IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received February 12, 2021, accepted February 23, 2021, date of publication April 5, 2021, date of current version April 19, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3071141

Artificial Intelligence Based Intrusion Detection System for IEC 61850 Sampled Values Under Symmetric and Asymmetric Faults

TAHA SELIM USTUN[1], (Member, IEEE), S. M. SUHAIL HUSSAIN[2], (Member, IEEE), LEVENT YAVUZ[3], (Member, IEEE), AND AHMET ONEN[3], (Member, IEEE)

[1]Fukushima Renewable Energy Institute, AIST (FREA), National Institute of Advanced Industrial Science and Technology (AIST), Koriyama 963-0298, Japan
[2]Department of Computer Science, School of Computing, National University of Singapore (NUS), Singapore 119077
[3]Department of Electrical and Electronics Engineering, Abdullah Gül University, 38080 Kayseri, Turkey

Corresponding author: Taha Selim Ustun (ustun@ieee.org)

ABSTRACT Modern power systems require increased connectivity to implement novel coordination and control schemes. Wide-spread use of information technology in smartgrid domain is an outcome of this need. IEC 61850-based communication solutions have become popular due to a myriad of reasons. Object-oriented modeling capability, interoperable connectivity and strong communication protocols are to name a few. However, power system communication infrastructure is not well-equipped with cybersecurity mechanisms for safe operation. Unlike online banking systems that have been running such security systems for decades, smartgrid cybersecurity is an emerging field. A recent publication aimed at equipping IEC 61850-based communication with cybersecurity features, i.e. IEC 62351, only focuses on communication layer security. To achieve security at all levels, operational technology-based security is also needed. To address this need, this paper develops an intrusion detection system for smartgrids utilizing IEC 61850's Sampled Value (SV) messages. The system is developed with machine learning and is able to monitor communication traffic of a given power system and distinguish normal data measurements from falsely injected data, i.e. attacks. The designed system is implemented and tested with realistic IEC 61850 SV message dataset. Tests are performed on a Modified IEEE 14-bus system with renewable energy-based generators where different fault are applied. The results show that the proposed system can successfully distinguish normal power system events from cyberattacks with high accuracy. This ensures that smartgrids have intrusion detection in addition to cybersecurity features attached to exchanged messages.

INDEX TERMS Smartgrid cybersecurity, SV message security, IEC 62351, intrusion detection, artificial intelligence, IEEE 14-bus system, renewable energy.

I. INTRODUCTION

Integration of Information Technology (IT) with power systems gave birth to smartgrids [1]. In this fashion, more measurement can be done, and better operational decisions can be made. Power systems are operated more efficiently with smaller margins, in contrast to traditional procedures. Additionally, such connectivity enables novel applications that require coordination of more than one equipment in the system [2]. For instance, coordination of electric vehicles with renewable energy-based generators to mitigate their intermittency requires continuous message exchanges between these entities [3]. Alternatively, virtual power plant concept where different generation and storage devices act collectively to represent a much larger generation plant heavily relies on successful communication between these devices [4].

Adaptive protection is, especially, became a very active field [5]. Recent advances in power generation technology such as inverter-interfaced generators, novel storage devices and smart inverters created a much more dynamic power system [6]. Ensuring safe operation in such complex systems where there is bilateral power flow and many active components is a very tall task. Therefore, power system protection

The associate editor coordinating the review of this manuscript and approving it for publication was Gustavo Olague.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

strategies started employing more measurement and communication [7], [8]. This is the cost of having active devices at distribution network which effect the traditional power flow assumptions. Since novel protection schemes heavily rely on measurement values and other data exchanged over communication networks, they are also very prone to cyberattacks such as false data injection (FDI). Protection systems, also energy management systems, can be manipulated by injection false data [9]. Thinking that something drastic is happening in the power network, protection measures can be activated, or non-optimal energy dispatch decisions can be made. These will cause financial losses as well as power system interruptions.

As the enabling technology, IEC 61850 communication standard has emerged as the leader in this field due to several advantages [10]. It offers a robust structure that allows object-oriented modeling. Thanks to its standardized data object approach, interoperability is ensured regardless of device model or manufacturer. Finally, it has fully developed message exchange protocols that can be used for different purposes such as periodic message update or event-triggered messages [11]. Literature sees a constant influx of device and system modeling based on IEC 61850 standard and this is only expected to increase [12].

However, it has been reported in the literature that this high connectivity creates many cybersecurity vulnerabilities in smartgrids [13]. Until very recently, communication in power systems was utilized in very exclusive and limited contexts. It was not open to third-party connections and the possibility of an outside connection was minute. Therefore, cybersecurity measures that are well-known in other domains are currently being deployed in power systems for the first time. Recently published IEC 62351 standard aims at equipping IEC 61850 messages with cybersecurity features such as message integrity and encryption [14]. There are different studies that focus on how these two standards can be merged and secure IEC 61850 messages can be sent [15]–[18].

These IT measures are excellent towards securing message exchanges. However, holistic cybersecurity design requires that additional schemes are also implemented [19]. For instance, currently, IEC 62351 does not have any recommendation towards intrusion detection in smartgrids. Theoretically, if a hacker successfully penetrates the first line of defense set by IEC 62351 measures, there is no system in place to detect this intrusion. To address this need, this paper proposes a machine learning based intrusion detection for IEC 61850 Sampled Value (SV) messages. As the name implies, these messages carry periodic samples of critical grid parameters such as bus frequency, voltage etc. Due to critical nature of the places of their use, e.g. measurements for power system protection, frequency and voltage control as well as energy management, SV messages can be exploited to render significant damage on the power system infrastructure.

There are different works in the literature that focus on IEC 61850-based communication security. There are works that focus on implementation of IEC 62351 recommendations such as authentication and message integrity [20]. In addition, there are works that focus on extending these security measures and investigate possibility of using other algorithms or encryption [5]. Nevertheless, all of these works focus on developing a first line of defense against manipulations such as man-in-the-middle attacks, replay and masquerade attacks. Holistic cybersecurity defense approach requires there are different mechanisms to prevent, detect and divert an attack.

Although there are some intrusion detection systems proposed in the literature [21], [22], these works focus on Supervisory Control and Data Acquisition Systems (SCADA). The works in [23], [24] develop an IDS for IEC 61850 SV messages. However, they only make use of some information carried within SV messages. This provides IT level security. Thorough cybersecurity requires a solution using Operational Technology (OT), i.e. involving the operational parameters when the messages are exchanged. Recent work in [25] focuses on IEC 61850 SV messages and uses machine-learning approach to detect spoofed packets with OT. However, there is no accuracy and timing tests performed which makes the reliability and feasibility of the proposed work dubious. The individual plots shown in that work have very high error rates. This is due to the fact that in [25] neural networks are used which is not a good selection for power system protection applications. This is also confirmed in the results section of this paper where neural networks algorithm has the lowest accuracy rate. Finally, there is no test under fault conditions which makes it very hard to predict the power system behavior in an accurate manner.

