This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TII.2019.2957140, IEEE Transactions on Industrial Informatics

A Privacy-Preserving Framework based Blockchain and Deep Learning for Protecting Smart Power Networks

Marwa Keshk, Benjamin Turnbull, Nour Moustafa, Dinusha Vatsalan and Kim-Kwang Raymond Choo

Abstract—Modern power systems depend on Cyber-Physical Systems (CPSs) to link physical devices and control technologies. A major concern in the implementation of smart power networks is to minimize the risk of data privacy violation (e.g., by adversaries using data poisoning and inference attacks). In this paper, we propose a privacy-preserving framework to achieve both privacy and security in smart power networks. The framework includes two main modules, namely: a two-level privacy module and an anomaly detection module. In the two-level privacy module, an enhanced Proof of Work (ePoW) technique based blockchain is designed to verify data integrity and mitigate data poisoning attacks, and a Variational AutoEncoder (VAE) is simultaneously applied for transforming data into an encoded format for preventing inference attacks. In the anomaly detection module, a Long Short Term Memory (LSTM) deep learning technique is used for training and validating the outputs of the two-level privacy module using two public datasets. The results highlight that the proposed framework can efficiently protect data of smart power networks and discover abnormal behaviors, in comparison to several state-of-the-art techniques.

Index Terms—Privacy preservation, blockchain, Proof-of-Work (PoW), deep learning, anomaly detection, CPS

1 Introduction

The modernization of power systems is of extreme interest particularly to technologically advanced nations such as Australia and U.S., as smart grids have the capacity to optimize energy consumption and provide efficient solutions. Cyber-physical systems (CPSs) can also be combined to establish smart power networks that integrate physical and communication technologies and their elements to increase the efficiency of power systems [1]. Individual CPSs comprise cyber, physical and cyber-physical elements, where cyber elements are those with no direct contact with the physical world, physical elements are those with no direct contact with cyber elements, and the third category includes devices that link between cyber and physical elements [2]. CPSs often include Supervisory Control and Data Acquisition (SCADA) systems as remote interfaces for monitoring and managing operations of CPSs [3].

Since CPS interconnectivity of power elements and network devices at different power nodes increases the complexity of power grids and SCADA systems , large amount of data is also generated [4]. Such data can be utilized in network management or as a source for security monitoring, facilitates analysis of power measurements, and so on. The integration of cyber and physical elements with smart power networks, however, introduces an additional attack vector. It is known that CPS systems can potentially be exploited by adversaries, including advanced persistent threat (APT) actors, with potentially devastating physical consequences [2], [5]. There are two kinds of attacks, namely: physical and cyber. Physical attacks are those that involve the direct tampering of physical component while cyber-attacks are generally executed using malware or malicious software or by gaining access to the components of network systems [3], [6] .

Existing conventional security solutions may not be fit-for-purpose in a smart grid environment. For example, using secure end-to-end encryption techniques can disrupt analytical approaches and produce high false alarm rates [2]. There is a broad range of potential attacks targeting a smart grid, such as active attacks and passive attacks [3], [5]. In the latter category, the attackers attempt to sniff (private) data from the CPSs, for example by analyzing publicly available data; while in active attacks (e.g., data poisoning and inference attacks) the attackers have the capability to alter data [7]. In data poisoning attacks, for example, the attackers attempt to add or alter normal data. This can impact on the performance of training of machine learning-based data analytic or intrusion detection systems [6]. False Data Injection Attacks (FDIA) is a common type of data poisoning attack in power system networks [7], [8]. Thus, ensuring data integrity is a key requirement in the operations of grid systems and their networks [1], [4].

Not surprisingly, there have been a large body of work relating to data privacy protection and detection and identification of cyber-attacks [6], [9]. There are, however, a

- *M. Keshk is with the School of Engineering and Information Technology, University of New South Wales, Canberra Campus, Northcott Dr, Canberra, ACT 2600, Australia E-mail: marwa.hassan@student.adfa.edu.au.*
- *B. Turnbull, N.Moustafa are with the School of Engineering and Information Technology, University of New South Wales, Canberra Campus, Northcott Dr, Canberra, ACT 2600, Australia E-mail: benjamin.turnbull@unsw.edu.au, nour.moustafa@unsw.edu.au.*
- *D.Vatsalan is with the Information Security and Privacy Group, Data61-CSIRO, Eveleigh, NSW 2015, Australia E-mail: dinusha.vatsalan@data61.csiro.au*
- *K.-K. R. Choo is with the Department of Information Systems and Cyber Security, University of Texas at San Antonio, San Antonio, TX 78249-0631, USA E-mail: Raymond.choo@fulbrightmail.org.*

number of challenges that have yet to be addressed (see also Section 2), and hence, in this paper we present our proposed privacy-preserving framework. The latter is designed to achieve data privacy and facilitate attack detection in smart power networks. The main contributions of the study are as follows. First, two-level of privacy techniques are proposed. The first level includes the development of an enhanced Proof of Work (ePoW) technique based on blockchain for authenticating data records and preventing data poisoning attacks from altering original data. The second level contains a Variational AutoEncoder (VAE) that is used for converting original data into an encoded format for mitigating inference attacks that could learn from system-based machine learning. The datasets of ICS power systems [10] and the UNSW-NB15 [11] are used for validating the framework. Second, anomaly detection-based deep learning is applied for evaluating data before and after applying the two-level privacy techniques. We use the Long Short-Term Memory (LSTM) deep learning technique due to its efficiency of detecting anomalies from time-series data such as data of smart power networks [12].

The rest of this paper is structured as follows. Relevant background and related work are presented in Section 2. In Sections 3 and 4, we introduce our proposed framework and our evaluation approach. In Section 5, we present and discuss the evaluation findings. Finally, the concluding remarks are provided in the last section.

2 Background and Related Work

This section includes the concepts and previous studies related to privacy-preservation including blockchain technology, and intrusion detection including deep learning algorithms, used in modern power systems and their networks.

2.1 Privacy-preserving approaches

Privacy-preservation is defined as the process of protecting confidential data from disclosure by unauthorised users while processing it over networks [6], [9], [13]. Privacy-preserving approaches can be classified into five types; encryption-based [14], [12], perturbation-based [12], [15], authentication-based [16], Differential Privacy (DP) approaches [6], [17] and blockchain-based approaches [1], [18]. Each of these is discussed separately.

