and Auxiliary servers are semi-trusted and independent. However, they might physically be the same machine however virtually divided into two servers.
- The wireless communication links between servers and meters are not fully reliable.
- The meters have small memory and computation capability.
- The control center/MDMS has the adequate computational ability.
- The meters keep the records of the position of neighboring meters, the frequency of transmission, packet size, and node identity. The frequency of transmission, node identity, and packet size are extracted from the packet header. The node position is derived from electromagnetic signals using RSS based localization as explained in previous section IV(A).
- Every meter transmits data at a constant transmission power.
- The data packet size is constant for every meter, and is 128 KB in the studied grid.
- After installing a new meter, it starts to record the position of the neighbor meters, frequency of data transmissions, node identity, and packet size.
- We assume that AMI network use high penetrating frequency bands such as 2.4 GHZ and 3.5 GHZ and the smarts are equipped with advanced signal processing techniques.

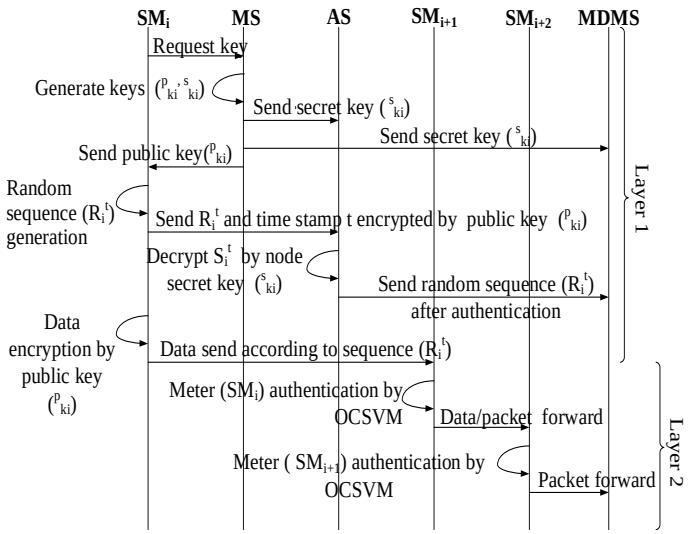

Fig. 3: Data flow among various components of AMI.

- The communication links among Master and Auxiliary servers and MDMS are fully trusted.
- It is noted that the security scheme adds latency to data communication. However, the delay is much less than expected data communication timeline of AMI, as energy bills and household privacy and 'ping' for on demand data have lower requirements in terms of timeline.

A. Attack Model

A simplified attack model for the system model shown in Fig. 1 is discussed here. Man-in-the-Middle (MITM) and replay are two of the most common attacks conducted in the AMI. While MITM leverages the vulnerabilities in the communication medium to intercept, swap, corrupt or steal sensitive data, the replay attacks simply capture a copy of legitimate information and replay it in the future, posing to be the same legitimate user. To this end, it is assumed in this paper that the attacks can happen only in L_1, L_2 or L_3, but not between TTP_M, TTP_A and MDMS. It is noted that other kinds of attacks such as compromised nodes and puppet attacks [54], [55] are out of scope in this study.

A critical assumption of this attack model is that the attacker's targets are solely on the communication channel between the devices, but not the devices themselves. It is assumed that the smart meter nodes, TTP_M, AP, TTP_A and MDMS are not compromised by the attacker, but only the links L_1, L_2 or L_3 are. The incentive for the attacker could be offsetting energy consumption to achieve lower bills, stealing sensitive information to monitor energy consumption profile and through that spy on consumer behavior, or corrupting encrypted data packets in L_1 or L_2. The proposed encryption scheme relies on the strength of the secret key sk, hence falling under the category of Public Key Encryption (PKI).

B. Data Traffic Flow

The data flow among the meters, servers, and MDMS (as shown in Fig. 3) is explained below in details.

STEP 1: Initialization
SM_i sends a request Req_i for a public key pk_i to the Master Server TTP_M. The Master server generates a public key-private key pair, (pk_i, sk_i), upon receiving the request. While pk_i is unicasted to SM_i for data encryption, sk_i is sent to the Auxiliary server TTP_A and MDMS. Key generation

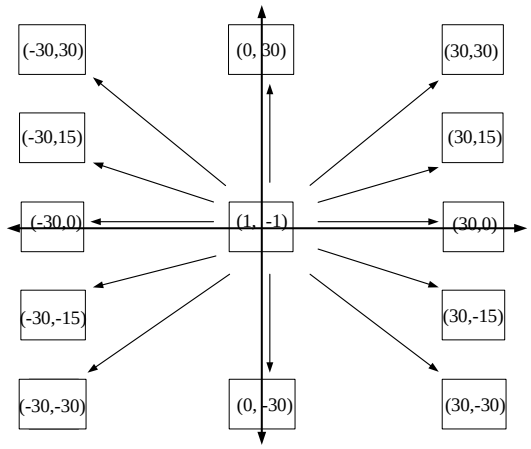

Fig. 4: Meter positions in Manhattan grid.