Following from the above, currently, there is no mechanism for detecting intrusion in power system communication networks employing SV messages. To address this knowledge gap, this paper proposes an intrusion detection system for SV messages. Needless to say, power systems always have events that require different equipment to respond. However, this natural behavior is different than the behavior of an attacker who has acquired access to critical infrastructure and intends to do as much harm as possible. The system employs machine-learning and is trained to discern this natural behavior of a power system from cyberattacks.

The major contributions of this work are as follows:

(a) A novel machine-learning based intrusion detection system is developed for IEC 61850 SV messages.
(b) Modified IEEE 14-bus system is utilized to obtain a dataset that represents operation with renewable energy-based generators under normal and fault conditions. This data set is used to train the proposed system. Then, the performance of the system is tested with test data where cyberattacks are included.
(c) Symmetric and asymmetric fault conditions are added to the dataset. The system is trained and tested for normal and attack conditions both with and without faults. Symmetric and asymmetric faults are successfully distinguished from false data injected by the intrude.

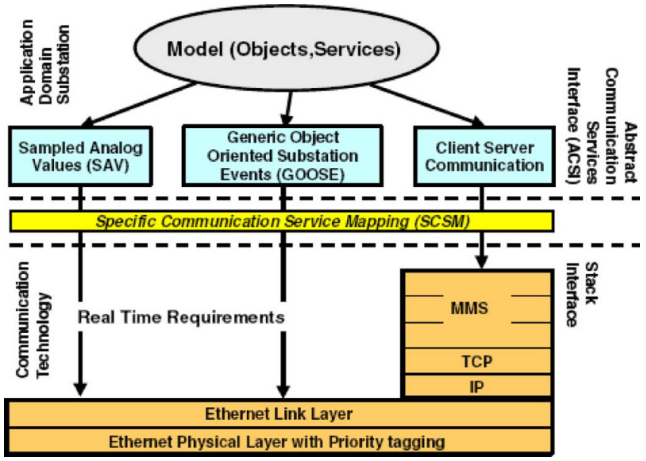

FIGURE 1. IEC 61850 communication stack.

(d) Different machine-learning algorithms are utilized, and their performances are contrasted. Results are reported to discuss which one of these algorithms is more suitable for intrusion detection in power system communication based on IEC 61850 SV messages. Evaluations are done in terms of training and attack detection times as well as attack detection accuracy.

The rest of the paper is organized as follows: Section 2 gives an overview of IEC 61850 SV messages, their structure and operation style. Section 3 presents the proposed intrusion detection algorithm. Section 4 gives details about performance experiments, sample data and test data. Finally, conclusions are drawn in Section 5.

II. IEC 61850 SV MESSAGES AND CYBERSECURITY VULNERABILITIES

IEC 61850 communication standard was initially developed to establish communication between substation devices [10]. However, it has received a lot of attention from researchers, engineers, and companies alike. Its initial domain is extended several times so that it can be used for power system communication with a much larger pool of available devices. Researchers have worked towards developing models for novel devices such as electric vehicles (EVs) [26] and smart meters [27] or new smartgrid applications such as virtual power plants [4], EV charging coordination schemes [3]. The main reasons behind such a positive uptake are object-oriented modeling that allows for simple yet strong device modeling, interoperable communication systems that do not depend on certain company or a technology as well as robust message exchange services that are developed for power system applications [28]. As shown in Figure 1, there are three services utilized. Generic Object-Oriented Substation Event (GOOSE) message is developed, as the name implies, as a means of exchanging information regarding an event that took place in the substation while Client-Server communication is used for ad-hoc message exchanges, notifications, and reporting.

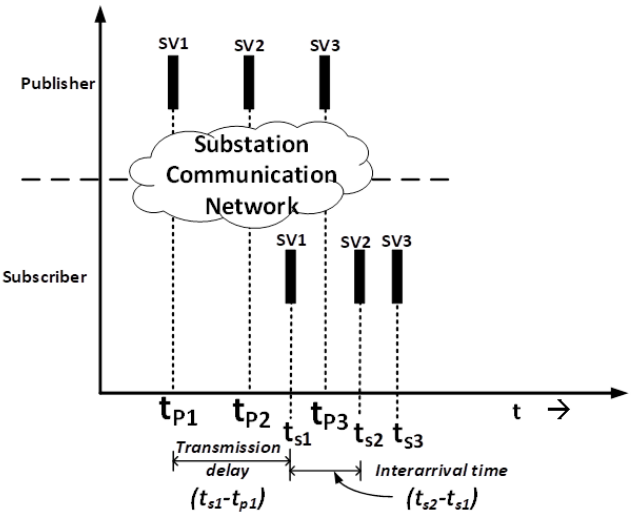

FIGURE 2. SV message exchange between a publisher and a subscriber.

As shown in Figure 2, Sample Value messages are used for periodic reporting of measurement values. Recently, their use has been extended to different smartgrid applications, yet the operational principles stayed the same. SV messages are preconfigured and, then, continuously transmitted from monitoring location to data processing point. Unlike GOOSE messages that are triggered when a predetermined event occurs, SV messages do not need a trigger to start transmission. Furthermore, they are designed to sample the target value with equal time windows. Therefore, the time period between two SV messages is always equal at publisher, although during transmission some delays may be introduced.

SV messages are traditionally used for substation protection devices. Due to high number of SV messages exchanged, handshaking procedures between sender and receiver are out of the question. This is not an issue in traditional substations that operate with exclusive communication networks that are not open to outside connection. However, recent advances in smart grid domain brought along novel uses of SV messages such as in demand side management or electric vehicle smart charging systems [3], [25]. The current structure of SV messages and the way in which these messages are transmitted have various cybersecurity vulnerabilities [29], [30]. The original use envisioned for these messages was limited to a proprietary substation that is not open to communication with the outside world. As the power system communication evolved and IEC 61850 standard is applied to information exchanged outside substation environment, these vulnerabilities became more apparent and relevant [31].