Privacy-preserving based on encryption approaches are used for encrypting data exchange using symmetric, asymmetric or homomorphic encryption methods [14]. Computing and analytic techniques demand unencrypted data, which could be manipulated by data poisoning or inference attacks. Moreover, despite advances in homomorphic encryption that allow simple arithmetic operations over encrypted data, no commercial applications currently use such techniques due to high computational resources and limited operational functions [6].

Privacy-preserving based perturbation approaches are used for transforming original data into new formats to conceal sensitive data using data transformation approaches, such as data projection and statistical measures [12]. The main challenge of these approaches is balancing privacy protection against data utility. Ideally, both are required, but these are also inverse and perfect privacy protection cannot co-exist with perfect data utility.

Privacy-preserving based on authentication approaches are utilised for providing user and system authentication mechanisms, such as single sign-on, federated identity and key management [16]. However, these approaches are not applicable to CPS protocols. Authentication-based techniques for privacy preservation cannot be applied to protecting data transferred over smart power networks due to the fact that it is not designed to operate on data, but only for authentication.

Differential Privacy (DP) approaches use efficient statistical methods, such as Gaussian and Laplace mechanisms to prevent inference and data poisoning attacks. DP approaches guarantee complete privacy as they have no assumption about the attacker's knowledge [6]. DP approaches ensure that the perturbed computations of specific data could not substantially alter when original data are updated [6], [19].

Privacy-preserving based on blockchain approaches apply the concept of blockchain, which is a peer-to-peer crypto connection to protect network nodes or data transactions [1]. The peers are from distributed networks, where every peer operates as a node of the network and can contribute to computing the solution to a hash-based puzzle problem confirming the transactions' integrity. Each transaction record is compressed as a block into the existing block chains. The recorded block contents are considered as a ledger. The entire blocks are synchronously updated to the entire network so that every peer retains a record of the same ledger [18]. Two popular miner techniques, namely, Proof of Work (PoW) and Proof of Stake (PoS) have been applied in Bitcoin and Ethereum, respectively, to verify the legitimacy of a transaction within blocks and to add new blocks [20]. To solve the puzzle, the PoW miner depends on the computation power while the PoS uses a deterministic algorithm that applies a hard fork to loose some blocks in some cases [21]. However, both techniques could be violated if a malicious miner has a computer power higher than 51% of the network; this is known as the 51% attack [21].

2.2 Intrusion Detection Systems

Intrusion Detection Systems (IDSs) have been widely used for monitoring and identifying intrusive activities from CPSs and their networks [6]. IDS approaches are categorised into three types: misuse-based, anomaly-based and hybrid of the two. A misuse-based IDS identifies only well-known attacks, while an anomaly-based IDS can effectively detect unknown attacks if its detection engine is well-designed to discriminate between normal and abnormal events [13]. From the perspective of power system networks, an IDS is an effective security system that can learn from transformed data generated at the control unit of power systems and unencrypted flow features of network traffic collected from smart power networks [6].

There are multiple methods used to develop an effective IDS. These include statistical learning, data mining, machine learning and deep learning [9]. Deep learning networks are broadly used for security applications such as malware

detection and IDSs, due to their capability to learn a computational process in depth. In intrusion detection, deep networks require information about legitimate data classes in the training phase to learn the weights of the network and obtain a model that can distinguish abnormal activities from normal ones. Deep learning includes generative and discriminative architectures. The generative architecture estimates joint probability distributions from observed data with their classes while discriminative architecture computes posterior distributions of classes conditioned on the observed data [16]. The LSTM is one of the useful generative models that can learn time series data of smart power networks [12].

2.3 Related studies

Privacy preserving and anomaly detection approaches are the main focus of this study, given that as the utility model that can effectively protect sensitive information and identify zero-day attacks from smart power networks. Several research studies have been carried out to assert data confidentiality and integrity in CPSs, along with applying a utility model to an IDS [7], [13], [16], [1], [12]. Lu et al. [4] proposed a privacy-preserving cosine similarity method to protect Big data in different systems such as power networks. Deng et al. [7] reviewed different methods and approaches for creating and preventing false data injection attacks in power systems. Gope et al. [22] proposed a privacy-preserving authentication scheme for securely analysing energy consumption between service providers and end users.

Gai et al. [13] proposed a blockchain-oriented approach to address the problem of privacy leakage in smart grids. In [16], a layer-wise perturbation and differential privacy based deep belief network techniques were developed for examining points of perturbation and accomplishing data privacy. Shen et al. [18] trained a support vector machine algorithm for detecting intrusive events from smart cities, along with authenticating data providers using a blockchain technique. More recently, Liang et al. [1] suggested a distributed blockchain-based framework for protecting modern power systems against cyber intrusions where blockchain technology was applied to power nodes. Their work was promising, but due to the tremendous computational resources required, it is unable to be directly applied to heterogeneous nodes found in smart power networks.

In our previous study [9], a Privacy-Preservation Intrusion Detection (PPID) mechanism was developed based on correlation coefficient and EM clustering algorithms for concealing confidential data and detect anomalous behaviours from power systems and their network data. Following that, in [6], a privacy-preserving anomaly detection framework was proposed for preserving sensitive data and detecting attacks from smart power networks. However, the two models cannot ensure data integrity against data poisoning and inference attacks while running anomaly detection systems. Consequently, this work addresses this problem through the use of two-level privacy methods using blockchain and variational autoencoder for differential privacy. For evaluating the reliability of the system utility, we then use LSTM as anomaly detection that has proven its capability of discovering abnormal events from time series data [12].

3 Proposed privacy-preserving framework

This section discusses the components of the proposed framework, including the proposed privacy-preserving and anomaly detection modules. The framework itself is outlined in Figure1.

The proposed methodology is comprised of input data collection, a two-level privacy preserving module, and an anomaly detection module using LSTM. The two-level privacy module is comprised of two components; the first level is a privacy-based blockchain, and the second level is a privacy-based Variational AutoEncoder (VAE). These components will be explained separately.

3.1 Privacy-preserving module

We propose a two-level privacy-preserving module for protecting data from disclosure while executing a utility system (e.g., anomaly detection) in smart power networks. The first level is the data integrity checker that uses the blockchain technology for authenticating data collections, and the second level includes the data transformation and generation model, which uses a VAE algorithm for converting original data into a new format that considerably prevents inference attacks from learning any information. The performance of the privacy-preserving model is evaluated using an anomaly detection system based on deep learning.