by asymmetric algorithm:

$$\mathcal{AE}_i = (K_i, \mathcal{E}_i, \mathcal{D}_i), i \in [1, n] \tag{13}$$

$$K_i \longrightarrow (pk_i, sk_i), i \in [1, n] \tag{14}$$

STEP 2: Encryption
The SM_i generates a random sequence (R_i^t) for timestamp instance t. It encrypts both the sequence and the timestamp using pk_i to generate a ciphertext, S_i, and sends it to TTP_A. This server receives S_i and decodes it using sk_i. It then performs two steps to validate the sequence. To safeguard against the replay attacks, TTP_A ensures if the timestamp instance t is received within a stipulated limit predefined for "freshness" of data. If it is within that limit, it considers the request, else it rejects the packet. To ensure whether the sequence was sent by a legitimate SM_i, the server re-encrypts the sequence it decoded along with the timestamp instance t using the pk_i it received from TTP_M. It compares the ciphertext it created with the one it initially received. A discrepancy between the two ciphertexts prompts TTP_A to reject the packet. A match indicates the packet is indeed legitimate. In this way, data transmitted from SM_i is verified. After that, TTP_A sends the sequence R_i^t to MDMS. At the same time, SM_i encrypts its consumption data by the following method:

$$C_i \longleftarrow \mathcal{E}(\{M_i, t\}, pk_i) \tag{15}$$

STEP 3: Data transmission
The encrypted data C_i is segmented into n packets of equal size, z_i^t.

$$(c_1, c_2, c_3, \cdots, c_n) \longleftarrow C_i \tag{16}$$

Then, the packets are ordered based on sequence R_i^t for transmission.

$$(c_3, c_1, c_4, \cdots, c_n) \xleftarrow{R_i^t} (c_1, c_2, c_3, \cdots, c_n) \tag{17}$$

The transmitting algorithm is explained in Algorithm 1:

STEP 4: Hop to hop data aggregation and forwarding
At first, meter SM_i is verified by SM_{i+1} using the node ID that is perceived from packet header. Afterwards, based on the history (i.e. previous data receiving history) of sender node's distance (d_{xy}), the frequency of data received (f_j) and packet size (z_j)- the node SM_{i+1} verifies source node SM_i and forwards data to the next node SM_{i+2}. OCSVM algorithm is used to authenticate in this process. The data aggregation and forwarding algorithm pseudocode is tabulated in Algorithm 2:

STEP 5: Data retrieval

Algorithm 1 Transmitting algorithm

Initialization:
Get ith meter's data packet C_i, $i \in \{1, 2, ..., N\}$ and random sequence $R_i^t = \{r_j\}$, $j = \{1, 2, 3, \cdots, n\}$, $\forall t \in T$ and $|R_i^t| \in \mathbb{N}$
if $R_i^{t-1} = R_i^t$ **then**
    Go to step 2
else
    Proceed to next step
end if
Segment packet $\mathcal{C} = \{c_j\} \longleftarrow C_i$, $j = \{1, 2, 3, \cdots, n\}$ and $|\mathcal{C}| = |R_i^t|$
Set index set $J_{\mathcal{C}} = \{l\}$, $l = 1, 2, 3, \cdots, n$ where $f : l \longrightarrow \mathcal{C}$ is the particular enumeration of $\mathcal{C}$
Update $f^{-1}(c_j) = r_j$, $j = \{1, 2, 3, \cdots, n\}$.
Calculate transmission probability, $P = \{p_j\}$ where $p_j = \frac{1}{r_j} = \frac{1}{f^{-1}(c_j)}$, $j = \{1, 2, 3, \cdots, n\}$.
Set index set $K_P = \{k\}$, $k = 1, 2, 3, ...n$ where $g : k \longrightarrow P$ is the particular enumeration of P
Sort index $l = f^{-1}(c_j) = g^{-1}(\max_{p_j \in P} p_j)$
Transmit packet $f(l)$ indexed packet
Update $\mathcal{C} = \mathcal{C} - \{f(l)\}$
if $\mathcal{C} \neq \Phi$ **then**
    Go to step 9
else
    End process
end if
End

Algorithm 2 Data aggregation and forwarding

Training:
For any two meters $\{SM_x, SM_y\} \in$ **SM** and time instant l, get regular data $I_{xy}^l = (d_{xy}^l, f_x^l, z_x^l), x \in \{1, 2, \cdots, N\}, y \in \{1, 2, \cdots, N\}$
For data set I_{xy}^l, define a family/boundary C_x^m through (8)
Meter/Packet authentication:
For time instant t, get relative distance $d_{i(i+1)}^t$ between source meter SM_i and data receiving meter SM_{i+1} through (6)
Calculate data transmission frequency f_i^t and packet size z_i^t for packet from meter i
For new data $I_{i(i+1)}^t = (d_{i(i+1)}^t, f_i^t, z_i^t)$, get decision by (12)
if $I_{i(i+1)}^t \in C_i^m$ **then**
    The data is within the boundary, forward to next meter SM_{i+2}
else
    Reject data from source SM_i which is flagged by algorithm as anomaly
end if
End

The MDMS receives the randomized and encrypted packets and decodes them by the secret key pk_i and random sequence R_i^t.
Reordering the data:

$$(c_1, c_2, c_3, \cdots, c_n) \xleftarrow{R_i^t} (c_3, c_1, c_4 \cdots, c_n) \tag{18}$$

Message unification:

$$C_i \longleftarrow (c_1, c_2, c_3, \cdots, c_n) \tag{19}$$

Decryption:

$$\{M_i, t\} \longleftarrow \mathcal{D}(C_i, sk_i) \tag{20}$$

VI. Simulation Results and Performance Analysis

In this section, we present the performance of RSS and OCSVM in our proposed AMI architecture. We also discuss about the security strength of our scheme and compare the performance of the OCSVM algorithm with another state of the art anomaly detection algorithm named Isolation Forest.

