Practical Challenges of Attack Detection in Microgrids Using Machine Learning

Ramotsoela, Daniel T.; Hancke, Gerhard P.; Abu-Mahfouz, Adnan M.

Published in:
Journal of Sensor and Actuator Networks

Published: 01/02/2023

Document Version:
Final Published version, also known as Publisher's PDF, Publisher's Final version or Version of Record

License:
CC BY

Publication record in CityU Scholars:
Go to record

Published version (DOI):
10.3390/jsan12010007

Publication details:

Citing this paper
Please note that where the full-text provided on CityU Scholars is the Post-print version (also known as Accepted Author Manuscript, Peer-reviewed or Author Final version), it may differ from the Final Published version. When citing, ensure that you check and use the publisher's definitive version for pagination and other details.

General rights
Copyright for the publications made accessible via the CityU Scholars portal is retained by the author(s) and/or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights. Users may not further distribute the material or use it for any profit-making activity or commercial gain.

Publisher permission
Permission for previously published items are in accordance with publisher’s copyright policies sourced from the SHERPA RoMEO database. Links to full text versions (either Published or Post-print) are only available if corresponding publishers allow open access.

Take down policy
Contact lbscholars@cityu.edu.hk if you believe that this document breaches copyright and provide us with details. We will remove access to the work immediately and investigate your claim.

Download date: 13/05/2024
Review

Practical Challenges of Attack Detection in Microgrids Using Machine Learning

Daniel T. Ramotsoela 1, Gerhard P. Hancke 2,3, and Adnan M. Abu-Mahfouz 3,4

1 Department of Electrical Engineering, University of Cape Town, Cape Town 7701, South Africa
2 Department of Computer Science, City University of Hong Kong, Hong Kong, China
3 Department of Electrical, Electronic and Computer Engineering, University of Pretoria, Pretoria 0002, South Africa
4 Council for Scientific and Industrial Research (CSIR), Pretoria 0184, South Africa

Abstract: The move towards renewable energy and technological advancements in the generation, distribution and transmission of electricity have increased the popularity of microgrids. The popularity of these decentralised applications has coincided with advancements in the field of telecommunications allowing for the efficient implementation of these applications. This convenience has, however, also coincided with an increase in the attack surface of these systems, resulting in an increase in the number of cyber-attacks against them. Preventative network security mechanisms alone are not enough to protect these systems as a critical design feature is system resilience, so intrusion detection and prevention system are required. The practical consideration for the implementation of the proposed schemes in practice is, however, neglected in the literature. This paper attempts to address this by generalising these considerations and using the lessons learned from water distribution systems as a case study. It was found that the considerations are similar irrespective of the application environment even though context-specific information is a requirement for effective deployment.

Keywords: microgrids; cyber–physical systems; industrial control systems; intrusion detection systems; machine learning; network security

1. Introduction

Traditional electricity supply systems rely on a central generating facility which feeds into the transmission and distribution networks [1]. These centralised systems are typically expensive, inefficient, and rely heavily on fossil fuels [2]. The fossil fuels on which these systems rely are not only a scarce resource, but they also contribute significantly towards pollution. In recent years, the concept of decentralising the generation capacity of electricity supply systems has become increasingly popular [3–5]. These distributed generation systems are more efficient and rely primarily on renewable energy sources instead of fossil fuels. Incorporating these distributed generation systems into the traditional electricity supply systems is challenging. This is because the latter relies on centralised generation but integration achieved by making use of microgrids [2]. In this setup, the consumer who traditionally only used power from the grid becomes a prosumer, who can now additionally add power to the grid [6]. These microgrids also leverage Internet of Things (IoT) technologies [7] to allow for the intelligent monitoring and control of the system. The use of these technologies can greatly improve the efficiency of the system. The microgrid concept will thus serve as the most critical component towards the realization of smart cities [8].

The energy sector has been identified as the most important of the 16 critical infrastructure sectors identified by the United States (US) Department of Homeland Security [9]. Critical infrastructure systems are those whose operations are paramount to the security and stability of a nation. The energy sector is regarded as the most critical because all of the other sectors are either directly or indirectly dependant of this sector. The energy sector
consistently being ranked the most targeted critical infrastructure sector in the US is further evidence of this [10]. It is thus a matter of national security for any nation to ensure that the operation of the electricity supply network is not compromised. This will become increasingly challenging as these critical infrastructure applications employ IoT technologies to improve system operations. This is because these technologies introduce cyber-threats which, if successfully carried out, could potentially have devastating economic, social and environmental consequences [11]. This means that an integral part of developing these critical infrastructure application is that they should be resilient against all potential threats, both intentional and unintentional [12].