First level: Privacy-based blockchain

In the first level of the two-level privacy module, the blockchain technology has been applied to authenticate meter nodes of smart power networks. As previously outlined, blockchain uses encryption methods on each block inside a ledger [1], [13]. However, such methods take significant time and resources for processing while validating and mining every transactional record in a dataset. Instead, we propose utilising a blockchain construct for ensuring data integrity. The proposed method verifies the chains of data records that prevent data poisoning attacks from altering data records.

To describe the proposed method, let's assume that a dataset d includes records $R = \{r_1, r_2, ..., r_n\}$, where n is the number of records involved in d. Each record r contains a generated message digest that includes a secure hash function (SHA) which asserts data integrity. There are different SHA functions, such as $SHA265$ and $SHA512$, which are one-way cryptographic functions with different structures and fixed bit-lengths as output [18]. The reason for using the message digest to data blocks (i.e., records) is that hash functions are difficult to be inverted to find the input message using the corresponding output of the message digest. Any modification to data blocks will generate different message digests, which would occur in the event of a data poisoning attack [1].

To generate blocks in the data transaction context, each block includes an index, a timestamp, a hash value of the previous record (i.e., previous_hash), the current hash of the data record (i.e., new_hash) and proof, as described in Table 1. The hashes of data blocks are linked to each other (i.e., Hash Chain), as shown in Figure 2. We use the $SHA512$ function, as formulated in Equation 1, for generating hash values as, at time of writing, is it infeasible to be inverted in real-time processing using brute-force attacks [1].

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TII.2019.2957140, IEEE Transactions on Industrial Informatics

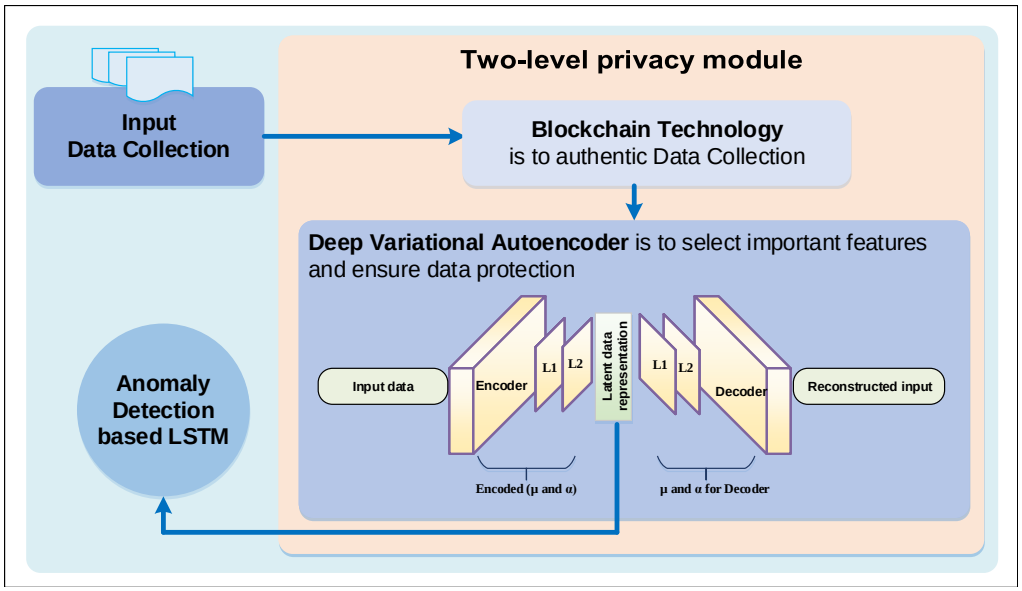

Figure 1. Proposed Deep Learning-based Blockchain Framework for protecting smart power networks

Table 1
Attributes of data blocks in the blockchain

Items	Meaning
Index	A sequent number of each block included in the ledger
Timestamp	The time generated while generated the message digest of a block
Data	Each record included in a dataset
Previous hash	The hash output of the previous block
Proof /nonce	The solution of the puzzle problem for the current block solved by a miner or problem solver
New hash	The hash result of the current block

It is essential in the blockchain to apply a consensus mechanism; a distributed rule set for the creation of new blocks and verifying the chain hash [21]. A puzzle problem is used to find a random number (nonce/proof) added to the hash value of a candidate block. This is shown in Equation 1. A problem solver or miner algorithm demands high computational power to find the proof of the candidate block. As depicted in Figure 2, the miner uses the proof of a particular block. For example, the miner uses the proof of Attribute 2 (99) in its mathematical hash function to find the linked hash of Attribute 2, which is located at Attribute 3.

$$Block - Hash = SHA512\,(data,\, proof) \qquad (1)$$

We propose an enhanced PoW (ePoW) miner algorithm that does not demand high computation power when compared with the classical PoW and PoS algorithms. This algorithm is presented in Figure 3. The algorithm includes two functions for creating hash blocks and estimating its proof based on the number of records in the dataset, where the number of records and proof are not equal zero, a proof of a new block can be created until the end of records included in the dataset is reached.

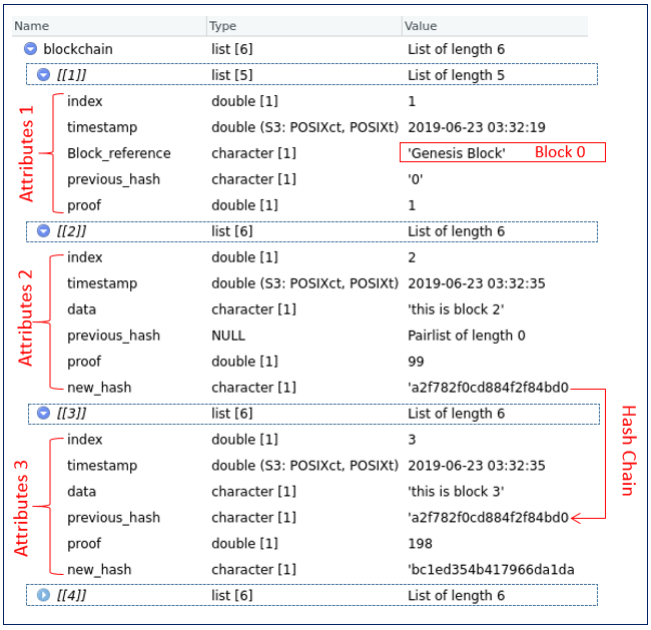

Figure 2. Privacy-preserving based blockchain module for ensuring the integrity of datasets

