communication between these devices [4]. Other examples include microgrid power management [5] and dynamic adaptive protection [6].”

“IEC 61850 communication standard has emerged as the leader in this field due to several advantages [7]. 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 [8]. Literature sees a constant influx of device and system modeling based on IEC 61850 standard and this is only expected to increase [9].”

However, it has been reported in the literature that this high connectivity creates many cybersecurity vulnerabilities in smart grids [10]. “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 [11]. There are different studies that focus on how these two standards can be merged and secure IEC 61850 messages can be sent [12–15].”

These IT measures are excellent towards securing message exchanges. However, holistic cybersecurity design requires that additional schemes are also implemented [16]. “For instance, currently, IEC 62351 does not have any recommendation towards intrusion detection in smart grids. 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 Generic Object-Oriented Substation Event (GOOSE) messages.” As the name implies, these messages are triggered by certain events in the power system and are transmitted to take necessary reactionary precautions, e.g., a trip command is sent to a circuit breaker via GOOSE message after a relay picks up excessive current readings. Novel uses of GOOSE messages have been proposed where these event-based messages are used to implement an energy management system and coordinate electric vehicle charging. Due to the critical nature of the places of their use, GOOSE 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 [17]. In addition, there are works that focus on extending these security measures and investigate possibility of using other algorithms or encryption [18]. 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 for GOOSE messages [19,20], these works focus on statistical analysis based on parameters of current GOOSE messages. However, for effective operation it is desired that the subscriber device has intelligence of attack scenarios to identify the faulty messages with more accuracy. The machine learning algorithms can be used for training the subscriber IEDs with the desired intelligence. In the literature, machine learning algorithms have been proposed for intrusion detection in Supervisory Control and Data Acquisition Systems (SCADA) [21,22]. Currently, there is no machine learning-based mechanism for detecting intrusion in power system communication networks employing GOOSE 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

The 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. Recently, its use has been extended outside substations, yet the operational principles stayed the same. GOOSE messages are triggered when a predetermined event occurs in the power system and a message is sent to subscribers which need to be alerted and react to this event. As shown in Figure 2, GOOSE messages are sent as a burst after the event and, then, settle down to cyclic messages. The reason behind this is the sensitive nature of the GOOSE message contents. Traditionally, they are used for substation protection devices, and to increase the delivery rate at the subscriber, a burst of the same messages is sent. When GOOSE messages are published for an event, if a message is lost or delayed in the network, the burst increases the chances of an event being reported in time by the next message. Since protection systems are very limited in time, handshaking procedures between sender and receiver are out of the question.

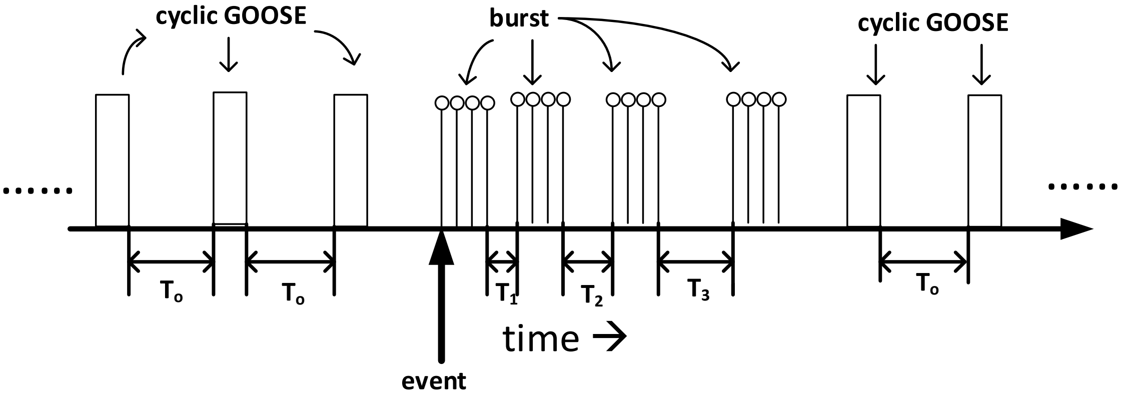


Figure 2. GOOSE message retransmission in IEC 61850.