Considering the critical nature of sector where these microgrids are deployed, it becomes clear that preventative cyber-security mechanisms alone will not be adequate to protect them. The resilience requirement identified in the previous paragraph also required the system to recover even when a threat is realized. When considering cyber-threats, this can be achieved by making use of intrusion detection and prevention systems (IDPS) [10]. These systems are deployed because preventative cyber-security mechanisms will inevitably fail and a resilient system should have measures to recover from these security failures. IDPSs involve both the detection of attacks and the reaction of the system once an attack has been detected in order to mitigate the potential damage [13].

A popular categorisation of these IDPSs is based on how the information is analysed, i.e., signature-based or behaviour-based [14]. Table 1 shows a comparative analysis of each category. When protecting critical infrastructure, the reactive nature of signature-based methods is not ideal because the consequences of successful cyber-attacks could be devastating. Attackers are also continuously developing new and innovative ways to compromise preventative security mechanisms. This is spurred on by the fast pace of technological development which means that these systems are constantly being upgraded to meet changing system requirements and adapt to new computing environments. The evolving nature of today’s technological world thus constantly introduces new security vulnerabilities which attackers can exploit. The result of this is that no matter how comprehensive an attack database is, there will always be exploits that are not accounted for because they will only exist in later iterations of the system. Evidence of this can be seen when looking at the attack on the Maroochy water treatment facility [15]. In this case, the attacker took advantage of vulnerabilities introduced by system upgrades. This resulted in significant monetary loss and environmental damage. IDSs thus need to be as adaptive and robust to change as the technological systems they are designed to protect.

Table 1. Behaviour-based vs. Signature-based detection.

<table>
<thead>
<tr>
<th></th>
<th>Signature-Based</th>
<th>Behaviour-Based</th>
</tr>
</thead>
<tbody>
<tr>
<td>Detection Speed</td>
<td>Has a fast detection time due to the use of a lookup</td>
<td>Has a slow detection time as it has to detect patterns of abnormal behaviour using machine learning</td>
</tr>
<tr>
<td>Accuracy</td>
<td>Lower false positive rate but higher false negative rate than behaviour-based methods.</td>
<td>Has a high false positive rate but lower false negative rate than signature-based methods</td>
</tr>
<tr>
<td>Attack Types</td>
<td>Can only detect previously identified attack signatures.</td>
<td>Can detect both unknown attacks and system faults.</td>
</tr>
<tr>
<td>Prior knowledge</td>
<td>Needs to maintain a large database of attacks in order to detect them</td>
<td>Needs prior knowledge of normal system behaviour in order to train models. Can also be trained using abnormal behaviour.</td>
</tr>
</tbody>
</table>
The current literature on this topic is fragmented and considers each industrial control system (ICS) application in isolation. The main contribution of this work is a generalisation of these concepts to demonstrate that all of these application share similar challenges and solution. The primary difference between them will be the architecture of the devices and application specific variables. The core principles will, however, remain the same, both at a system and device level. This type of analysis is currently missing from the literature and can provide some meaningful insights into the current body of work. Our contribution in this survey paper is summarised below.

- We analyse state-of-the-art research and extrapolate the challenges and limitation of intrusion detection in ICSs in general.
- We use water distribution systems as a case study and generalise the practical consideration for the implementation of these schemes. This is then applied to Microgrids based on the vulnerabilities identified in the state-of-the-art research.
- We then analyse and discuss the state-of-the-art research of IDPSs in Microgrids taking the above into consideration with a particular focus on the practical implications.

2. Cyber–Physical Systems

The operation of cyber–physical systems is realised through industrial control systems (ICS) [16]. The purpose of ICSs is to limit the amount of human interaction required by automating the control of the devices in the system [17]. The Supervisory Control And Data Acquisition (SCADA) system monitors and supervises the control-level devices from a higher level of abstraction. This allows human interactions through a human machine interface (HMI). The connection between the SCADA and control levels can be direct via a LAN or remote via a WAN. They can also make use of specialised programmable logic controllers (PLC) known as remote terminal units (RTU). SCADA systems are not responsible for the actual control logic or functionality but can set the parameters for control [18]. These parameters can be adjusted automatically using predetermined system-wide constraints or directly by a plant manager through an HMI. This means that an attacker who is able to compromise the SCADA system will have complete control over the system functionality.