Second-level: Privacy-based Variational autoencoder (VAE)

While the ePow algorithm runs, the second level of privacy is applied to the attribute 'data' in Function 1, shown in Figure 3. This attribute refers to the datasets of smart power networks. In this level, we select important features for the purpose of securely training and validating a utility model. A Variational AutoEncoder (VAE) is used, which is a feed-forward model used for encoding an input X into new data codes using a set of weighted parameters [23]. The VAE is a generative model of classical autoencoder that can generate new samples from data based on their prior distribution.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TII.2019.2957140, IEEE Transactions on Industrial Informatics


```
# Function creates a hashed "block"
 Function creat_block (proof) {
   If (index = 0)
      previous_hash = 0
       hashed_block = digest (index, timestamp, data,
                          previous_hash, proof, new_hash,
                                    "sha512")
    return (hashed_block)
  }

# Function of an enhanced PoW (ePoW)
Function (last_proof) {
  proof <- last_proof + 1
  while (! (proof % n == 0 and proof % last_proof == 0))
    # n is the number of records
    proof <- proof + 1
  return (proof)
  }

# insert blocks to the chain
for (i in 1: n) {
    if (i == 1)
       proof =1  # Block 0
       add_block = Function creat_block ()
    else
      last_proof = i
      ePoW = Function (last_proof)
      add_block = Function creat_block (ePoW) }
```

Figure 3. An enhanced PoW (ePow) algorithm for authenticating data records

This technique considerably improves the utility system of anomaly detection. The model can achieve better protection against inference attacks when compared with classical and denoising autoencoder models that are deterministic for only learning the latent structure of features and do not generate new samples [24]. The VAE is trained to encode the input dataset d without the class labels into a hidden representation through a set of adapted weights; then the data codes are reconstructed using a set of generative weights, which are obtained from the latent data representation, as depicted in Figure 1.

The VAE is trained on a dataset d that includes$X$ features and n records, which are stochastically generated with a latent variable z. The stochastic process includes two key steps: prior and likelihood distributions. The prior distribution $p_\theta(z)$ estimates the hidden values of z_i. Then, the x_i observed value is generated using the conditional probability distribution $p_\theta(x \mid z)$. Supposing $p_\theta(z)$ is a prior distribution and $p_\theta(x \mid z)$ is a distribution that is estimated using the general parametric relations of $p_\theta(z)$ and $p_\theta(x \mid z)$ distributions. The marginal likelihood of the X features can be calculated using the sum of the likelihoods of data points, as given by

$$\log\, p_\theta(x_1, x_2, x_3 ....., x_N) = \sum_{i=1}^{N} \log\, p_\theta(x_i) \tag{2}$$

$$\log\, p_\theta(x_i) = D_{KL}(q_\varphi(z \mid x_i)\, K\, p_\theta(z \mid x_i)) + L(\theta, \varphi; x_i) \tag{3}$$

where $D_{KL}(q_\varphi\, K\, p_\theta)$denotes the Kullback–Leibler divergence [1] among the estimated posterior $q_\varphi(z \mid x_i)$ and the true posterior $p_\theta(z \mid x_i)$, and its value is a non-negative value, as estimated by:

$$D_{KL}(q_\varphi \parallel p_\theta) = \sum_{i=1}^{N} p_\theta(x_i)\, \log \frac{p_\theta(x_i)}{q_\varphi(x_i)} \tag{4}$$

$L(\theta, \varphi; x_i)$ refers to the variational lower bound on the marginal likelihood for each single data point i , as formulated by

$$\begin{aligned} L(\theta, \varphi; x_i) = &-D_{KL}(q_\varphi(z \mid x_i) \parallel p_\theta(z)) \\ &+E_{q_\varphi(z|x_i)}\left[\log\, p_\theta(x_i \mid z)\right] \end{aligned} \tag{5}$$

such that $q_\varphi(z \mid x_i)$ is fitted using a normal distribution, which estimates the true input distribution. The constructed latent distributions could reduce the probability of variations in the original input. The model is optimised using a $L2$ regularisation function to transform original data into a new data format [24]. The generated data cannot be violated using inference attacks because the model can generate new data samples using many potential distributions rather than using the original data.

3.2 Anomaly detection Module based LSTM

Once the data blocks are successfully verified using the ePoW algorithm, a Recurrent Neural network (RNN) algorithm is applied as a utility system of anomaly detection with retaining data privacy. The RNN is a deep learning algorithm for classifying a sequence of data. It is an extension of the conventional feed-forward neural network, but with recurring relations for better modelling. In this case, it is used to classify data from smart power networks. Assuming we have a sequence of input observations $X = (x_1, x_2, ..., x_T)$ for $t = 1$ to T, the RNN estimates the hidden and output vector sequences $H = (h_1, h_2, ..., h_T)$ and $Y = (y_1, y_2, ..., y_T)$, accordingly as follows [2]:

$$h_t = \sigma(W_{xh} x_t + W_{hh} h_{t-1} + b_h) \tag{6}$$

$$y_t = W_{hy} h_t + b_y \tag{7}$$

where σ is a nonlinear function, W denotes a weight matrix and b is a bias.

This work utilises Long Short Term Memory (LSTM) for classifying time-series data of smart power networks. LSTM has several advantages in this space. It has a complicated structure, which confers it to memorise information for a longer time, then use this information for prediction. It can also challenge the vanishing and exploding gradient problems through the use of three gates [3]. Figure 4 shows the connections in a single LSTM cell, where it has three gates (i.e., input gate i, forget gate f and output gate o) that control the information flow and a cell state c. For computing, the values of i,f,o and c, the following equations can be calculated,

$$i_t = \sigma(W_{xi} x_t + W_{hi} h_{t-1} + W_{ci} c_{t-1} + b_i) \tag{8}$$

$$f_t = \sigma(W_{xf} x_t + W_{hf} h_{t-1} + W_{cf} c_{t-1} + b_f) \tag{9}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TII.2019.2957140, IEEE Transactions on Industrial Informatics

$$c_t = f_t c_{t-1} + i_c \tanh(W_{xc} x_t + W_{hc} h_{t-1} + b_c) \quad (10)$$

$$h_t = o_t \tanh(c_t) \quad (11)$$

where σ is the logistic sigmoid function and W_{ci},W_{cf} and W_{co} refer to weight matrices for peephole connections. The i gate determines the proportion of input data, f gate decides to pass previous memory h_{t-1} or not and o gate determines whether the output can be passed or not. While calculating the cell state c, the input ratio has great effect.