A. The Proposed Algorithm's Performance

To get insights into the localization of meters, we consider, Manhattan grid building topology [56] in which the distance between two meters (i.e. house) is 30m. We consider an area of interest (AOI) of 30m$\times$ 30m where a new meter is located at $(1, -1)$ position surrounded by known position meters as illustrated in Fig. 4. For optimization problem in localization, PSO [57] was used whereas residential path loss model was considered for path loss calculation. The transmission power of each meter is 10 dBm, and iteration number and population size of PSO are 100 and 30, respectively. The simulation results of position and power of an unknown positioned meter surrounded by 4 known positioned meters and environment with path loss constant 3 are tabulated in Table II.

TABLE II: Mean Square Error (MSE) in Localization for Different Noise Variances

Noise Variance	X	Y	Reference Power	MSE
2	-3.3059	-2.2506	10.0005	4.4839
4	-2.8715	-0.2784	9.9997	3.9324
6	7.3873	-1.9575	9.9841	6.4586
8	-4.8517	-5.9137	10.0175	7.6411
10	0.1728	-8.5027	9.9789	7.5482
12	2.3678	-8.8504	9.9818	7.9687

With an increase in the number of neighboring meters, the Mean Square Error (MSE) from the exact position of the meter decreases. This means that for the more number of neighboring meters, localization error for a meter will be lower. Secondly, for the increase of noise variance, MSE also increases. This implies that localization error for a meter is higher for the increased interference or noise. These are illustrated in Fig. 5(a). For the increase of both neighboring meters and path loss exponent, the MSE decreases from the exact position of meter which is reflected in Fig. 5(b). Path loss exponent (decaying rate of signal) is associated with the obstacle in the path of electromagnetic signal propagation. Therefore, for the presence of buildings, walls, trees, etc., the error in determining the location of meters will be higher. Furthermore, as meters are mounted on a stationary pole/wall, and in good conditions the environment around them is stable, the error for a meter by RSS based localization technique will be nearly uniform.

In the second part of the simulation, OCSVM is implemented in Python using scikit-learn library [58]. As a power system is a critical national infrastructure, it is very hard to get real-world power system data, especially cyber attack data. Since there is no appropriate public dataset for our simulation

This article has been accepted for publication in IEEE Open Access Journal of Power and Energy. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/OAJPE.2024.3393971

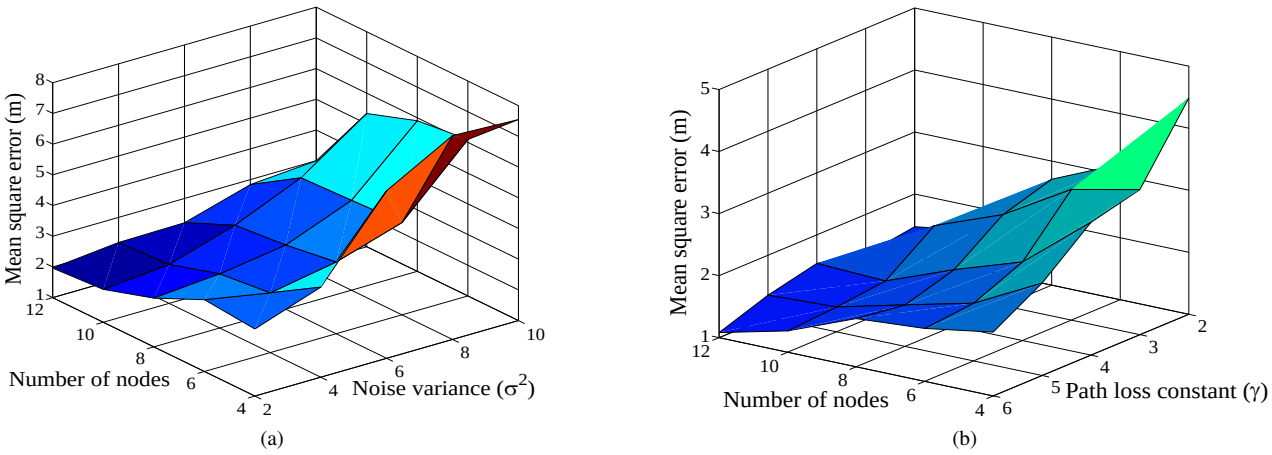

Fig. 5: (a) Mean square error for different number of nodes and noise variances. (b) Mean square error for different number of nodes and path loss exponent.

and obtaining the real data is not feasible, we generated a synthesized dataset simulating/mimicking the normal data packet structure due to the scarcity of the malicious packets and the unknown structure of the attacks. One of the important features of OCSVM is that if it is trained with the normal dataset, it can classify anything as malicious if it fails outside of the normal behavior. The structure of the normal packets and their transmission behavior is generated based on the defined standard of the network with a small variation. Regarding the current network standard, each packet data was generated with the following information: meter's position, the frequency of transmission, and the packet size. Some real-world data transferring from two smart meters in an AMI network of a local utility company is illustrated in Table III.