The second function of the SCADA system is data acquisition which allows for the effective monitoring of the system by giving an overview of the system state from a higher level of abstraction [19]. This information can be made available in a separate historian server and used to analyse aspects such as plant efficiency and possible defects. This information can also be used in real-time to run an IDS to identify malicious behaviour in the system. An attacker who has taken over the SCADA system would, however, have superuser privileges which would enable him/her to switch off such protective functionality. From this discussion, it is clear that considerable resources need to be allocated towards the protection of SCADA systems. This is because attacks at this layer would be almost impossible to detect for an IDS running in the ICS. In this case, the network anomaly detection schemes running on the corporate network would be more ideally suited to the task. Threats to the SCADA system from malicious internal attackers are, however, easier to carry out and have more points of entry than their remote counterparts [20].

ICSs rely heavily on the communication between the interconnected devices to realise system functionality [21]. As a result, it is necessary to define relevant communication protocols which allow for safe and reliable communication within the system. For many years, the primary focus has been on the internal communication of the system with particular focus on the communication between the field and control devices. This group of protocols are collectively referred to as fieldbus communication protocols [18]. More recently, as systems move towards the cyber–physical space communication has become a key component at all levels of the system hierarchy. This incorporates both internal and external communication networks with Ethernet being the dominant standard. This means that industrial control networks are now similar conceptually to conventional computer networks. They, however, remain structurally different owing to the conflicting priorities...
of the two systems. The architecture of industrial control networks is usually much deeper
than conventional IT networks. The major differences between the two network types
are also shown in Table 2. The most notable difference is the reliability and real-time
requirements of ICSs which are not major issues in IT networks. These conflicting priorities
also need to be considered when looking at the security of both systems.

Table 2. Industrial control network vs. conventional IT network.

<table>
<thead>
<tr>
<th></th>
<th>Industrial Control Network</th>
<th>Conventional IT Network</th>
</tr>
</thead>
<tbody>
<tr>
<td>System Reliability</td>
<td>Critical</td>
<td>Issues tolerated</td>
</tr>
<tr>
<td>Real-time Requirements</td>
<td>Stringent</td>
<td>lax</td>
</tr>
<tr>
<td>Determinism</td>
<td>Required</td>
<td>Not Required</td>
</tr>
<tr>
<td>Data Size</td>
<td>Small</td>
<td>Large</td>
</tr>
<tr>
<td>Data Transmission</td>
<td>Periodic and aperiodic</td>
<td>“best effort”</td>
</tr>
<tr>
<td>Time-series Events</td>
<td>Yes</td>
<td>Not usually</td>
</tr>
</tbody>
</table>

The traditional Microgrid architecture uses a centralised unidirectional hierarchical
network architecture with three layers [22]. Each of these layers have different operational
requirements and thus make use of different communication technologies and standards
to meet the required specifications. For example, at the level of the load controller, the
speed requirement is in the range of milliseconds to minutes. At the Microgrid central
controller and distribution management system, the specification is in the range of minutes
to hours and hours to days respectively. There is currently a trend towards a decentralised
bidirectional architecture that incorporates aspects of the internet architecture to meet the
current demands of Microgrid communication system. Traditional systems make use of
wired technologies such as RS232, Ethernet and power line communication. There has,
however, been trend towards wireless communications in recent years.

The authors in [23] evaluated a number of communication standards and technologies
used in Microgrids. They found that the wireless technologies were better suited for the
application environment than their wired counterparts. In particular, low powered wide
area networks (LPWAN) were found to be quite promising with their low deployment cost
and large communication range. Factors such as spectrum sharing, gateway placement and
network security were identified as key challenges for deployment. Their low data rates
also meant they were not ideal in cases of emergency which could affect the redundancy
requirement of the Microgrids. Cellular networks can also be used to achieve this long
range communication with faster data rates although these have a higher deployment
cost than LPWAN technologies [22]. Irrespective of the link-layer technology used there
is a requirement of the network to have redundancy built into the system to maintain the
resilience requirement of Microgrids. The implementation of peer-to-peer networks will
also increasingly be the norm as the architecture becomes more decentralised.