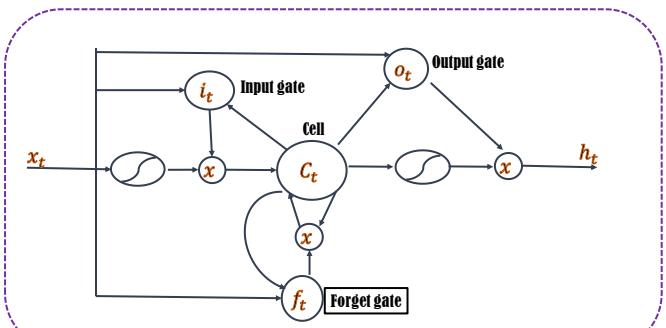

Figure 4. LSTM for anomaly detection

4 Datasets and Evaluation Metrics

To investigate the performance of the privacy-preserving framework for smart power networks, we used the datasets of Power System [10] and UNSW-NB15 [11], as the first includes control attributes of power systems while the second represents network attributes.

The Power system dataset is a multiclass dataset, involving 37 scenarios that include 8 natural events, 28 intrusive events and 1 no event. The UNSW-NB15 dataset includes a combination of current normal and attack records. The features published in our previous study [6] were used to validate the proposed framework and provide a fair comparison with recent peer techniques of privacy-preservation and anomaly detection. 300,000 random samples of legitimate and attack observations are chosen from each dataset for assessing the performance of the proposed framework. To evaluate the reliability of the proposed framework and other competing techniques without the associated bias due to the normal and abnormal classes in the datasets, the average from the the 5-fold cross-validation results is used as the performance.

As explained in [6], to evaluate privacy-preserving techniques, we apply the measures of the privacy level index (P_{index}), the dissimilarity level ($DISS$) and the information loss (IL). To evaluate the performance of anomaly detection before and after applying the privacy-preserving technique, we use $Accuracy$, False Alarm Rate (FAR), and accuracy vs. loss measures. For this work, the framework was developed in the 'R programming language' and are run on Ubuntu 18.04 LTS with a GPU Quadro P6000 and 32 GB RAM.

5 Experimental Results and Evaluations

5.1 The two-level privacy-preserving process

As per the proposed framework, the ePoW technique is applied as the first level of privacy to the datasets of power system and UNSW-NB15 for verifying data records and preventing data poisoning attacks from manipulating their records. Figure 5 represents a sample of the power system data to illustrate the ePow technique. The attributes of the blockchain were created for each record in the dataset, where the previous hash of index i should be equal to the new hash of index $i+1$ when the miner solves the proof.

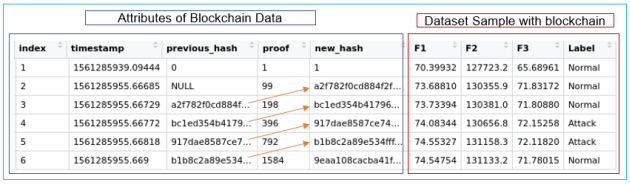

Figure 5. Example of applying the ePow technique for authenticating data records

While running the ePoW technique, the VAE technique is applied as the second level of privacy (i.e., differential privacy) on the eight features of each dataset published in [10], [11]. This is to protect data against inference attacks that can learn sensitive information about models. The hyper-parameters of executing the VAE technique on the two datasets are listed in Table 2. The technique includes 8 input features in the input layer, along with 2 hidden layers in the encoder and the decoder layers.

Table 2
Adopted VAE parameters

Settings	Hyper-parameters
Input layer	8 input features
Encoder	Two hidden layers: 1) **Layer 1**: 50 units, a 0.2 dropout rate, a Relu function 2) **Layer 2 (output)**: 2 units, a Softmax function
Decoder	Two hidden layers: 1) **Layer 1**: 50 units, a 0.2 dropout rate, a Relu function 2) **Layer 2 (output)**: 8 units of decoding the input features
VAE model	**loss**= 'binary_correntopy' , **optimizer**= 'adam', **epochs**=25, **batch_size**= 50, **metrics**= 'accuracy'

The results in terms of accuracy (acc) and $loss$ revealed that the VAE model could effectively learn from both datasets, as shown in Figures 6 and 7. For 25 epochs, the VAE technique achieves about 92.1% accuracy and 0.5% loss on the dataset of the power system, while the model accomplishes 99.8% accuracy and 0.01% loss for validating the UNSW-NB15 dataset. The aim of VAE is not for detecting attack vectors but transforming the data into a new shape that can be used for learning attack behaviours without degrading the system's performance.

The proposed two-level privacy techniques are compared with other four techniques using the three privacy

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TII.2019.2957140, IEEE Transactions on Industrial Informatics

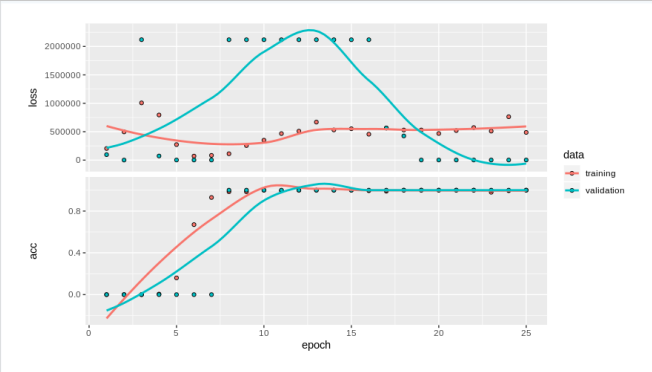

Figure 6. The accuracy (acc) vs the loss for training and validating the VAE using the power system dataset