For instance, as shown in Figure 1, SV messages are directly mapped onto Ethernet layer, skipping TCP/IP, and making transmission much faster. However, the downside is that there is no traditional sender and receiver addresses that can be used to protect messages and prevent cyberattacks. It is true that the SV message structure as shown in Figure 3 includes destination and sources addresses, but

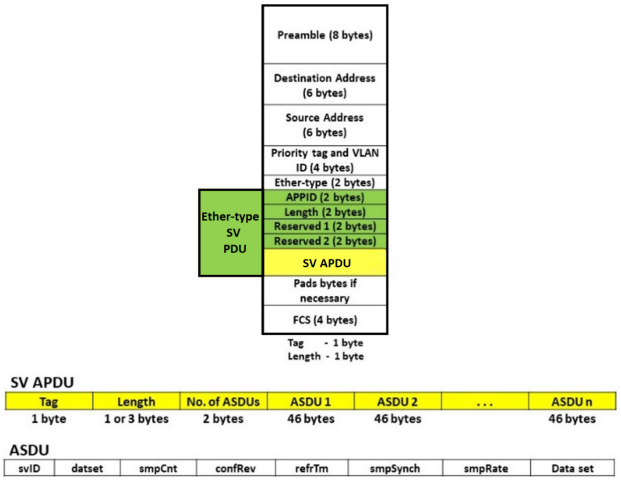

FIGURE 3. Message structure of SV.

TABLE 1. GOOSE and SV messages target address ranges.

Message type	Address Range
GOOSE	01:0C:CD:01:00:00 to 01:0C:CD:01:01:FF
SV	01:0C:CD:04:00:00 to 01:0C:CD:04:01:FF

these cannot be used for such purposes. The reason is that, firstly, these are Media Access Control (MAC) addresses and self-declared. Secondly, the destination address is not a real device's address. It is utilized to differentiate SV streams from each other and can take any value within the range specified in IEC 61850 standard as shown in Table 1.

Thirdly, these messages do not include any cybersecurity mechanism whether it be message integrity, authentication or encryption. They are exchanged over the net with full visibility and can be read and viewed by any party [32]. The messages do not have any built-in mechanism to authenticate the sender which leaves the doors open for any imposter attack [33]. Similarly, there is virtually nothing stopping an entity from capturing an SV message exchanged in the network, editing its contents and retransmitting it as a part of a replay or masquerade attack [33]. Some of these issues are identified and IEC 62351 Cybersecurity standard has been issued as a complementary to IEC 61850 communication standard. The proposed cybersecurity mechanisms are still in their infancy and require a lot of work to be widely implemented in power system communication infrastructure.

Nevertheless, IEC 62351 cybersecurity standard only recommends use of communication layer security mechanisms, such as implementing hash algorithms to check message integrity or using digital signatures to authenticate senders. There is no input on operational layer security. To ensure fully secure communication, a holistic cybersecurity approach is needed. For instance, if a hacker circumvents the security checks implemented at communication layer and gains access to the network, there is no system in place to detect this breach. Considering the sensitive nature of SV message contents and that they are used to trigger actions in devices, this is a big problem. They can be utilized by parties with malicious intent to inject false data into the system with the aim of disrupting desired operation. In order to fill this knowledge gap, a machine-learning based intrusion detection algorithm is developed in the next section.

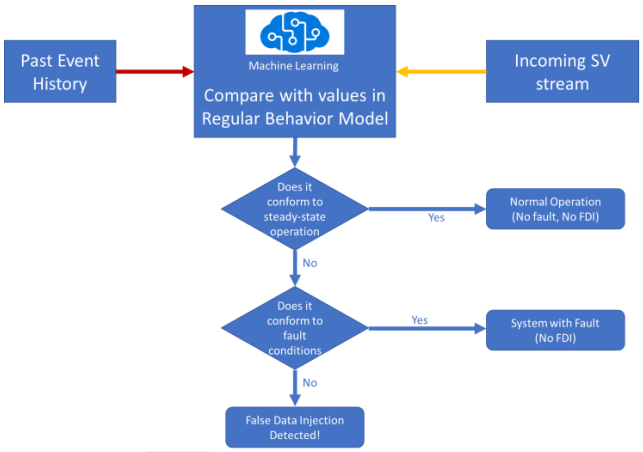

FIGURE 4. Machine learning-based intrusion detection system for SV messages.

III. MACHINE-LEARNING BASED INTRUSION DETECTION ALGORITHM

SV messages are giving a snapshot of the entire power system by continuously sampling system parameters and sending them to control centers. These system parameters change with respect to events taking place in network such as load increase, generation loss or faults. However, these changes are not arbitrary and depend on the system topology, operating conditions and components. In other words, similar events under similar conditions should trigger reactions that resemble each other. Furthermore, an event at a certain location, e.g. generation loss or load increase at a certain bus, not only effects SV messages sent from that particular location but from neighboring sampling sites as well. Therefore, there is a certain behavioral pattern that every power system has for a particular event. This can be observed, reported and detected. On the other hand, hackers launch cyberattacks from a point where they can breach and gain unauthorized access. It is inconceivable that a hacker gains access to all measurement devices in a network. Therefore, when a hacker launches FDI attack at the compromised access node, incoming SV values would be inconsistent with the rest of the SV streams present in the network. This discrepancy shows that there is an intruder in the system who is trying to inflict some damage with falsely injected data.

Based on these facts, it is possible to design an intrusion detection system as shown in Figure 4. SV streams are constantly supplied for measurement and control purposes. In parallel with power system operation, an event analysis is performed for these streams. Based on the event history, i.e. previous events, SV streams are subjected to scrutiny and compared with the regular behavior of the power system. If the event history shows that this event is likely to be a legitimate event, then the normal operation continues.

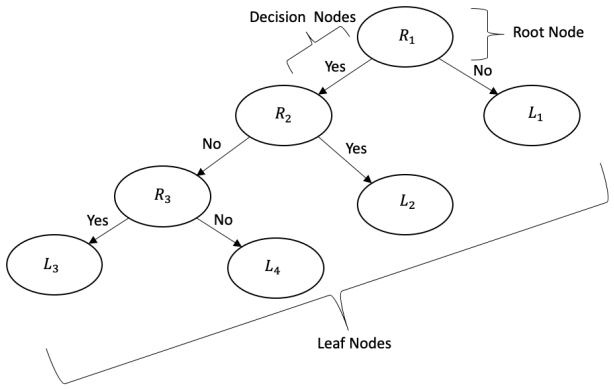

FIGURE 5. Decision tree operation structure.

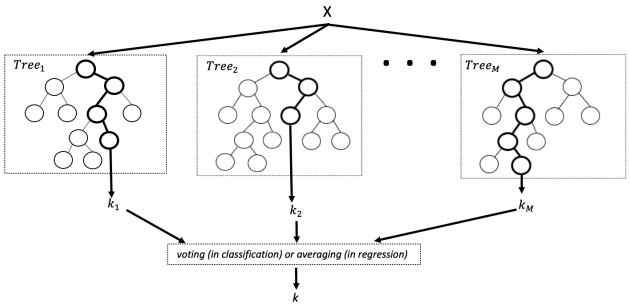

FIGURE 6. A random forest with several decision trees.

Otherwise, the accumulating evidence indicates that there is an intruder in the system and the alarm is raised.