The generic process flow for cyber–physical systems is as follows. Sensors gather
information from the monitored physical environment and transmit the information to the
cyber/digital system. This information is transmitted using the industrial control network
described above, which has both internal and external communication links. In the cyber
layer, the data undergoes digital processing and the system makes a decision based on this
sensed data. This allows the system to alter the physical environment as per the system
requirements. This means that decisions can be made locally, within the SCADA network
or it could be facilitated by an external entity. The decision is implemented by making
use of actuators which allow the system to interact with the physical environment. The
actuators are thus able to manipulate the physical environment in order to achieve system
goals based on data that was sourced from sensor readings.

During a cyber-attack, an attacker could manipulate sensor readings in a coordinated
manner to force the system to malfunction. In water system infrastructure for example, an
attacker could change the tank level reading for a particular tank to manipulate the system
into pumping more and more water into that tank until it overflows. A malfunctioning
sensor which erroneously reads the tank level as lower than the actual value would also lead to the same overflow caused by the malicious attack. Even though only one of the scenarios was intentional, the resultant consequences are identical.

The consequences of malfunctioning sensors and cyber-attacks can be even more severe than a tank overflow. Consider the Boeing 737 MAX saga which resulted in two plane crashes and hundreds of people regrettably losing their lives. Although the investigations have not yet been finalised, preliminary results indicate that both plane crashes were caused by faulty sensor readings which caused the manoeuvring characteristics augmentation system (MCAS) to malfunction [24]. The sensors the MCAS was reliant on were error-prone which thus resulted in the system causing the plane to nose-dive instead of stabilising it as intended. The critical nature of that application environment thus required additional mechanisms to account for these sensor errors. This is because of the disastrous consequences that could result from a system failure. This emphasises the point that IDSs and fault detection mechanisms should be developed while taking careful consideration of the needs of the particular application environment.

An important consideration between fault detection systems and IDSs is that the system faults will be easier to detect. This is because with the latter, a malicious attacker will attempt to evade detection by employing stealthy techniques [25]. These stealthy techniques are usually employed with the sole purpose of circumventing any automated detection mechanisms. In the case of an unintentional system fault, this will not be the case and anomalous events are likely to deviate significantly from the norm which would make detection easier to achieve. This means that while the two concepts are similar, and the result of their occurrence will likely lead to the same consequence, the resultant algorithms are different owing to the presence of a potentially stealthy malicious actor.

3. Intrusion Detection Systems

In network security there is a heavy emphasis on preventative security mechanisms such as firewalls, cryptography and access control [26]. These preventative measures provide an external security perimeter to prevent an attacker from gaining access to the system. An intrusion occurs when an attacker bypasses these external security mechanisms in an attempt to compromise at least one of the three key pillars of network security (confidentiality, integrity and availability) [27]. If an attacker is able to bypass the preventative measures, intrusion detection and prevention systems can be used to limit the potential damage. IDPSs involve both the detection of attacks and the reaction of the system once an attack has been detected in order to mitigate the potential damage [13]. Figure 1 shows the security mechanisms that protect system resources. In order for an attack to be successful an attacker first has to bypass the preventative security mechanisms. Once the attacker has access to the system he/she will have to evade the IDPS in order to gain unfettered access to the desired system resources.

![Figure 1. Security mechanisms that protect system resources.](image-url)
The focus of this work is on the detection of malicious activity which is the cornerstone of IDPSs. These IDSs can be categorised in a number of different ways with one of the most popular being the type of information being utilised by the system, i.e., Network-based [28] or Host-based [29]. Network-based intrusion detection analyses network traffic of an entire network or subnetwork to determine whether an intrusion has occurred. Host-based systems utilise operational parameters specific to the host/application and is generally concerned with the integrity of a specific device. In a typical IT application, each device would have a dedicated host-based intrusion detector while the network connecting the devices uses a network anomaly detector. The latter gives an overall view of the entire system and will thus be effective at detecting contextual anomalies. Using this configuration, compromised hosts which affect network operations can also be detected but this would primarily be limited breaches in integrity and availability. The confidentiality of the data passing through the affected devices cannot be guaranteed unless a host-based detector is also running on the device. In IoT applications, this is not always possible because of the resource constraints of the host devices [30].