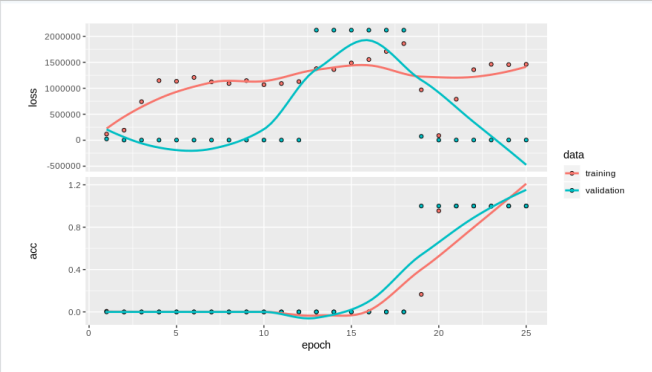

Figure 7. The accuracy (acc) vs the loss for training and validating the VAE using the UNSW_NB15 dataset

metrics of P_{index}, $DISS$ and IL on both datasets, as demonstrated in Table 3. The RDP technique [17] transforms data using a rotating noise perturbation, but it cannot entirely convert the original data. The PPFSCADA technique [25] divides the original data into vertical parts and then uses k-mean clusters to transform them. However, these parts are executed for all the features without ranking their significance in terms of defining attack vectors. The PCA-DR technique [15] replaces the original attributes by a small number of uncorrelated attributes. It is similar to the PPAD-CPS framework except that the latter improves privacy and security using multiple perturbation processes, and then applies an anomaly detection method to discover attacks using the permutated data. The proposed two-level techniques can protect data against altering data based on the ePoW technique that is not used in any technique before, and then protects against identifying sensitive information using the VAE technique. Therefore, it achieves the highest values for the three privacy metrics.

5.2 Anomaly detection based LSTM

The evaluation of the proposed two-level privacy model is also evaluated as a utility system of the anomaly detection based LSTM model. The parameters of LSTM is set as illustrated in Table 4. The LSTM model is applied to both datasets for classifying normal and attack classes on both datasets. The results, after applying the two-level privacy techniques, are promising. The model achieves approximately 95.2% accuracy and a 0.17% loss using the dataset of the power system. This is depicted in Figure 8. On the UNSW-NB15 dataset as shown in Figure 9, the results are slightly higher, achieving a 98.1% accuracy and a 0.19% loss while testing the model.

Table 3
Comparisons of the two-level privacy techniques with other four ones on both datasets.

Privacy method	**Power data**	**UNSW-NB15**	**Power data**	**UNSW-NB15**	**Power data**	**UNSW-NB15**
	P_{index}(%)		$DISS$ (%)		IL (%)	
RDP [17]	43.56	46.16	52.45	54.61	57.41	59.16
PPFSCADA [25]	51.33	58.89	49.73	58.75	52.14	56.43
PCA–DR [15]	57.78	62.34	67.29	69.87	72.19	74.50
PPAD-CPS [6]	67.43	81.35	68.87	73.52	75.20	79.16
Two-level privacy	**72.32**	**82.90**	**77.76**	**78.34**	**80.41**	**82.03**

Table 4
LSTM parameters for anomaly detection

Settings	**Hyper-parameters**
Input layer	encoded features resulted from the VAE model
Hidden layers	Two hidden layers: 1) **Layer 1**: 80 units, a tanh activation function 2) **Layer 2 (output)**: 40 units a tanh activation function
Output layer	2 units (normal/attack), a softmax activation function
LSTM model	loss= 'binary_correntopy' , optimizer= 'adam' batch_size= 50, epochs=100, metrics= 'accuracy' '

5.3 Comparisons and discussions

The LSTM model is compared with the other four techniques demonstrated in Table 5. The detection accuracy and False Alarm Rate (FAR) show that the anomaly detection technique outperforms others after applying the proposed two-level privacy models. On the dataset of the power system, the proposed anomaly detection-based LSTM technique achieves the highest accuracy (96.27%) and the lowest FAR (2.93%) that are close to the results of the PPAD-CPS framework [6] while the proposed technique achieves significantly better results (i.e., a 99.8% accuracy and a 0.01 FAR) compared with others on the UNSW-NB15 dataset. The LSTM technique is also applied to the two datasets before applying the two-level privacy module. The results revealed that the module can achieve optimum accuracy (100%) on both datasets. However, the results are better than other techniques while applying the proposed privacy-preserving modules which is the aim of the study for preventing data poisoning and inference attacks while applying a utility system such as anomaly detection.

We also evaluate the computational processing time of the framework with the other competing approaches, in order to determine how the time required by each approach to process data. Based on the findings, we determine that the proposed framework (i.e., LSTM-based privacy) requires

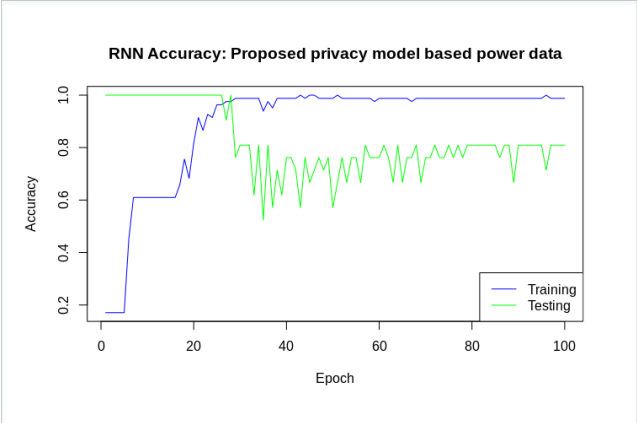

Figure 8. The accuracy of the Anomaly detection based RNN LSTM model after applying privacy-preservation on the dataset of power system.

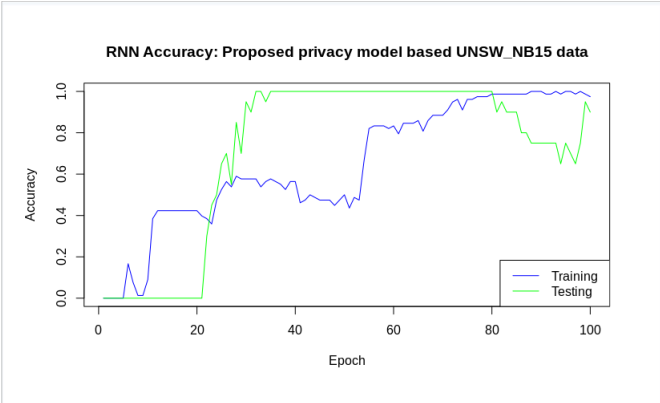

Figure 9. The accuracy of the Anomaly detection based RNN LSTM model after applying privacy-preservation on the power system dataset.

approximately 73 seconds to build a normal profile by training around 14,000 data observations. This performance is relatively close to that of the PPAD-CPS technique. However, the other three approaches require an average of 80 to 87 seconds, as shown in Table 5.