Still, none of the messages in Figure 2 are identical and they have numbers that are used to differentiate between GOOSE messages that belong to different events as well as GOOSE messages that belong to the same event and are repetitions of each other. Figure 3 shows the contents of a GOOSE message as described by IEC 61850. Inside the GOOSE Application layer Protocol Data Unit (APDU), there are two distinct parameters utilized for this purpose. The parameter "*stNum*" is utilized to keep track of status changes, i.e., events. On the other hand, "*sqNum*" represents the sequence number for a single *stNum*. Therefore, GOOSE messages that belong to the same event and are repetitions in the same sequence have the same *stNum*, while *sqNum* increases in time. Similarly, when a new event occurs in the system, e.g., Figure 2, *stNum* is incremented by one while *sqNum* is reset to 1. This means a new event has occurred and the first message for this event is sent with *sqNum* = 1. These parameters are pivotal in monitoring the events in a power system and will be used in the proposed intrusion detection system later.

"The current structure of GOOSE messages and the way in which these messages are transmitted have various cybersecurity vulnerabilities [26]. The traditional 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 [27]."

knowledge gap, a machine learning-based intrusion detection algorithm is developed in the next section.

3. Machine Learning-Based Intrusion Detection Algorithm

As mentioned in the previous section, GOOSE messages have two parameters that are utilized to track sequences of messages pertaining to the same event as well as status changes that occur with individual events. These parameters can be utilized to detect whether the system is operating as usual or an intruder with a malicious intent has gained access to the system.

The original use of GOOSE messages is intended for sending tripping signals from relays to circuit breakers. This means GOOSE messages for new events should only be issued if there is a fault in the system. Therefore, it is expected that in a healthy system GOOSE messages should have very high *sqNum* values and *stNum* values should not change very often, i.e., events should not be very frequent and mostly cyclic GOOSE messages should be present in the network.

Conversely, if a hacker gains access to a power system communication network, they would like to inflict as much damage as possible in a short period of time. In such a scenario, hackers would send several GOOSE messages to instruct power system equipment to trip, power off or change operation in a way that it would hinder the operation of the system or cause a black-out. This would mean that GOOSE messages are sent very frequently, and events occur with very little separation. It would be possible to observe this phenomenon as very small *stNum* values and very frequently increasing *sqNum* values. It is also possible to observe frequent resets of *stNum* values as every new sequence of GOOSE message starts with *sqNum =1* when *stNum* is incremented. Additionally, as depicted in Figure 2, the communication network will be flooded with burst-type messages of new GOOSE sequences, contrary to stable operation where mostly cyclic messages are transmitted.

"Based on these facts, it is possible to design an intrusion detection system as shown in Figure 4. When an event occurs in the system, respective GOOSE message is issued. In parallel with power system operation, an event analysis is performed for this event. Based on the event history, i.e. previous events, the most recent event is 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. Otherwise, the accumulating evidence indicates that there is an intruder in the system and the alarm is raised."

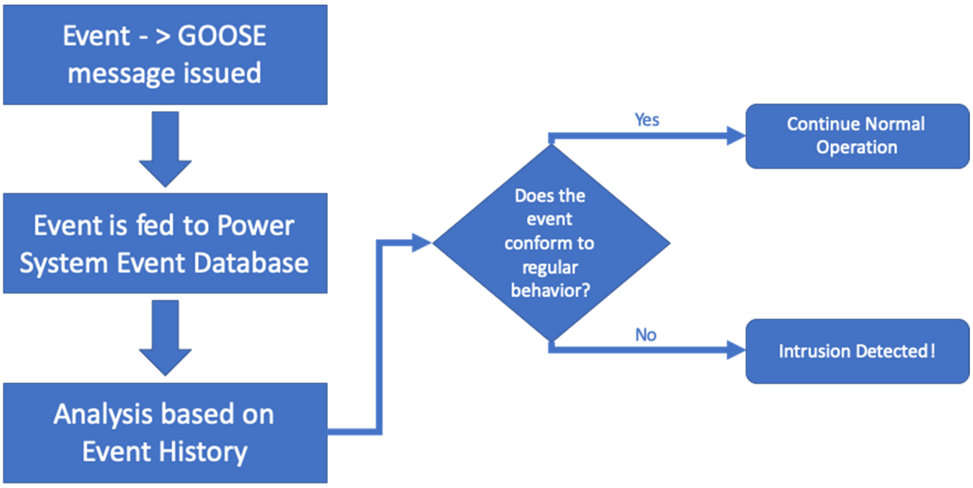

Event - > GOOSE message issued
Event is fed to Power System Event Database
Analysis based on Event History
Does the event conform to regular behavior?
Yes
No
Continue Normal Operation
Intrusion Detected!