TABLE III: SMART METER REAL WORLD DATA SAMPLE

Device-Name	Read-Start-Time	Usage Unit	Meter ID	Coordinate
G0034501624	5/1/2013 0:00 1:00	0.3218 kWh	98747434	8668675880
G0034501624	5/1/2013 1:00 2:00	0.2757 kWh	98747434	8668675880
G0034501624	5/1/2013 2:00 3:00	0.3561 kWh	98747434	8668675880
G0034501637	5/1/2013 0:00 1:00	0.4587 kWh	98747434	8728604300
G0034501637	5/1/2013 1:00 2:00	0.4101 kWh	98747434	8728604300
G0034501637	5/1/2013 2:00 3:00	0.1346 kWh	98747434	8728604300

Based on the real world utility data (referred to Table III), we defined data transmission frequency equal to one hour. The distance between two meters is assumed to be 30m considering Manhattan Grid. The packet size is directly estimated upon the delivery of packet, and the standard packet size is considered 128KB based on the network topology. The training data is generated with 3 degrees of standard deviation from normal distribution of meter distance, data transmission frequency, and data packet size where the mean of meter distance, data transmission frequency, and packet size are 30m, one hour, and 128KB, respectively.

The OCSVM model was trained on the 70% of the data and the model was tuned on the validation set, 10% of the remaining data, with an exhaustive grid search, resulting in the RBF (Guassian) kernel with ν and γ both equal to 0.01. The model was tested by the rest of the data. The mapped decision boundaries of the OCSVM with the best parameter settings is shown in Fig. 6. Red lines show the decision boundaries and yellow dots are the packets from unauthorized sources. The training error[2], false positive rate[3] and false negative rate[4] of OCSVM are $4/200$, $3/20$, and $2/20$, respectively. On the other hand, the training error, false positive rate and false negative rate for Isolation Forest are $16/200$, $6/20$, and $1/20$, respectively. Therefore, it can be conceived that OCSVM can discriminate the authorized and unauthorized (malicious) packets almost precisely in comparison to it's counterpart Isolation Forest algorithm.

To evaluate the model's performance, confusion matrix and receiver operating characteristic (ROC) curve [59], [60] are provided. The statistics of true negative, true positive, false positive, and false negative of OCSVM and Isolation Forest are illustrated in Fig. 7. Referred to the figure, the overall accuracy of OCSVM is $(\frac{37+18}{60}) \times 100 = 91.6\%$. On the other hand, the accuracy of Isolation Forest is $(\frac{34+19}{60}) \times 100 = 88.33\%$.

Receiver Operator Characteristic (ROC) curve plots the true positive rate (TPR) against the false positive rate (FPR) at various threshold. ROC for this experiment with different tweaked parameters is shown in Fig. 8. It is noted that the area under the curve slightly changes based on the different hyper parameter setting, and for $\gamma = 0.01$ and $\nu = 0.01$, the maximum area of 0.95 is achieved; therefore the aforementioned parameters are selected as the best setting for the model.

B. Security Strength Analysis of a Data Packet

Let us assume, a meter generates a consumption unit packet of size 128KB (1024^3 bit) which is divided into 4000 blocks with each block size of 256 bit. If each block is encrypted by 256 bit public/asymmetric key and transmitted according to a random sequence, then the entropy of each block is 256. The security strength of the data block is 2^{256}.

Furthermore, the security strength of a 256 bit public key is $2^{256/2}$.

So, for 4000 random sequenced packets and 256 bit public key,

Total security strength of the 128KB meter data = $4000 * (2^{256} + 2^{256/2})$

Hence, a hacker needs maximum $4000 \times (2^{256} + 2^{256/2})$ number of iterations (tries) to decrypt a message, which is impractical.

[2]The training error is the ratio between the number of normal data that falls outside of the boundary erroneously and total number of data.

[3]The false positive rate is the ratio between the number of negative events wrongly categorized as positive (false positives) and the total number of actual negative events (regardless of classification).

[4]The false negative rate is the ratio between the number of positive events wrongly categorized as false (false positives) and the total number of actual negative events (regardless of classification)

This article has been accepted for publication in IEEE Open Access Journal of Power and Energy. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/OAJPE.2024.3393971