Figure 2 shows a generic framework for anomaly detection using machine learning. The first important step is data-preprocessing which involves filtering, data imputation and feature extraction [31]. This step is paramount in the viability of any machine learning algorithm in the application environment. It is also important in this step to understand the data as it relates to the application in order to make appropriate design choices. Once the features have been extracted, the model needs to be trained using some prior knowledge about the system/data. This prior knowledge can be in the form of establishing a normal profile for the data such that outliers can be identified as anomalies. An alternative use of prior knowledge is making use of data samples of both normal and anomalous states in order to train the model to distinguish between the two. The challenge with this alternative approach is the lack of data samples that represent the anomalous states [15]. If not adequately accounted for, this could result in a model that is biased towards the classification of the normal state. Once the model is trained, it can then be used to classify unseen data to evaluate how well it works. In this step it is important to also make use of the data samples that represent the anomalous state, even if they were not used during the training phase. A challenge with this algorithm evaluation phase is selecting metrics that are appropriate for the application environment. This is required in order to adequately analyse the results.

Figure 2. Generic Anomaly Detection Framework [15].
4. Vulnerability assessment

Like all critical infrastructure applications, the resilience of the electrical grid is an important component when considering deployment. In Microgrids, this has particular been a central focus as the distributed nature of the systems inherently makes them ideal tools for a resilient electrical grid. There have thus been many studies assessing how resilient they are to a variety of external factors [32–35]. The resilience to random yet predictable events (such as normal weather conditions) can be modelled relatively accurately. It can also be accounted for such that there is a very low risk they could adversely affect the system. Anomalous events are, however, a lot more difficult to model and account for and thus pose a much more significant risk to the system [32]. These events have been the primary focus of the research into the resilience of Microgrids. These anomalous events can be both unintentional, as is the case with natural disasters, or intentional, such as malicious attacks. The focus of this work is on the latter although good IDS should be able to detect both kinds of anomalous events.

A distinction will now be made between system-level vulnerabilities and device-level vulnerabilities [36]. The former are application dependent and will differ depending on the type of critical infrastructure being deployed. An example of this is how a Microgrid application operates differently to a water distribution system. This means that the two systems will be vulnerable to different exploits. The device-level vulnerabilities are mostly independent of the specific application and their differences are primarily vendor-specific [37]. This means that in this case the main consideration is the architecture of the device (which is vendor-specific) and not the SCADA system (which is application-specific). These types of vulnerabilities are difficult to protect against because some information is proprietary and not publicly available. Additionally these devices are resource constrained and operate in time-critical applications, so there is very little margin to introduce protective measures [38]. A popular example of these device-level vulnerabilities is that the device could download and run maliciously altered code.

Another important consideration is that microgrids occur at the end points of the smart grid architecture [39]. This means that these systems directly interact with endpoint devices so a breach in security has major ramifications for user privacy. User data could not only reveal sensitive information, but usage patterns could allow an attacker to deduce which appliances are being used and even when a resident is not home. The prosumer model of the smart grid, which allows consumers to sell excess electricity back into the grid, requires real-time and accurate information about the user. This introduces additional challenges because the anonymity of the user must be maintained in this bidding process due to the reasons mentioned [40]. From the context of intrusion detection, it is very difficult to detect breaches of confidentiality because they are passive attacks. The actions of the attacker, however, can be logged by host-based IDS and if they differ significantly from that of a legitimate user they can be flagged.

These endpoint devices can also be used as the launching pad for active attacks on the grid because of the two-way communication line. Researchers have shown, for example, that should a smart meter be infected with a worm and if that worm were allowed to spread, then tens of thousands of smart meters could potentially be infected within twenty-four hours [41]. Depending on the worm payload, this breach could result in all three of the key network security objectives (confidentiality, integrity and availability) being compromised. Malicious software has become increasingly stealthy in recent years in an attempt the shield them from detection from signature-based detectors. The resources constraints of these endpoint devices also make implementing other types of IDS more challenging in the application environment.