Figure 4. Developed intrusion detection system for GOOSE messages.

It goes without saying that every power system, or sub-system such as a microgrid or a sub-station, has different behavior. Therefore, the comparison performed in Figure 4 needs to be particular to each system, not generic. This requires analyzing the past events

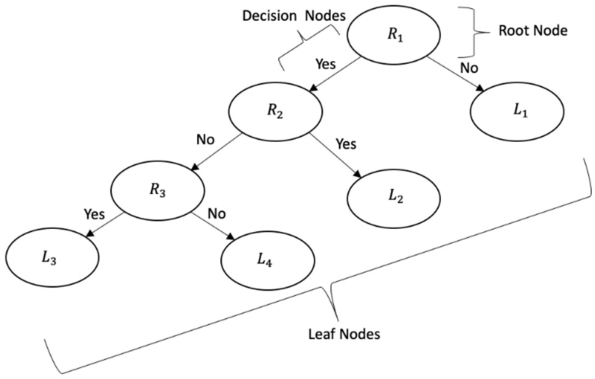


Figure 6. Decision tree operation structure.

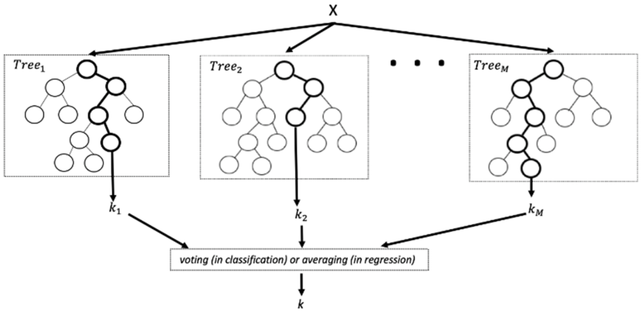


Figure 7. A random forest with several decision trees.

SVMs are utilized to process input data and decide which of the two classes they belong to. As a non-probabilistic binary linear classifier, an SVM develops a model which is utilized to assign new inputs to each category. The SVM model represents inputs as points in space and separates them into two categories where the incoming data are assigned to appropriate class in space. Due to the nature of the intrusion detection system proposed herein, the SVM approach is selected as it lends itself to the application of the proposed system. Any incoming data are processed and classified as regular operation or attack by a hacker. The use of SVMs in non-linear data requires kernel tricks or kernel numbers. In this study, the selected kernel type is Radial Basis Function (RBF), while the gamma value, i.e., kernel coefficient, is 0.125. RBF is one of the most commonly used kernel methods for nonlinear support in SVMs. The operation principle of SVMs is depicted in Figure 8.

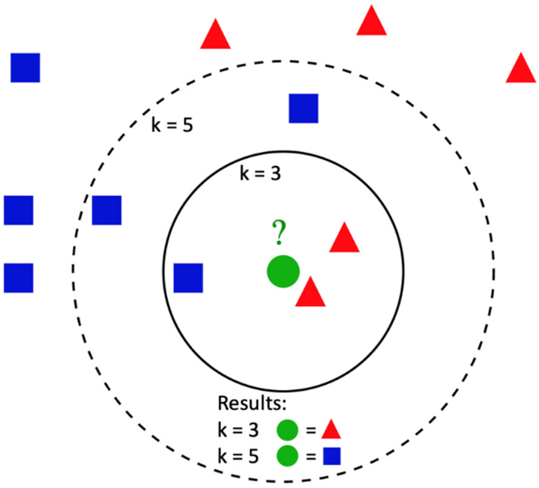


Figure 8. Impact of k value on k-NN algorithm’s output.