To explain why the proposed framework outperforms other techniques for preserving privacy and identifying attack behaviours, we consider several perspectives based on its potential design. The two-level privacy modules can achieve full protection by authenticating data transactions and transforming the original data into a new format for training and validating machine learning techniques. In the first-level privacy-based blockchain, the ePoW technique achieves the goal of verifying the data integrity, and if data poisoning attacks alter any record, the record can be easily discovered based on the proof and hash chain utilised in the ePoW technique.

The second-level privacy based the VAE technique can protect data by transforming it into another shape and using the encoded data for validating anomaly detection as an example of measuring the efficiency of detecting anomalies. Due to the high performances of deep learning in detecting attack behaviours, the LSTM technique is selected due to its norm of classifying time-series data such as datasets of smart power networks. The results highlight that the LSTM technique can efficiently classify legitimate and suspicious records after encoding the data using the two-level privacy-preserving techniques.

Table 5
Comparison of anomaly detection based LSTM with four techniques.

Techniques	Datasets				Time
	Power System		UNSW-NB15		
	DR (%)	FPR (%)	DR (%)	FPR (%)	Sec
NNR [26]	88.35	9.45	86.76	11.63	83
FSVM [27]	94.44	3.98	91.73	8.48	87
CART [28]	94.93	4.61	93.45	6.53	80
PPAD-CPS [6]	96.26	3.52	93.82	6.75	70
LSTM-based Privacy	**96.27**	**2.93**	**99.80**	**0.01**	73

6 Conclusion and Future Work

In this paper, we introduced a privacy-preserving framework based on blockchain and deep learning methods, in order to protect datasets of smart power networks and detect potential attacks. Our framework comprises two-level of privacy mechanisms. The first level includes an ePoW technique for verifying data integrity while the second level involves a VAE technique for encoding data and transforming it to a new format. The use of two levels of privacy achieves better performance compared with recent methods, and we demonstrated that it is also effective in preventing data poisoning and inference attacks from manipulating original datasets of smart power networks. An anomaly detection method based Long Short Term Memory (LSTM) was then evaluated on the datasets before and after applying the two-level privacy techniques. The results revealed that the technique can outperform other techniques in terms of accuracy and false alarm rate.

Future extension will include applying the framework on different real-world datasets from smart power networks, in order to evaluate its scalability and utility.

References

[1] G. Liang, S. R. Weller, F. Luo, J. Zhao, and Z. Y. Dong, "Distributed blockchain-based data protection framework for modern power systems against cyber attacks," *IEEE Transactions on Smart Grid*, 2018.

[2] J. Cortés, G. E. Dullerud, S. Han, J. Le Ny, S. Mitra, and G. J. Pappas, "Differential privacy in control and network systems," in *2016 IEEE 55th Conference on Decision and Control (CDC)*. IEEE, 2016, pp. 4252–4272.

[3] H. Song, G. A. Fink, and S. Jeschke, *Security and Privacy in Cyber-Physical Systems*. Wiley Online Library, 2017.

[4] R. Lu, H. Zhu, X. Liu, J. K. Liu, and J. Shao, "Toward efficient and privacy-preserving computing in big data era," *IEEE Network*, vol. 28, no. 4, pp. 46–50, 2014.

[5] B. Li, R. Lu, K. R. Choo, W. Wang, and S. Luo, "On reliability analysis of smart grids under topology attacks: A stochastic petri net approach," *TCPS*, vol. 3, no. 1, pp. 10:1–10:25, 2019.

[6] M. Keshk, E. Sitnikova, N. Moustafa, J. Hu, and I. Khalil, "An integrated framework for privacy-preserving based anomaly detection for cyber-physical systems," *IEEE Transactions on Sustainable Computing*, 2019.

[7] R. Deng, G. Xiao, R. Lu, H. Liang, and A. V. Vasilakos, "False data injection on state estimation in power systems—attacks, impacts, and defense: A survey," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 2, pp. 411–423, 2016.

[8] B. Li, R. Lu, W. Wang, and K. R. Choo, "Distributed host-based collaborative detection for false data injection attacks in smart grid cyber-physical system," *J. Parallel Distrib. Comput.*, vol. 103, pp. 32–41, 2017.

[9] M. Keshk, N. Moustafa, E. Sitnikova, and G. Creech, "Privacy preservation intrusion detection technique for scada systems," in *2017 Military Communications and Information Systems Conference (MilCIS)*. IEEE, 2017, pp. 1–6.

[10] "Power systems datasets," May 2017. [Online]. Available: https://sites.google.com/a/uah.edu/tommy-morris-uah/ics-data-sets

[11] N. Moustafa, B. Turnbull, and K.-K. R. Choo, "An ensemble intrusion detection technique based on proposed statistical flow features for protecting network traffic of internet of things," *IEEE Internet of Things Journal*, 2018.

[12] J. Lu and R. K. Wong, "Insider threat detection with long short-term memory," in *Proceedings of the Australasian Computer Science Week Multiconference*. ACM, 2019, p. 1.

[13] K. Gai, Y. Wu, L. Zhu, M. Qiu, and M. Shen, "Privacy-preserving energy trading using consortium blockchain in smart grid," *IEEE Transactions on Industrial Informatics*, 2019.

[14] M. S. Rahman, I. Khalil, A. Alabdulatif, and X. Yi, "Privacy preserving service selection using fully homomorphic encryption scheme on untrusted cloud service platform," *Knowledge-Based Systems*, 2019.

[15] R. V. Banu and N. Nagaveni, "Evaluation of a perturbation-based technique for privacy preservation in a multi-party clustering scenario," *Information Sciences*, vol. 232, pp. 437–448, 2013.

[16] T. A. Adesuyi and B. M. Kim, "A layer-wise perturbation based privacy preserving deep neural networks," in *2019 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*. IEEE, 2019, pp. 389–394.

[17] S. d. M. OLIVEIRA and O. R. Zaiane, "Privacy preserving clustering by data transformation." in *Embrapa Informática Agropecuária-Artigo em anais de congresso (ALICE)*. In: SIMPÓSIO BRASILEIRO DE BANCO DE DADOS, 18., 2003, Manaus. Anais . . . , 2010.

[18] M. Shen, X. Tang, L. Zhu, X. Du, and M. Guizani, "Privacy-preserving support vector machine training over blockchain-based encrypted iot data in smart cities," *IEEE Internet of Things Journal*, 2019.