It goes without saying that every power system, or sub-system such as a microgrid or a sub-station, has different behavior. Therefore, comparison performed in Figure 4 needs to be particular to each system, not generic. This requires analyzing the past events and developing a behavior model as shown in the middle. Machine learning is utilized to develop a pattern for any given power system. To add to diversity, machine learning algorithm is trained with data pertaining to system with and without fault conditions. If there is a fault, these are also not trivial three-phase to ground faults that are mostly used in power flow simulations. In real life, asymmetric faults are more common than symmetric ones. Therefore, the algorithm is trained with asymmetric fault data as well as symmetric fault data. As a result, the developed detection system can distinguish between no fault operation, fault operation and FDI attack situations. In case of a fault, system can also detect what type of a fault is observed.

Several machine learning algorithms are utilized, and their performances are compared in the next section. Before getting into test results and their analysis, an overview of these algorithms is given in the next sub-section.

A. DIFFERENT MACHINE LEARNING ALGORITHMS UTILIZED

In order to measure success of prediction and contrast their performances, several algorithms are utilized in the proposed system. These are Decision Tree (DT), Random Forest (RF), Extremely Randomized Trees (XRT) and Artificial Neural Network (ANN) algorithms.

Shown in Figure 5, DT algorithm utilizes decision tress with branches and leaves. In this fashion, it extracts conclusions from observations related to a particular item. In this approach, observations are represented as branches while the conclusions are the leaves. The algorithm is designed to progress towards the leaves. Since the goal of DT is to draw some conclusions and estimate the value of a target node, it is deemed suitable for the developed intrusion detection system where values for SV messages are estimated in a broad sense.

A collection of DTs constitutes a RF. In other words, RFs utilize several DTs to make a decision and individual decisions from each DT are processed to reach a final conclusion in RF, as shown Figure 6. Decisions are made by following the most efficient path in each DT. RF is a bagging algorithm and it can be utilized to address over-fitting or accuracy issues encountered in DTs. The number of DTs is not limited and can be set as wished. In this particular study, 100 trees are used in RF.

XRT is similar to RF in the sense that an ensemble of trees is used instead of a single one. Certainly, there are differences between the two. Firstly, XRT uses the entire sampling data while RF allows for bootstrap data use. The more structural difference is that RF chooses the optimum cut points to split nodes whereas XRT does this randomly. XRT selects the best cut point among the randomly split nodes. This approach makes XRT both randomized and optimized.

In some XRT and RF implementations, bagging and boosting ensemble methods are utilized. This stems from the principle that many learners perform better than a single one. By combining them, it is possible to create a better learning algorithm. As shown in Figure 7, the approach is very similar for these two methods. Individual weak learners are trained, and a collective decision is made to reach the output. Only difference is that bagging trains the learners separately while boosting does this sequentially. Output of bagging relies on the mean prediction and cannot give precise values for the model. But it can reduce over-fitting and maintain accuracy for data that is not available. On the other end of the spectrum, boosting increases accuracy but is prone to overfitting. An approach where these two are combined can be used to attain high accuracy and avoid overfitting. In this study, XRT uses bagging mechanism as it is more suitable for FDI attack on power system values. Boosting may fail in this application since it is more fit for cases where data add on each other.

ANN models biological neural networks in an artificial way; hence the name. It utilizes artificial neurons that process the input and provide an output. All the processing is done inside the algorithm called hidden layer which can include only a single set of neurons or more. The non-linear functions embedded inside the neurons help model a certain behavior and create a relationship between input and output data. Considering that the proposed intrusion detection system is also

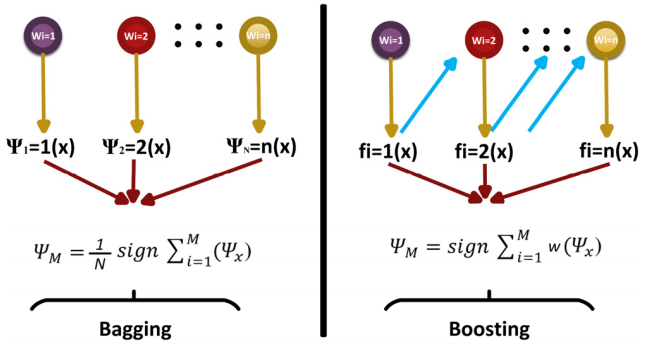

FIGURE 7. Bagging and boosting methods used in XRT.

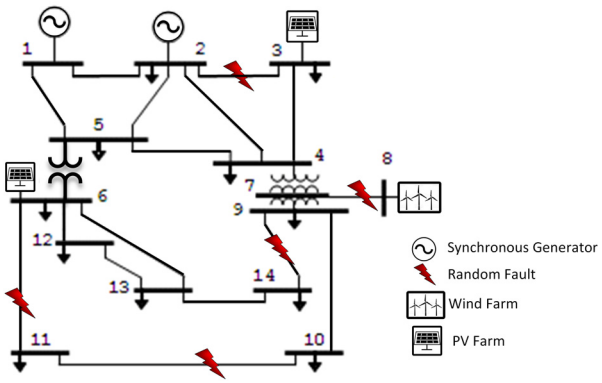

FIGURE 8. Modified 14-bus system with renewable energy based generators.

trying to achieve this between SV stream data and operating conditions, ANN is selected as one of the tested machine learning algorithms.

Next section presents the training data, test data and the test results for all the algorithms discussed above.

IV. INTRUSION DETECTION PERFORMANCE TESTS

In order to investigate the accuracy of the proposed intrusion detection system for SV messages, several tests have been performed. As shown in Figure 8, a modified IEEE 14-bus system is obtained by deploying some renewable energy-based generators. These distributed generators create a more dynamic power system operation profile and makes it more difficult to successfully predict accurately. Considering the current trends for increased renewable energy share in generation portfolio, such systems will become more common in the future. Therefore, the tests are performed on this difficult yet more realistic topology.

Firstly, the system is run under normal conditions without any faults or FDI attacks. System parameters are sampled so that a normal operation model can be constructed. Then, random faults are applied to one of the five designated fault locations shown in Figure 8. There is only one fault at a single location, if a fault is applied. Since symmetric faults account for only 2-5 % of all faults in the grid [34], asymmetric faults are also considered. The fault type is randomly selected from any one of the following:

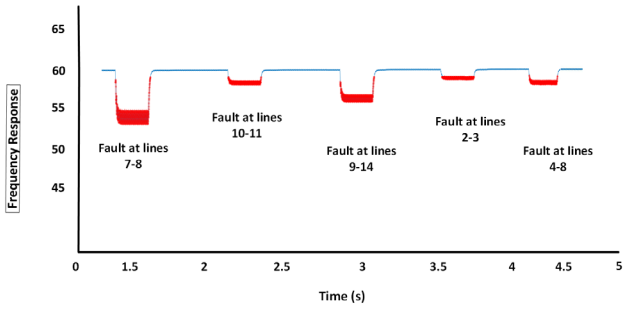

FIGURE 9. Frequency at bus 9 for LG fault at different locations.