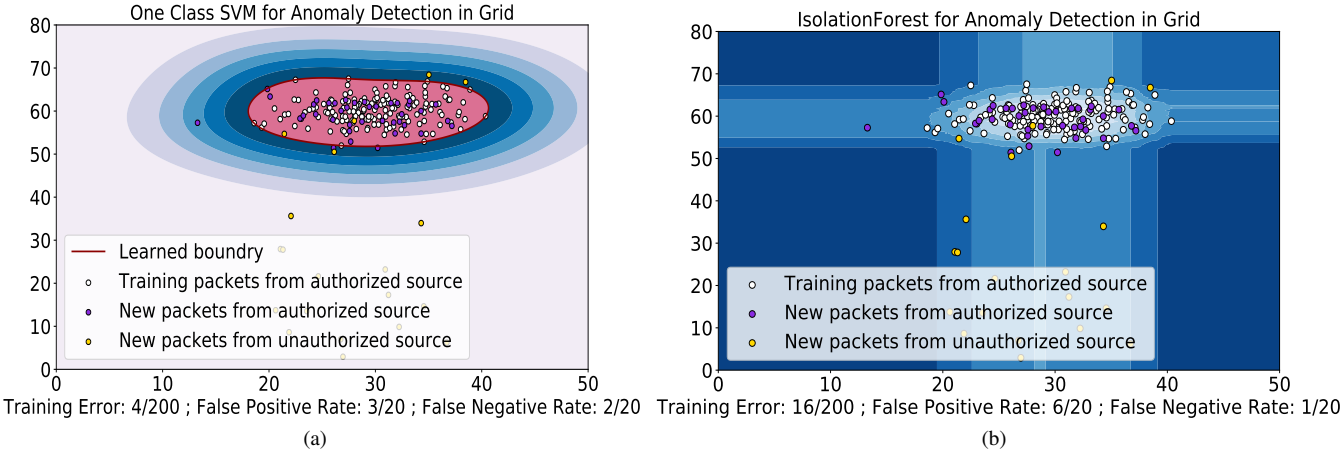

Fig. 6: Anomaly Detection with a) One Class SVM and b) Isolation Forest Algorithms

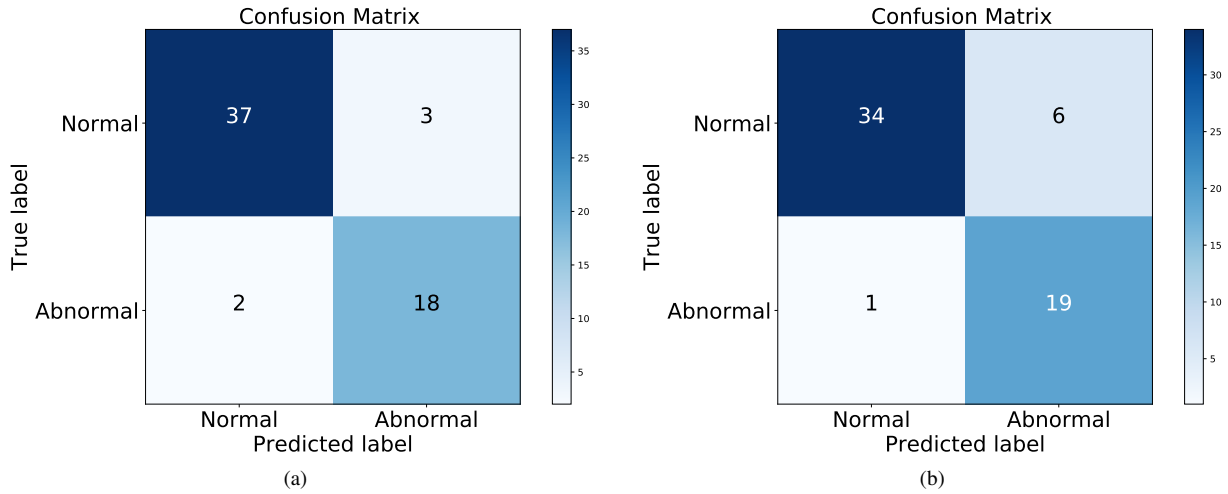

Fig. 7: Confusion Matrix for a) One Class SVM, and b) Isolation Forest

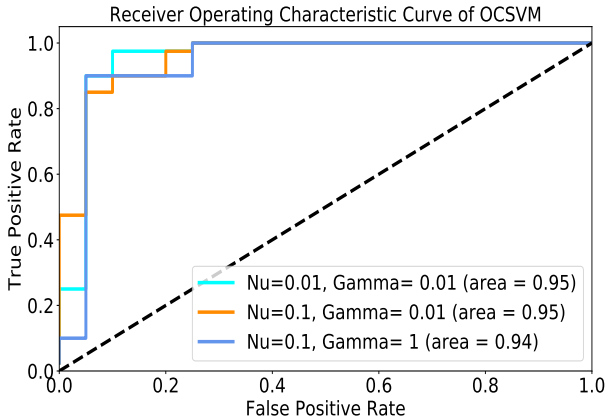

Fig. 8: ROC curve for OCSVM with different parameter settings. (Nu=ν; Gamma=γ)

VII. Conclusion

In our security scheme, a novel authentication management model comprised of two-level security method has been proposed with data encryption and node authentication. In the data encryption level, encryption by asymmetric keys and randomization of data packets have been proposed. In the conventional key management system, only data encryption is used. On the other hand, in our scheme, randomization of packets along with data encryption ensures enhanced data security. Another contribution of our scheme is the introduction of node-to-node authentication by OCSVM, which utilizes three features-frequency of data reception from a specific meter, packet size, and meter position. The features of data frequency and packet size can easily be extracted from the packet's header. To capture the last feature which detects the position of the sender of the packet, RSS model is used. In the case of authorized source this value would be very close or identical but in the case of attack, it varies.

For TTP-to-smart meter communication, we use a bi-directional communication similar to meter data communication of conventional AMI network. As communication from meters and servers occurs once per every session initialization between meter and control center, the data traffic of the normal meter data flow from meter to control center is not hampered. Furthermore, since a random sequence along with the asymmetric key is used to retrieve data in the control center, it also helps to authenticate data incoming from the smart meter. Additionally, being a cluster of meters is served by a TTP (a Master and an Auxiliary server), prudent design

of clusters can make our approach scalable easily.