[19] X. Liu, Z. Li, and Z. Li, "Impacts of bad data on the pmu based line outage detection," *arXiv preprint arXiv:1502.04236*, 2015.

[20] J. Huang, L. Kong, G. Chen, M.-Y. Wu, X. Liu, and P. Zeng, "Towards secure industrial iot: Blockchain system with credit-based consensus mechanism," *IEEE Transactions on Industrial Informatics*, 2019.

[21] D. Puthal, N. Malik, S. P. Mohanty, E. Kougianos, and C. Yang, "The blockchain as a decentralized security framework [future directions]," *IEEE Consumer Electronics Magazine*, vol. 7, no. 2, pp. 18–21, 2018.

[22] P. Gope and B. Sikdar, "An efficient privacy-preserving authentication scheme for energy internet-based vehicle-to-grid communication," *IEEE Transactions on Smart Grid*, 2019.

[23] Y. Burda, R. Grosse, and R. Salakhutdinov, "Importance weighted autoencoders," *arXiv preprint arXiv:1509.00519*, 2015.

[24] J. An and S. Cho, "Variational autoencoder based anomaly detection using reconstruction probability," *Special Lecture on IE*, vol. 2, pp. 1–18, 2015.

[25] A. Fahad, Z. Tari, A. Almalawi, A. Goscinski, I. Khalil, and A. Mahmood, "Ppfscada: Privacy preserving framework for scada data publishing," *Future generation computer systems*, vol. 37, pp. 496–511, 2014.

[26] R. A. R. Ashfaq, X.-Z. Wang, J. Z. Huang, H. Abbas, and Y.-L. He, "Fuzziness based semi-supervised learning approach for intrusion detection system," *Information Sciences*, vol. 378, pp. 484–497, 2017.

[27] M. A. Ambusaidi, X. He, P. Nanda, and Z. Tan, "Building an intrusion detection system using a filter-based feature selection algorithm," *IEEE transactions on computers*, vol. 65, no. 10, pp. 2986–2998, 2016.

[28] R. Petersen, "Data mining for network intrusion detection: A comparison of data mining algorithms and an analysis of relevant features for detecting cyber-attacks," 2015.

Marwa Keshk is a PhD candidate at UNSW of Canberra at the school of Engineering and Information Technology (Australian Centre for Cyber Security)and is a research candidate at Data61-CSIRO, Australia. Also, She received her Master degree in Computer Science from UNSW of Canberra in 2017. She completed her Bachelor degree of Computer Science in 2012 from the Faculty of Computer and Information, Helwan University, Egypt. Her areas of interests include Cyber Security, privacy preservation, Evolutionary Computation, Artificial intelligence techniques and Statistical methods.

Benjamin Turnbull is a Senior Lecturer at the University of New South Wales at the Australian Defence Force, Canberra. His research focuses on the intersection of cyber-security, simulation, scenario-based learning and the security of heterogeneous devices and future networks. He is also a Certified Information Systems Security Professional (CISSP). Ben has been working in digital forensics, network security and simulation for 17 years. His previous work as a defence research scientist saw him develop and deploy new technologies to multiple clients, globally.

Nour Moustafa is a Research Associate at UNSW's Australian Centre for Cyber Security. He received his Bachelor and Master degree of Computer Science in 2009 and 2014, respectively, from the Faculty of Computer and Information, Helwan University, Egypt. He completed his PhD degree in the field of Cyber Security from the University of New South Wales-Canberra in 2017. Nour started his academic career as an Assistant Lecturer at the Faculty of Computers and Information - Helwan University, Egypt in 2011. Besides his academic experience, Nour worked as a Senior Developer in the development field for developing .Net web and desktop applications for seven years. His areas of interests include Cyber Security, in particular, Network Security, host- and network- intrusion detection systems, statistics, Deep learning and machine learning techniques. He is interested in designing and developing threat detection and forensic mechanisms to the Industry 4.0 technology for identifying malicious activities from cloud computing, fog computing, Internet of Things (IoT) and industrial control systems over virtual machines and physical systems.

Dinusha Vatsalan is a Research Scientist at Data61-CSIRO, Australia, and an Honorary Lecturer in the Research School of Computer Science at the Australian National University. Her research interests are mainly in privacy preserving techniques, including privacy in data matching and mining, privacy in social media, privacy preserving couunting in stream data analytics, privacy risk evaluation and prediction, health informatics, and population informatics.

Kim-Kwang Raymond Choo (SM'15) received the Ph.D. degree in information security from Queensland University of Technology, Australia, in 2006. He currently holds the Cloud Technology Endowed Professorship at The University of Texas at San Antonio (UTSA). In 2016, he was named the Cybersecurity Educator of the Year - APAC (Cybersecurity Excellence Awards are produced in cooperation with the Information Security Community on LinkedIn), and in 2015 he and his team won the Digital Forensics Research Challenge organized by Germany's University of Erlangen-Nuremberg. He is the recipient of the 2019 IEEE Technical Committee on Scalable Computing (TCSC) Award for Excellence in Scalable Computing (Middle Career Researcher), 2018 UTSA College of Business Col. Jean Piccione and Lt. Col. Philip Piccione Endowed Research Award for Tenured Faculty, Outstanding Associate Editor of 2018 for IEEE Access, British Computer Society's 2019 Wilkes Award Runner-up, 2019 EURASIP Journal on Wireless Communications and Networking (JWCN) Best Paper Award, Korea Information Processing Society's Journal of Information Processing Systems (JIPS) Survey Paper Award (Gold) 2019, IEEE Blockchain 2019 Outstanding Paper Award, IEEE TrustCom 2018 Best Paper Award, ESORICS 2015 Best Research Paper Award, 2014 Highly Commended Award by the Australia New Zealand Policing Advisory Agency, Fulbright Scholarship in 2009, 2008 Australia Day Achievement Medallion, and British Computer Society's Wilkes Award in 2008. His research has been funded by the National Science Foundation, NASA, CPS Energy, LGS Innovations, Texas National Security Network Excellence Fund, Australian Government National Drug Law Enforcement Research Fund, Australian Government Cooperative Research Centre for Data to Decision, auDA Foundation, Government of South Australia, BAE Systems stratsec, Australasian Institute of Judicial Administration Incorporated, Australian Research Council, etc. He is also a Fellow of the Australian Computer Society, and Co-Chair of IEEE Multimedia Communications Technical Committee's Digital Rights Management for Multimedia Interest Group.