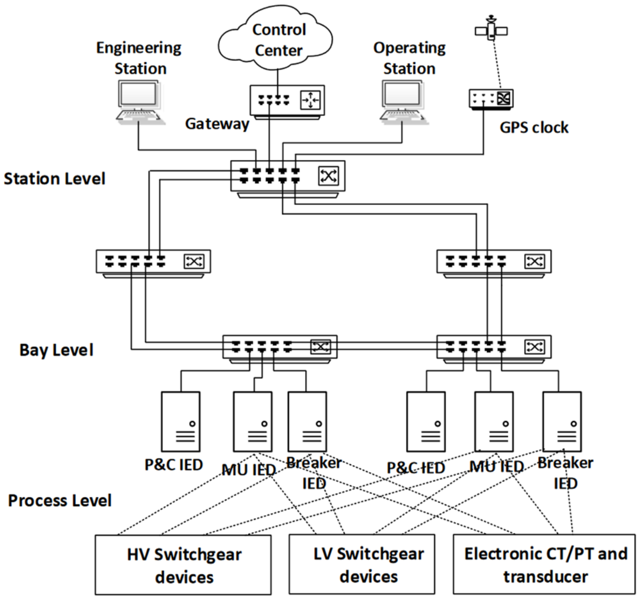
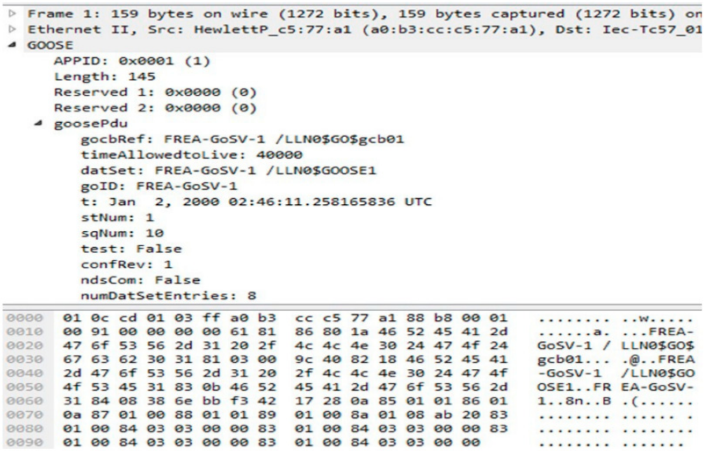


Figure 10. Power system setup for GOOSE messages.



```
▷ Frame 1: 159 bytes on wire (1272 bits), 159 bytes captured (1272 bits) on
▷ Ethernet II, Src: HewlettP_c5:77:a1 (a0:b3:cc:c5:77:a1), Dst: Iec-Tc57_01
◢ GOOSE
    APPID: 0x0001 (1)
    Length: 145
    Reserved 1: 0x0000 (0)
    Reserved 2: 0x0000 (0)
  ◢ goosePdu
      gocbRef: FREA-GoSV-1 /LLN0$GO$gcb01
      timeAllowedtoLive: 40000
      datSet: FREA-GoSV-1 /LLN0$GOOSE1
      goID: FREA-GoSV-1
      t: Jan  2, 2000 02:46:11.258165836 UTC
      stNum: 1
      sqNum: 10
      test: False
      confRev: 1
      ndsCom: False
      numDatSetEntries: 8
0000  01 0c cd 01 03 ff a0 b3  cc c5 77 a1 88 b8 00 01   ........ ..w.....
0010  00 91 00 00 00 00 61 81  86 80 1a 46 52 45 41 2d   ......a. ...FREA-
0020  47 6f 53 56 2d 31 20 2f  4c 4c 4e 30 24 47 4f 24   GoSV-1 / LLN0$GO$
0030  67 63 62 30 31 81 03 00  9c 40 82 18 46 52 45 41   gcb01... .@..FREA
0040  2d 47 6f 53 56 2d 31 20  2f 4c 4c 4e 30 24 47 4f   -GoSV-1  /LLN0$GO
0050  4f 53 45 31 83 0b 46 52  45 41 2d 47 6f 53 56 2d   OSE1..FR EA-GoSV-
0060  31 84 08 38 6e bb f3 42  17 28 0a 85 01 01 86 01   1..8n..B .(......
0070  0a 87 01 00 88 01 01 89  01 00 8a 01 08 ab 20 83   ........ ...... .
0080  01 00 84 03 03 00 00 83  01 00 84 03 03 00 00 83   ........ ........
0090  01 00 84 03 03 00 00 83  01 00 84 03 03 00 00      ........ .......
```

Figure 11. GOOSE message capture where stNum = 1 and sqNum = 10.

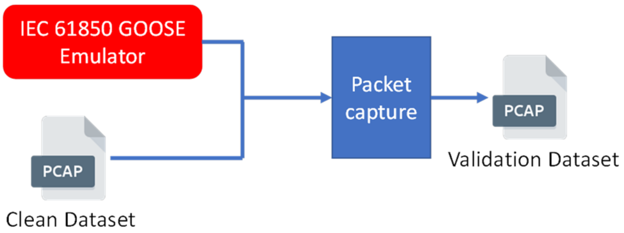


Figure 12. Experimental set up for generating validation dataset.