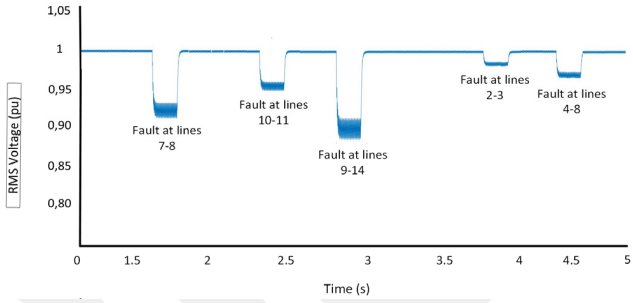

FIGURE 10. Voltage at bus 9 for LG fault at different locations.

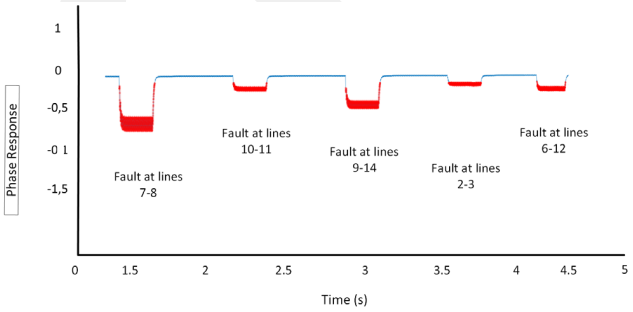

FIGURE 11. Voltage phase angle at bus 9 for LG fault at different locations.

1. Three phase line to ground fault (3L-G, symmetric)
2. Line to ground fault (LG, asymmetric)
3. Line to Line fault (LL, asymmetric)
4. Double line to ground (2L-G, asymmetric)

Data collected under these conditions represent the system behavior with different faults at various locations. Figures 9-11 depict impact of LG fault at different locations on frequency, voltage value and phase angle of bus 9. It can be observed that the same event has different repercussions at different measurement locations.

Once this aggregated dataset is acquired, FDI attack data is injected to replicate a scenario where a hacker has gained unauthorized access to the system. In that case, a hacking algorithm has been developed to find a gap and crack the security system. However, this is not blind random value injection. The developed attacking algorithm takes original data and injects the hacked data based on this real value. Then, based on the output of ML-based detection system, the attack algorithm tweaks this attack data. If it is easily detected,

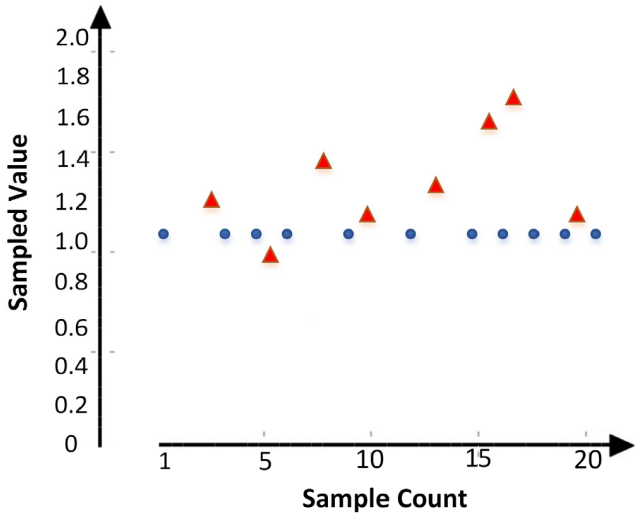

FIGURE 12. Random value FDI attack on voltage values in SV values.

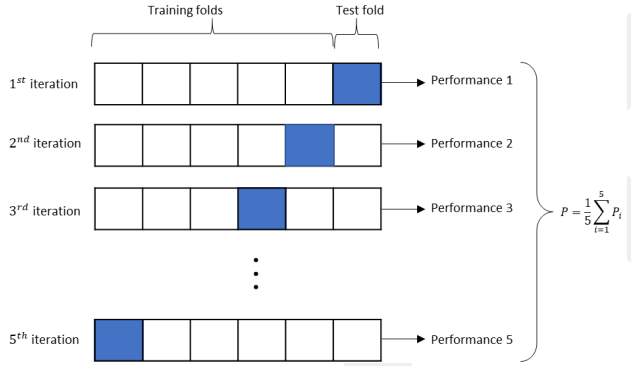

FIGURE 13. Cross validation approach with several iterations.

then a random value closer to *real value* is injected. This is repeated until the hacked data can pass through detection. This trains the detection algorithm and increases its prediction success. In order to evaluate this performance, one of the three system parameters (bus voltage, phase angle and frequency) is selected at random, and random value FDI attack is applied as explained above and depicted in Figure 12.

This final dataset has normal operation, fault operation and FDI attack data inside. The algorithms are trained and tested with this dataset. In order to increase the reliability of test results cross validation is performed with a cross validation value of 5. This means executing five distinct iterations on the dataset. This PSCAD-Python co-simulation creates 21600 data points. Almost %10 of all columns has been attacked on different time zones and different bus locations. However, ML algorithms tend to memorize the dataset instead of learning and predicting process. This is called overfitting. In order to avoid this problem, cross validation method has been applied as shown in Figure 13.

As shown, the overall data is split into seven equal portions. In each iteration a different portion is designated as the test fold while the rest is used as training data. The benefit of this approach is that it mixes the training and test folds over the

TABLE 2. Accuracy precision and F1 scores for different FDI attack and fault combinations.

		Accuracy (%)			
FDI Attack	**Fault**	**ANN**	**DT**	**RF**	**XRT**
No	Yes	85.31	86.13	94.27	97.73
Yes	No	89.12	85.23	92.22	97.13
Yes	Yes	83.17	82.79	90.37	94.69
		Precision (%)			
FDI Attack	**Fault**	**ANN**	**DT**	**RF**	**XRT**
No	Yes	78.15	81.13	83.71	92.74
Yes	No	80.23	85.23	86.54	91.37
Yes	Yes	75.83	83.13	85.79	90.45
		F1 Score (%)			
FDI Attack	**Fault**	**ANN**	**DT**	**RF**	**XRT**
No	Yes	80.41	88.73	90.79	91.73
Yes	No	78.25	86.85	89.20	90.13
Yes	Yes	77.17	84.89	90.37	90.69

entire dataset. This eliminates the possibility of any lucky situations that may arise from a specific way of splitting the dataset. Every portion gets to be utilized as a test fold, thereby subjecting the proposed intrusion detection system to all possible combinations.