By integrating Supervisory Control and Data Acquisition (SCADA) data with AMI systems, utilities can gain deeper insights into their infrastructure's performance and optimize resource allocation in real-time. However, integrating SCADA with AMI comes with inherent risks that need careful consideration. One significant risk is the potential for cybersecurity threats and vulnerabilities. When SCADA systems, which control critical infrastructure, are interconnected with AMI, which gathers sensitive consumption data, the attack surface for cyber threats expands. Hackers could exploit vulnerabilities in one system to gain unauthorized access to the other, potentially disrupting operations, manipulating data, or causing physical damage. Another risk is the complexity of integration between SCADA and AMI. SCADA and AMI systems may use different communication protocols, data formats, and security standards. Additionally, there is a risk of operational disruption during the integration process. To mitigate these risks in integrated SCADA and AMI, the utility should adopt a comprehensive approach to cybersecurity, including regular risk assessments, threat monitoring, employee training, and the implementation of industry best practices and standards such as NERC CIP (North American Electric Reliability Corporation Critical Infrastructure Protection) for the energy sector.

Additionally, a collaboration between smart meters and the distribution dispatching center is essential for ensuring the security and reliability of the power grid. The utilities can implement comprehensive and proactive approaches that encompass both technological solutions and robust cybersecurity protocols, including encryption and authentication, intrusion detection systems, secure firmware updates, and collaborative defense strategies to resist attack signals.

ACKNOWLEDGMENT

This article is based on the findings of the Ph.D. Dissertation titled "Spectrum Sharing, Latency, and Security in 5G Networks with Application to IoT and Smart Grid" [1] authored by Imtiaz Parvez.

REFERENCES

[1] I. Parvez, "Spectrum Sharing, Latency, and Security in 5G Networks with Application to IoT and Smart Grid," Ph.D. dissertation, Florida International University, 2018.

[2] M. Ghorbanian, S. H. Dolatabadi, M. Masjedi, and P. Siano, "Communication in smart grids: A comprehensive review on the existing and future communication and information infrastructures," *IEEE Systems Journal*, vol. 13, no. 4, pp. 4001–4014, 2019.

[3] N. Suhaimy, N. A. M. Radzi, W. S. H. M. W. Ahmad, K. H. M. Azmi, and M. A. Hannan, "Current and future communication solutions for smart grids: A review," *IEEE Access*, vol. 10, pp. 43 639–43 668, 2022.

[4] P. Yi, A. Iwayemi, and C. Zhou, "Developing ZigBee deployment guideline under WiFi interference for smart grid applications," *IEEE trans. on smart grid*, vol. 2, no. 1, pp. 110–120, 2011.

[5] C. Huang, C.-C. Sun, N. Duan, Y. Jiang, C. Applegate, P. D. Barnes, and E. Stewart, "Smart meter pinging and reading through ami two-way communication networks to monitor grid edge devices and ders," *IEEE Transactions on Smart Grid*, vol. 13, no. 5, pp. 4144–4153, 2022.

[6] M. A. Alomar, "An IOT based smart grid system for advanced cooperative transmission and communication," *Physical Communication*, vol. 58, p. 102069, 2023.

[7] I. Parvez, A. I. Sarwat, L. Wei, and A. Sundararajan, "Securing metering infrastructure of smart grid: A machine learning and localization based key management approach," *Energies*, vol. 9, no. 9, p. 691, 2016.

[8] M. Benmalek, Y. Challal, and A. Derhab, "An improved key graph based key management scheme for smart grid ami systems," in *2019 IEEE Wireless Communications and Networking Conference (WCNC)*, 2019, pp. 1–6.

[9] Y. Hanna, M. Cebe, S. Mercan, and K. Akkaya, "Efficient group-key management for low-bandwidth smart grid networks," in *2021 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm)*, 2021, pp. 188–193.

[10] Y. Yan, Y. Qian, and H. Sharif, "A secure and reliable in-network collaborative communication scheme for advanced metering infrastructure in smart grid," in *Proc. IEEE Int. conf. on Wireless Communications and Networking Conference (WCNC)*, March 2011, pp. 909–914.

[11] K. Yu, M. Arifuzzaman, Z. Wen, D. Zhang, and T. Sato, "A Key Management Scheme for Secure Communications of Information Centric Advanced Metering Infrastructure in Smart Grid," *IEEE Transactions on Instrumentation and Measurement*, vol. 64, no. 8, pp. 2072–2085, Aug 2015.

[12] A. V. D. M. Kayem, H. Strauss, S. D. Wolthusen, and C. Meinel, "Key Management for Secure Demand Data Communication in Constrained Micro-Grids," in *2016 30th International Conference on Advanced Information Networking and Applications Workshops (WAINA)*, March 2016, pp. 585–590.

[13] L. S. Beevi, G. Merlin, and G. MoganaPriya, "Security and privacy for smart grid using scalable key management," in *2016 International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT)*, March 2016, pp. 4716–4721.

[14] I. Parvez, A. Islam, and F. Kaleem, "A key management-based two-level encryption method for AMI," in *PES General Meeting — Conference Exposition, 2014 IEEE*, July 2014, pp. 1–5.

[15] P. Kumar, Y. Lin, G. Bai, A. Paverd, J. S. Dong, and A. Martin, "Smart grid metering networks: A survey on security, privacy and open research issues," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 3, pp. 2886–2927, 2019.