During normal operation, events occur with some time separation between them. This can be observed from the rate of change in *stNum* values as well as the final value that *sqNum* reaches. In attack data that are injected into the dataset, the *stNum* value increases pretty rapidly with a high rate of change, while the corresponding *sqNum* values stay uncharacteristically low.

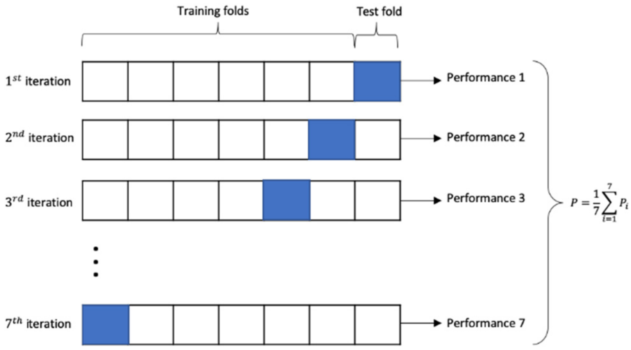


Figure 15. Cross validation approach with several iterations.

Table 2. Accuracy, detection rate and false alarm rate of different machine learning algorithms.

Algorithm	Accuracy	Detection Rate (DR)	False Alarm Rate (FAR)
Adaboost	0.9487	0.8507	0.0194
DT	0.9448	0.9231	0.0408
RF	0.9519	0.8657	0
k-NN	0.9448	0.9231	0.0408
SVM	0.9512	0.8718	0.0338

Table 3. Training, test and overall timing of different machine learning algorithms.

	I7 (Time in Seconds)			R-Pi 3 (Time in Seconds)		
Algorithm	**Overall**	**Training**	**Test**	**Overall**	**Training**	**Test**
Adaboost	71.85	71.84	0.0169	699.07	698.95	0.069
DT	1.36	1.36	0.0009	19.57	19.50	0.031
RF	59.7	59.69	0.0080	855.16	855.07	0.054
k-NN	13.41	13.42	0.0019	215.30	215.27	0.003
SVM	3427.69	3427.69	0.0029	31655.86	31655.81	0.016

Secondly, it is possible to observe that RF, SVM and Adaboost have the highest accuracy among the algorithms. Considering that Adaboost is designed the boost the performance of other classifiers and decision stumps are utilized as the underlying classifier, such an outcome is expected. On the other hand, DT has a comparable accuracy due to the simple nature of the prediction that is made. In a system with several inputs and interdependent variables, Adaboost yields higher accuracy.

Detection Rates (DR) are acceptable, with the highest value being 92.31% for both DT and k-NN. These two algorithms have very high False Alarm Rates (FAR), which brings down their overall accuracy. The lowest FAR is reported for Adaboost, which seems to have the best balance of accuracy, DR and FAR for this application.

The most important aspect of the test results is the timing values, as shown in Table 3. There are three parameters: (i) training time, the time required to train the system; (ii) testing time, the time required to run the algorithm and detect an attack; and (iii) overall time required for training and testing. The training and testing times are completely distinct and relate to different steps of operation. Training can be performed offline or before the deployment of the system. Therefore, it does not have a direct impact on the system operation when GOOSE 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 GOOSE message and decide whether it is a normal message or an attack message, as shown in Figure 5. It is also important to note

Author Contributions: Conceptualization, T.S.U. and S.M.S.H.; Data curation, A.U.; Formal analysis, T.S.U.; Funding acquisition, T.S.U.; Investigation, T.S.U.; Methodology, S.M.S.H.; Software, A.U.; Writing—Revision, A.O., M.M.R. and D.M., Supervision, T.S.U. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Ministry of Energy, Transportation and Industry, METI, Japan.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