Performance tests have been performed in Python on a platform with Intel Core i7 @ 2.80 GHz with 32 GB RAM and the results are reported in Table 2 and 3. Firstly, it is safe to say that the proposed intrusion detection system is validated with these results. Regardless of the machine learning algorithm used, the system distinguishes regular operation from cyberattacks regardless of there being a fault in the system. The accuracy of the used algorithms has been a mixed bag where ANN and DT had about 85 % accuracy for no attack situations. This value dropped all the way down to around 83 % where there are attacks and faults simultaneously. On the other hand, RF and XRT reported very high accuracy values where XRT had 97.7 % and 94.7 % for no attack case and the case where both attacks and faults are present, respectively.

Precision and F1 Scores are also shown in Table 2. These results are exactly as expected. For example, if the model marks the signals that need to come to monitoring system as attack (false positive = FP), the power system will always raise an alarm. However, filtering the correct it is important is correct signal must be filtered. In this case, the high/low precision value is an important criterion for choosing the model as presented in Figure 14. Receiver Operating Characteristic(ROC) curve has been created for binary classification and shows that the accuracy is the best option.

The poor performance of ANN is due to the nature of the implementation. The input values are not complex, power system parameter readings sent by SV values, and outputs are pretty straightforward. ANN is more suitable for more stochastic and complex applications such as image processing. DT is a very simple machine-learning algorithm and its performance is similar to that of ANN. RF and XRT, on the

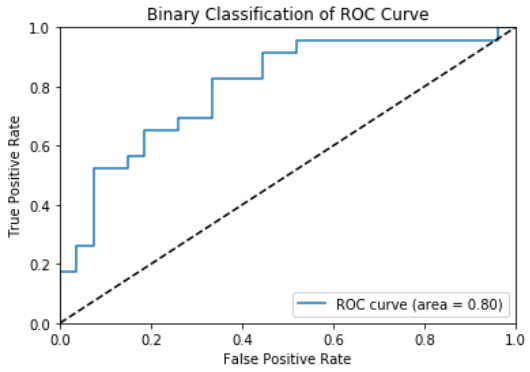

FIGURE 14. ROC curve.

TABLE 3. Training and FDI detection times of different algorithms.

	FDI Attack	Fault	Time Performance (sec)			
			ANN	DT	RF	XRT
Test conditions (3 inputs)	Yes	Yes	0.318 $x10^{-3}$	0.049 $x10^{-3}$	0.213$x10^{-3}$	0.105$x10^{-3}$
Training	Yes	Yes	1.06	0.231	0.92	0.451

other hand, are improved implementations of DT and this can be clearly observed from the reported high accuracy values. XRT's randomized node splitting has given it an edge over RF where this is done optimally.

Timing performances are reported in Table 3 for the case where both FDI attack and fault are present. Most important aspect of the test results is the times required to train the system, i.e. training time, and the time required to run the algorithm and detect an attack, i.e. attack detection time. These two times are completely distinct and relate to different steps of operation. Training can be done offline or before the deployment of the system. Therefore, it does not have a direct impact on the system operation when SV messages are received in real-time. On the other hand, attack detection time pertains to real-time operation of the proposed system. It corresponds to the time it takes for the system to process an incoming SV message and decide whether it is a normal message or an attack message, as shown in Figure 4.

IEC 61850 standard does not stipulate a certain time required for SV messages to be sent. That's because the sampling rate may differ from application to application. In cases where SV is used for slowly changing values, e.g. state of charge in EV battery, sampling rate can be one in several minutes [35]. However, this arbitrary definition of SV messages has been hard to implement in the field. IEC 61850 users' group has agreed on a lighter edition of SVs, IEC 61850-9-2LE, where the sampling rate is limited to two choices: 80 and 256 samples per cycle [36], [37]. For a 60 Hz system this means the time between two SV messages is 208 μs for 80 samples and 65.1 μs for 256 samples.

Analyzing the performance test data in Table 3, it can be observed that DT can be used for intrusion detection in a system running 9-2LE SV messages. For 80 samples, there is ample time for processing the incoming SV messages in real-time for a possible attack. If 256 samples are used, the window is much smaller but still the processing time is less than separation between the messages. ANN and RF cannot be used for either 9-2LE sampling rate in real-time. XRT, which has the highest accuracy rate, can be used for 9-2LE SV messages if sampling rate is 80. In this case, XRT needs half of the time required for an SV message to arrive before the next SV message comes. This is the best combination of accuracy and speed for the tested cases.

It is also possible that attack detection can run in real-time as SV messages arrive and every n^{th} SV message is subject to scrutiny. For instance, if RF is used for 9-2LE SV message with 256 samples, it can process every 5^{th} SV message to check for possible FDI attack and intrusion. When and if any attack is detected, alarm can be raised, and SV stream is blocked. Only a couple messages will slip through until an attack is detected. This approach is especially important where the users may opt to use a much higher sampling rate than 9-2LE, at least in theory IEC 61850-9-2 (not LE) allows that. In such cases, the proposed intrusion detection system can be used in an asynchronous fashion. Still, any FDI attack can be successfully detected and system operation can be secured.

Finally, all of the algorithms have relatively short training times, considering that training is done offline. This opens a path to pseudo-online training approach where the system may collect data and retrain itself on a specific time window, e.g. 1 month or 3 months. This will add value to the proposed system as it can learn the changing behavior of the power system and adjust its training. This will create a much more dynamic intrusion detection system that can respond to changing trends in the power system.

Test results show that DT and XRT have much smaller training and detection times. However, DT's accuracy is low whereas XRT reported excellent accuracy values. Therefore, XRT can be deemed as the most suitable algorithm for proposed intrusion detection system since it offers the best combination of higher accuracy and less time required.

V. CONCLUSION

Smartgrid applications are getting more popular where different devices need to communicate and coordinate. For this to happen, a reliable infrastructure is needed. There have been efforts towards providing an interoperable communication platform for such purposes. However, implementation of cybersecurity mechanisms to secure information exchange on such large-scale has lagged behind. There is imminent need for achieving cybersecurity in power system, a cyber-physical system where message exchanges may have real, physical implications.

IEC 61850's SV messages are widely used for sampling and reporting system parameters. This makes them highly critical in cybersecurity assessments. This paper develops a machine learning-based intrusion detection system for SV messages. Based on the nature of SV messages, the system

is able to differentiate between *usual operation* from *attacks*. The resilience of the system is high since it is not affected by the presence of a fault current in the system. Furthermore, the system is able to distinguish between a symmetrical and asymmetrical fault as well. Performance tests have been performed with realistic smartgrid communication dataset. Different machine-learning algorithms are utilized to see their suitability for such use. Results show that the developed system can successfully detect cyber-attacks based on SV messages. Although the performance of algorithms differs, all machine-learning algorithms yield acceptable results and no over-fitting is observed.