[16] A. Ghosal and M. Conti, "Key management systems for smart grid advanced metering infrastructure: A survey," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 3, pp. 2831–2848, 2019.

[17] N. I. Haque, M. H. Shahriar, M. G. Dastgir, A. Debnath, I. Parvez, A. Sarwat, and M. A. Rahman, "Machine learning in generation, detection, and mitigation of cyberattacks in smart grid: A survey," *arXiv preprint arXiv:2010.00661*, 2020.

[18] M. Ali, E. Al-Shaer, and Q. Duan, "Randomizing AMI configuration for proactive defense in smart grid," in *IEEE Int. Conf. on Smart Grid Communications (SmartGridComm)*, Oct 2013, pp. 618–623.

[19] B. Yuce, M. Mourshed, Y. Rezgui, and O. F. Rana, "Preserving prosumer privacy in a district level smart grid," in *2016 IEEE International Smart Cities Conference (ISC2)*, Sept 2016, pp. 1–6.

[20] C. Efthymiou and G. Kalogridis, "Smart Grid Privacy via Anonymization of Smart Metering Data," in *Proc. First IEEE Int. Conf. on Smart Grid Communications (SmartGridComm)*, Oct 2010, pp. 238–243.

[21] S. Tonyali, K. Akkaya, N. Saputro, and A. S. Uluagac, "A reliable data aggregation mechanism with Homomorphic Encryption in Smart Grid AMI networks," in *2016 13th IEEE Annual Consumer Communications Networking Conference (CCNC)*, Jan 2016, pp. 550–555.

[22] A. Abdallah and X. Shen, "A Lightweight Lattice-based Homomorphic Privacy-Preserving Data Aggregation Scheme for Smart Grid," *IEEE Transactions on Smart Grid*, vol. PP, no. 99, pp. 1–1, 2016.

[23] M. Kamal and M. Tariq, "Light-weight security and blockchain based provenance for advanced metering infrastructure," *IEEE Access*, vol. 7, pp. 87 345–87 356, 2019.

[24] L. Zhu, M. Li, Z. Zhang, C. Xu, R. Zhang, X. Du, and N. Guizani, "Privacy-preserving authentication and data aggregation for fog-based smart grid," *IEEE Communications Magazine*, vol. 57, no. 6, pp. 80–85, 2019.

[25] M. Benmalek, Y. Challal, A. Derhab, and A. Bouabdallah, "Versami: Versatile and scalable key management for smart grid ami systems," *Computer Networks*, vol. 132, pp. 161–179, 2018.

[26] S. Khasawneh and M. Kadoch, "Hybrid cryptography algorithm with precomputation for advanced metering infrastructure networks," *Mobile Networks and Applications*, vol. 23, pp. 982–993, 2018.

[27] L. Zhang, L. Zhao, S. Yin, C.-H. Chi, R. Liu, and Y. Zhang, "A lightweight authentication scheme with privacy protection for smart grid communications," *Future generation computer systems*, vol. 100, pp. 770–778, 2019.

[28] S. R. Bajekal, J. S. Arun, M. D. Nix, and K. K. Yellepeddy, "Event-driven, asset-centric key management in a smart grid," Aug. 24 2017, uS Patent App. 15/445,087.

[29] D. Wu and C. Zhou, "Fault-tolerant and scalable key management for smart grid," *IEEE Transactions on Smart Grid*, vol. 2, no. 2, pp. 375–381, June 2011.

[30] H. Nicanfar and V. C. M. Leung, "Smart grid multilayer consensus password-authenticated key exchange protocol," in *2012 IEEE International Conference on Communications (ICC)*, June 2012, pp. 6716–6720.

[31] Z. Sun and J. Ma, "Efficient key management for advanced distribution automation system," in *2010 2nd IEEE InternationalConference on Network Infrastructure and Digital Content*, Sept 2010, pp. 794–798.

[32] V. Sureshkumar, S. Anandhi, R. Amin, N. Selvarajan, and R. Madhumathi, "Design of robust mutual authentication and key establishment security protocol for cloud-enabled smart grid communication," *IEEE Systems Journal*, vol. 15, no. 3, pp. 3565–3572, 2020.

[33] H. Shekhawat and D. S. Gupta, "Quantum-defended lattice-based anonymous mutual authentication and key-exchange scheme for the smart-grid system," in *International Conference on Hybrid Intelligent Systems*. Springer, 2022, pp. 1132–1142.

[34] G. Nabeel, J. Zage, S. Kerr, E. Bertino, N. Kulatunga, and U. N. M. Duren, "Cryptographic Key Management for Smart Power Grids," CERIAS Technical Report, Tech. Rep., Februarry, 2012.

[35] H. So, S. Kwok, E. Lam, and K.-S. Lui, "Zero-Configuration Identity-Based Signcryption Scheme for Smart Grid," in *Proc. First IEEE Int. Conf. on Smart Grid Communications (SmartGridComm)*, Oct 2010, pp. 321–326.

[36] Y. Yan, Y. Qian, and H. Sharif, "A secure and reliable in-network collaborative communication scheme for advanced metering infrastructure in smart grid," in *Proc. IEEE Int. Conf. on Wireless Communications and Networking Conference (WCNC)*, March 2011, pp. 909–914.