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

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. 4 December 2011. Available online: https://bit.ly/3r7Zv8T (accessed on 18 December 2020).
2. Gangale, F.; Vasiljevska, J.; Covrig, F.; Mengolini, A.; Fulli, G. *Smart Grid Projects Outlook 2017: Facts, Figures and Trends in Europe*; EUR 28614 EN; Publications Office of the EU: Luxembourg, 2017. [CrossRef]
3. Ustun, T.S.; Hussain, S.M.S.; Kikusato, H. IEC 61850-Based Communication Modeling of EV Charge-Discharge Management for Maximum PV Generation. *IEEE Access* **2019**, *7*, 4219–4231. [CrossRef]
4. Nadeem, F.; Aftab, M.A.; Hussain, S.S.; Ali, I.; Tiwari, P.K.; Goswami, A.K.; Ustun, T.S. Virtual Power Plant Management in Smart Grids with XMPP Based IEC 61850 Communication. *Energies* **2019**, *12*, 2398. [CrossRef]
5. Ustun, T.S.; Hussain, S.M.S. IEC 61850 Modeling of UPFC and XMPP Communication for Power Management in Microgrids. *IEEE Access* **2020**, *8*, 141696–141704. [CrossRef]
6. Ferrari, V.; Lopes, Y. Dynamic Adaptive Protection based on IEC 61850. *IEEE Lat. Am. Trans.* **2020**, *18*, 1302–1310. [CrossRef]
7. International Electrotechnical Commission. *IEC TR 61850-1:2013, Communication Networks and Systems for Power Utility Automation—Part 1: Introduction and Overview*; International Standard: Geneva, Switzerland, 2013.
8. International Electrotechnical Commission. *IEC TR 61850-8-1:2011, 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*; International Standard: Geneva, Switzerland, 2020.
9. Aftab, M.A.; Hussain, S.S.; Ali, I.; Ustun, T.S. IEC 61850 based substation automation system: A survey. *Int. J. Electr. Power Energy Syst.* **2020**, *120*, 106008. [CrossRef]
10. Otuoze, A.O.; Mustafa, M.W.; Larik, R.M. Smart grids security challenges: Classification by sources of threats. *J. Electr. Syst. Inf. Technol.* **2018**, *5*, 468–483. [CrossRef]
11. International Electrotechnical Commission. *IEC 62351-Power Systems Management and Associated Information Exchange—Data and Communications Security*; International Standard: Geneva, Switzerland, 2020.
12. Ustun, T.S.; Farooq, S.M.; Hussain, S.M.S. Implementing Secure Routable GOOSE and SV Messages Based on IEC 61850-90-5. *IEEE Access* **2020**, *8*, 26162–26171. [CrossRef]
13. Hussain, S.M.S.; Farooq, S.M.; Ustun, T.S. A Method for Achieving Confidentiality and Integrity in IEC 61850 GOOSE Messages. *IEEE Trans. Power Deliv.* **2020**, *35*, 2565–2567. [CrossRef]
14. Farooq, S.M.; Hussain, S.M.S.; Kiran, S.; Ustun, T.S. Certificate Based Security Mechanisms in Vehicular Ad-Hoc Networks based on IEC 61850 and IEEE WAVE Standards. *Electronics* **2019**, *8*, 96. [CrossRef]
15. Farooq, S.M.; Hussain, S.M.S.; Kiran, S.; Ustun, T.S. Certificate Based Authentication Mechanism for PMU Communication Networks Based on IEC 61850-90-5. *Electronics* **2018**, *7*, 370. [CrossRef]
16. Asghar, M.R.; Miorandi, D. A Holistic View of Security and Privacy Issues in Smart Grids. In *Smart Grid Security. SmartGridSec 2012*; Cuellar, J., Ed.; Lecture Notes in Computer Science; Springer: Berlin/Heidelberg, Germany, 2013; Volume 7823. [CrossRef]
17. Hussain, S.S.; Farooq, S.M.; Ustun, T.S. Analysis and Implementation of Message Authentication Code (MAC) Algorithms for GOOSE Message Security. *IEEE Access* **2019**, *7*, 80980–80984. [CrossRef]
18. Farooq, S.M.; Hussain, S.S.; Ustun, T.S. Performance Evaluation and Analysis of IEC 62351-6 Probabilistic Signature Scheme for Securing GOOSE Messages. *IEEE Access* **2019**, *7*, 32343–32351. [CrossRef]
19. Quincozes, S.E.; Albuquerque, C.; Passos, D.; Mossé, D. A survey on intrusion detection and prevention systems in digital substations. *Comput. Netw.* **2021**, *184*, 107679. [CrossRef]
20. Hong, J.; Liu, C.; Govindarasu, M. Detection of cyber intrusions using network-based multicast messages for substation automation. In Proceedings of the Innovative Smart Grid Technologies (ISGT), Washington, DC, USA, 19–22 February 2014; IEEE: Piscataway, NJ, USA, 2014; pp. 1–5.