Using algorithms other than the ones in this paper or using different parameter values can be a future extension of this work. Nevertheless, the current results show that the proposed intrusion detection system can successfully detect unauthorized access via SV message analysis. Future work may focus on integrating this system with a honeypot.

REFERENCES

[1] Communication from The Commission To The European Parliament, The Council, The European Economic and Social Committee and The Committee of The Regions. *Smart Grids: From Innovation to Deployment*, document 52011DC0202, Brussels, 2011. Accessed: Dec. 18, 2020. [Online]. Available: https://bit.ly/3r7Zv8T

[2] F. Gangale, J. Vasiljevska, F. Covrig, A. Mengolini, and G. Fulli, *Smart Grid Projects Outlook 2017: Facts, Figures and Trends in Europe*, Standard EUR 28614 EN, 2017, doi: 10.2760/701587.

[3] T. S. Ustun, S. M. S. Hussain, and H. Kikusato, "IEC 61850-based communication modeling of EV charge-discharge management for maximum PV generation," *IEEE Access*, vol. 7, pp. 4219–4231, 2019.

[4] F. Nadeem, M. A. Aftab, S. M. S. Hussain, I. Ali, P. K. Tiwari, A. K. Goswami, and T. S. Ustun, "Virtual power plant management in smart grids with XMPP based IEC 61850 communication," *Energies*, vol. 12, no. 12, p. 2398, Jun. 2019.

[5] V. Ferrari and Y. Lopes, "Dynamic adaptive protection based on IEC 61850," *IEEE Latin Amer. Trans.*, vol. 18, no. 7, pp. 1302–1310, Jul. 2020, doi: 10.1109/TLA.2020.9099773.

[6] T. S. Ustun and Y. Aoto, "Analysis of smart inverter's impact on the distribution network operation," *IEEE Access*, vol. 7, pp. 9790–9804, 2019.

[7] S. Shaoqun, Z. Yongli, H. Min, and Y. Hong, "Multiagent and WAN based adaptive coordinated protection system," in *Proc. IEEE/PES Transmiss. Distrib. Conf. Expo., Asia Pacific*, Dalian, China, Aug. 2005, pp. 1–6.

[8] T. S. Ustun, "Design and development of a communication-assisted microgrid protection system," Ph.D. dissertation, School Eng. Sci., Victoria Univ., Melbourne, VIC, Australia, 2013.

[9] M. Khalaf, A. Hooshyar, and E. El-Saadany, "On false data injection in wide area protection schemes," in *Proc. IEEE Power Energy Soc. Gen. Meeting (PESGM)*, Portland, OR, USA, Aug. 2018, pp. 1–5.

[10] *Communication Networks and Systems for Power Utility Automation—Part 1: Introduction and Overview*, Standard IEC TR 61850-1:2013, International Standard, International Electrotechnical Commission, 2013.

[11] *Communication Networks and Systems for Power Utility Automation—Part 8—1: Specific Communication Service Mapping (SCSM)—Mappings to MMS (ISO 9506-1 and ISO 9506-2) and to ISO/IEC 8802-3*, Standard IEC TR 61850-8-1:2011, International Standard, International Electrotechnical Commission, 2020.

[12] M. A. Aftab, S. M. S. Hussain, I. Ali, and T. S. Ustun, "IEC 61850 based substation automation system: A survey," *Int. J. Elect. Power Energy Syst.*, vol. 120, Sep. 2020, Art. no. 106008.

[13] A. O. Otuoze, M. W. Mustafa, and R. M. Larik, "Smart grids security challenges: Classification by sources of threats," *J. Elect. Syst. Inf. Technol.*, vol. 5, no. 3, pp. 468–483, 2018.

[14] *Power Systems Management and Associated Information Exchange—Data and Communications Security*, Standard IEC 62351, International Standard, International Electrotechnical Commission, 2020.

[15] S. M. Farooq, S. M. S. Hussain, and T. S. Ustun, "Performance evaluation and analysis of IEC 62351-6 probabilistic signature scheme for securing GOOSE messages," *IEEE Access*, vol. 7, pp. 32343–32351, 2019, doi: 10.1109/ACCESS.2019.2902571.

[16] S. M. S. Hussain, S. M. Farooq, and T. S. Ustun, "A method for achieving confidentiality and integrity in IEC 61850 GOOSE messages," *IEEE Trans. Power Del.*, vol. 35, no. 5, pp. 2565–2567, Oct. 2020, doi: 10.1109/TPWRD.2020.2990760.

[17] S. M. Farooq, S. M. S. Hussain, S. Kiran, and T. S. Ustun, "Certificate based security mechanisms in vehicular ad-hoc networks based on IEC 61850 and IEEE WAVE standards," *Electronics*, vol. 8, no. 1, p. 96, Jan. 2019.

[18] S. Farooq, S. Hussain, S. Kiran, and T. Ustun, "Certificate based authentication mechanism for PMU communication networks based on IEC 61850-90-5," *Electronics*, vol. 7, no. 12, p. 370, Dec. 2018.

[19] M. R. Asghar and D. Miorandi, "A holistic view of security and privacy issues in smart grids," in *Smart Grid Security* (Lecture Notes in Computer Science), vol. 7823, J. Cuellar, Ed. Berlin, Germany: Springer, 2013, doi: 10.1007/978-3-642-38030-3_4.

[20] S. M. Suhail Hussain, S. Mullapathi Farooq, and T. Selim Ustun, "Analysis and implementation of message authentication code (MAC) algorithms for GOOSE message security," *IEEE Access*, vol. 7, pp. 80980–80984, 2019, doi: 10.1109/ACCESS.2019.2923728.

[21] A. F. S. Prisco and M. J. Freddy Duitama, "Intrusion detection system for SCADA platforms through machine learning algorithms," in *Proc. IEEE Colombian Conf. Commun. Comput. (COLCOM)*, Cartagena, Colombia, Aug. 2017, pp. 1–6, doi: 10.1109/ColComCon.2017.8088210.

[22] R. R. Barbosa and A. Pras, "Intrusion detection in SCADA networks," in *Mechanisms for Autonomous Management of Networks and Services* (Lecture Notes in Computer Science), vol. 6155, B. Stiller and F. D. Turck, Eds. Berlin, Germany: Springer, 2010. [Online]. Available: https://doi.org/10.1007/978-3-642-13986-4_23.

[23] Y. Yang, K. McLaughlin, L. Gao, S. Sezer, Y. Yuan, and Y. Gong, "Intrusion detection system for IEC 61850 based smart substations," in *Proc. IEEE Power Energy Soc. Gen. Meeting (PESGM)*, Boston, MA, USA, Jul. 2016, pp. 1–5, doi: 10.1109/PESGM.2016.7741668.