[37] Q. Li and G. Cao, "Multicast authentication in the smart grid with one-time signature," *IEEE Transactions on Smart Grid*, vol. 2, no. 4, pp. 686–696, Dec 2011.

[38] T. B. Florian Skopika, Zhendong Maa and H. Grüneis, "A Survey on Threats and Vulnerabilities in Smart Metering Infrastructures ," *International Journal of Smart Grid and Clean Energy*, 2013.

[39] M. Kaveh and M. R. Mosavi, "A lightweight mutual authentication for smart grid neighborhood area network communications based on physically unclonable function," *IEEE Systems Journal*, vol. 14, no. 3, pp. 4535–4544, 2020.

[40] H. Nicanfar, P. Jokar, and V. C. M. Leung, "Smart grid authentication and key management for unicast and multicast communications," in *2011 IEEE PES Innovative Smart Grid Technologies*, Nov 2011, pp. 1–8.

[41] H. Ma, C. Wang, G. Xu, Q. Cao, G. Xu, and L. Duan, "Anonymous authentication protocol based on physical unclonable function and elliptic curve cryptography for smart grid," *IEEE Systems Journal*, 2023.

[42] D. Abbasinezhad-Mood, A. Ostad-Sharif, M. Nikooghadam, and S. M. Mazinani, "A secure and efficient key establishment scheme for communications of smart meters and service providers in smart grid," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 3, pp. 1495–1502, 2019.

[43] R. Bhatia and V. Bodade, "Defining the framework for wireless-AMI security in smart grid," in *Proc. Int. conf. on Green Computing Communication and Electrical Engineering (ICGCCEE)*, March 2014, pp. 1–5.

[44] M. Thomas, I. Ali, and N. Gupta, "A secure way of exchanging the secret keys in advanced metering infrastructure," in *Proc. IEEE Int. Conf. on Power System Technology (POWERCON)*, Oct 2012, pp. 1–7.

[45] P.-H. Hsu, W. Tang, C. Tsai, and B.-C. Cheng, "Two-Layer Security Scheme for AMI System in Taiwan," in *Proc. Ninth IEEE Int. Symp. on Parallel and Distributed Processing with Applications Workshops (ISPAW)*, May 2011, pp. 105–110.

[46] F. T. Liu, K. M. Ting, and Z.-H. Zhou, "Isolation-based anomaly detection," *ACM Trans. Knowl. Discov. Data*, vol. 6, no. 1, pp. 3:1–3:39, Mar. 2012.

[47] ——, "Isolation forest," in *Proc. IEEE Inter. Conf. on Data Mining*. IEEE, 2008, pp. 413–422.

[48] S. Wang, R. Inkol, and B. R. Jackson, "Relationship between the maximum likelihood emitter location estimators based on received signal strength (RSS) and received signal strength difference (RSSD)," in *2012 26th Biennial Symposium on Communications (QBSC)*, May 2012, pp. 64–69.

[49] I. Parvez, M. Jamei, A. Sundararajan, and A. Sarwat, "RSS based loop-free compass routing protocol for data communication in advanced metering infrastructure (AMI) of Smart Grid," in *Proc. on IEEE Sym. on Computational Intelligence Applications in Smart Grid (CIASG)*, Dec 2014, pp. 1–6.

[50] I. Parvez, A. I. Sarwat, L. Wei, and A. Sundararajan, "Securing metering infrastructure of smart grid: A machine learning and localization based key management approach," *Energies*, vol. 9, no. 9, 2016.

[51] B. Schölkopf, J. C. Platt, J. Shawe-Taylor, A. J. Smola, and R. C. Williamson, "Estimating the support of a high-dimensional distribution," *Neural computation*, vol. 13, no. 7, pp. 1443–1471, 2001.

[52] V. N. Vapnik and V. Vapnik, *Statistical learning theory*. Wiley Press, New York, 1998, vol. 1.

[53] B. Schölkopf, R. C. Williamson, A. J. Smola, J. Shawe-Taylor, and J. C. Platt, "Support Vector Method for Novelty Detection," in *NIPS*, vol. 12, 1999, pp. 582–588.

[54] P. Kumar, Y. Lin, G. Bai, A. Paverd, J. S. Dong, and A. Martin, "Smart grid metering networks: A survey on security, privacy and open research issues," *IEEE Communications Surveys Tutorials*, vol. 21, no. 3, pp. 2886–2927, 2019.

[55] S. Tufail, I. Parvez, S. Batool, and A. Sarwat, "A survey on cybersecurity challenges, detection, and mitigation techniques for the smart grid," *Energies*, vol. 14, no. 18, p. 5894, 2021.

[56] J. Markkula and J. Haapola, "LTE and hybrid sensor-LTE network performances in smart grid demand response scenarios," in *Proc. IEEE Conf. on Smart Grid Communications (SmartGridComm)*, Oct 2013, pp. 187–192.

[57] J. Kennedy and R. Eberhart, "Particle swarm optimization," 1995.

[58] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[59] K. A. Heller, K. M. Svore, A. D. Keromytis, and S. J. Stolfo, "One class support vector machines for detecting anomalous windows registry accesses," in *Proc. of the workshop on Data Mining for Computer Security*, vol. 9, 2003.

[60] T. Fawcett, "An introduction to ROC analysis," *Pattern recognition letters*, vol. 27, no. 8, pp. 861–874, 2006.