[24] Y. Yang, H.-Q. Xu, L. Gao, Y.-B. Yuan, K. McLaughlin, and S. Sezer, "Multidimensional intrusion detection system for IEC 61850-based SCADA networks," *IEEE Trans. Power Del.*, vol. 32, no. 2, pp. 1068–1078, Apr. 2017, doi: 10.1109/TPWRD.2016.2603339.

[25] M. El Hariri, E. Harmon, T. Youssef, M. Saleh, H. Habib, and O. Mohammed, "The IEC 61850 sampled measured values protocol: Analysis, threat identification, and feasibility of using NN forecasters to detect spoofed packets," *Energies*, vol. 12, no. 19, p. 3731, Sep. 2019.

[26] P. Nsonga, S. M. S. Hussain, I. Ali, and T. S. Ustun, "Using IEC 61850 and IEEE WAVE standards in ad-hoc networks for electric vehicle charging management," in *Proc. IEEE Online Conf. Green Commun. (OnlineGreenComm)*, Piscataway, NJ, USA, Nov. 2016, pp. 39–44, doi: 10.1109/OnlineGreenCom.2016.7805404.

[27] N. Liu, J. Chen, H. Luo, and W. Liu, "A preliminary communication model of smart meter based on IEC 61850," in *Proc. Asia–Pacific Power Energy Eng. Conf.*, Wuhan, China, Mar. 2011, pp. 1–4, doi: 10.1109/APPEEC.2011.5748756.

[28] H. J. Kim, C. M. Jeong, J.-M. Sohn, J.-Y. Joo, V. Donde, Y. Ko, and Y. T. Yoon, "A comprehensive review of practical issues for interoperability using the common information model in smart grids," *Energies*, vol. 13, no. 6, p. 1435, Mar. 2020.

[29] *Power Systems Management and Associated Information Exchange—Data and Communications Security—Part 6: Security for IEC 61850*, Standard IEC 62351-6, International Standard, International Electrotechnical Commission, 2020.

[30] C. Diago and A. Forshaw, "Cybersecurity for shared infrastructure substation networks with IEC 61850 GOOSE and sampled values," *J. Eng.*, vol. 2018, no. 15, pp. 1195–1198, Oct. 2018.

[31] J. Cai, Y. Zheng, and Z. Zhou, "Review of cyber-security challenges and measures in smart substation," in *Proc. Int. Conf. Smart Grid Clean Energy Technol. (ICSGCE)*, Chengdu, China, Oct. 2016, pp. 65–69, doi: 10.1109/ICSGCE.2016.7876027.

[32] S. M. S. Hussain, T. S. Ustun, and A. Kalam, "A review of IEC 62351 security mechanisms for IEC 61850 message exchanges," *IEEE Trans. Ind. Informat.*, vol. 16, no. 9, pp. 5643–5654, Sep. 2020, doi: 10.1109/TII.2019.2956734.

[33] T. S. Ustun, S. M. Farooq, and S. M. S. Hussain, "A novel approach for mitigation of replay and masquerade attacks in smartgrids using IEC 61850 standard," *IEEE Access*, vol. 7, pp. 156044–156053, 2019, doi: 10.1109/ACCESS.2019.2948117.
[34] J. J. Grainger and W. D. Stevenson, *Power System Analysis*. New York, NY, USA: McGraw-Hill Tata, 2003, p. 380.
[35] *Communication Networks and Systems for Power Utility Automation—Part 9-2: Specific Communication Service Mapping (SCSM)—Sampled Values Over ISO/IEC 8802-3*, Standard IEC TR 61850-9-2:2011, International Electrotechnical Commission, 2020.
[36] S. M. Farooq, S. M. S. Hussain, and T. S. Ustun, "S-GoSV: Framework for generating secure IEC 61850 GOOSE and sample value messages," *Energies*, vol. 12, no. 13, p. 2536, Jul. 2019.
[37] *Implementation Guideline for Digital Interface to Instrument Transformers Using*, Standard IEC 61850-9-2, UCA International Users Group, 2004. Accessed: Dec. 25, 2020. [Online]. Available: https://bit.ly/2KVubtc

TAHA SELIM USTUN (Member, IEEE) received the Ph.D. degree in electrical engineering from Victoria University, Melbourne, VIC, Australia. He is currently a Researcher with the Fukushima Renewable Energy Institute, AIST (FREA), where he leads the Smart Grid Cybersecurity Laboratory. Prior to that, he was a Faculty Member with the School of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA, USA. He has been invited to run specialist courses in Africa, India, and China. He has delivered talks for the Qatar Foundation, the World Energy Council, the Waterloo Global Science Initiative, and the European Union Energy Initiative (EUEI). His research has attracted funding from prestigious programs in Japan, Australia, EU, and North America. His current research interests include power systems protection, communication in power networks, distributed generation, microgrids, electric vehicle integration, and cybersecurity in smart grids. He is also a member of the IEC Renewable Energy Management WG 8 and IEC TC 57 WG17. He also serves on the Editorial Board of IEEE Access, IEEE Transactions on Industrial Informatics, *Energies*, *Electronics*, *Electricity*, *World Electric Vehicle Journal*, and *Information* Journal.

S. M. SUHAIL HUSSAIN (Member, IEEE) received the Ph.D. degree in electrical engineering from Jamia Millia Islamia (a Central University), New Delhi, India, in 2018. He is currently a Research Fellow with the Department of Computer Science, National University of Singapore, Singapore. Prior to that, he was an AIST Postdoctoral Researcher with the Fukushima Renewable Energy Institute, AIST (FREA), Koriyama, Japan. His research interests include power system communication, cybersecurity in power systems, substation automation systems, IEC 61850 standards, electric vehicle integration, and smart grid. He was a recipient of IEEE Standards Education Grant approved by the IEEE Standards Education Committee for implementing project and submitting a student application paper for the term 2014–2015. He is also a Guest Editor of the IEEE Transactions on Industrial Informatics.

LEVENT YAVUZ (Member, IEEE) received the degree in physics and electrical electronic engineering from Erciyes University. He is currently a Research Assistant with the Abdullah Gül University. He established a start-up by the Tubitak Foundation in ODTU Teknokent, as an Entrepreneur. He received three patents. His research interests include machine learning, deep reinforcement learning, virtual power plant, microgrids, solar generation forecasting, nuclear energy, and renewable energy systems.

AHMET ONEN (Member, IEEE) received the B.Sc. degree in electrical-electronics engineering from Gaziantep University, in 2005, the M.S. degree in electrical-computer engineering from Clemson University, in 2010, and the Ph.D. degree from the Electrical and Computer Engineering Department, Virginia Tech, in 2014. He is currently working as an Associate Professor with Abdullah Gül University. His research interests include transmission network and smart grids, big data in power systems, renewables and integrations, power system optimization, microgrid, and virtual power plants.