The introduction of wireless personal area network (WPAN) technologies such as Zigbee to microgrids also adds significant vulnerabilities to the system [42]. The added convenience and easier deployment of wireless technologies when compared to their wired counterparts comes at a cost of an increase in the attack surface of the system. This is because when using a wireless link, at attacker no longer needs to have physical access to
the system and can launch an attack remotely. These WPAN technologies are going to be an integral part of the envisioned smart cities of the future. They are typically deployed in resource constrained environments in applications that require low data rates because they are inexpensive to deploy. The characteristics that make them ideal for the application environment are also the characteristics that make them more vulnerable to attacks. They are particularly susceptible to denial of service (DoS) because of these characteristics [43]. From an intrusion detection perspective, DoS attacks are easy to detect but difficult to defend against, especially in a resource constrained devices. What is even more challenging is that in operational technology, the availability of the system is the most important of the three key network security objectives.

The challenge in protecting a system against DoS attacks is that the attacker does not need to cause a complete loss of available. In time critical industrial control systems, significantly slowing down the system could have devastating consequences [44]. The key difference in this case is that the security goals of confidentiality and integrity require the prevention of unauthorized access, which can be assessed on a binary scale. This means an attacker has either gained unauthorized access to the system resources or they have not. The security goal of availability, however, requires the preservation of authorized access. The analysis of whether or not this has been maintained is more nuanced and can differ between applications. This is one of the reasons why DoS attacks are easy to detect but difficult to defend against. A key design feature of Microgrids is that they should have resilience built into the system to mitigate the impact of such events [12]. The security protocols of the IDPS would thus need to account for this once an attack has been detected by the IDS. A potential solution to this is having redundancy built into the system.

The last of the three network security objectives which will be considered is data or system integrity. From the perspective of the system, there is no difference between a system fault and a data integrity attack. This is because whether system data have been altered maliciously or naturally, the result remains the same. One of the most prominent case studies that demonstrate this is the MCAS malfunction from the Boeing 737 saga mentioned earlier. The consequences of maliciously altered and faulty data are clearly the same, but from an intrusion detection perspective the key difference is that a malicious actor can attempt to evade detection. This means that the systems built-in fault detection mechanisms may not be adequate because an attacker with an intimate knowledge of the system can tailor the attack to evade detection.

Data integrity attacks are more serious than confidentiality attacks but usually not as severe as attacks on system availability in the application environment. The three computer security objectives are, however, linked because in order for confidentiality to be compromised, the attacker may need to compromise system or data integrity. This is also the case for distributed DoS attacks where compromised devices can be turned into Zombies by making use of Bots [45]. An attack on data or system integrity can also directly lead to a loss in availability as was the case in the attack of the Iranian nuclear program by the Stuxnet worm in 2010 [46]. Traditional DoS attacks are typically short term but these types of data integrity attacks can have devastating long term effects and loss of availability.

In the application environment, data integrity attacks can vary in the severity of impact. These attacks are, however, the most dangerous because the system relies on reliable data to make decision. The system interacts with the physical world using actuators, but the nature of the interaction is dependent on the sensed data. This means that an attacker equipped with knowledge about the application environment would be able to manipulate the system into achieving a desired outcome. In microgrids, various attacks have been demonstrated of which can affect the system’s energy efficiency [47] and generation costs [48]. The system can also be manipulated for the financial benefit of the attacker [49]. In the worst case, these data integrity attacks have been shown to result in a voltage collapse [50]. Countermeasures against these types of attacks typically rely on reputation-based models to determine which datapoints and/or system nodes can be trusted at any given time period.
Another important consideration is that of the cyber–physical cloud [51]. Figure 3 shows the generic process flow of this configuration. The growing complexity of CPS applications has meant that the procurement and maintenance of the required computing resources has become increasingly complex and expensive. By outsourcing some of these capabilities to cloud providers, the system can be scaled up and down as required to enable a more efficient use of resources. This can result in the significant reduction in the operational costs of the system [52]. The drawback of this approach is that it increases the attack surface of the system and makes the system more vulnerable to external attacks [53]. This approach does, however, also present an opportunity to provide increased security in the system. This can be conducted through the use of innovations such as the digital twin model which can not only aid with the detection, but also the mitigation of cyber-attacks [54].

![Cyber–physical cloud process flow](image)

Figure 3. Cyber–physical cloud process flow.

Figure 4 summarises the three key network security objectives as discussed in this section. In conventional IT networks, maintaining the network confidentiality can be considered the most important of the three network security goals. In microgrids, the most important goal is that of available as can be seen in the focus on the resilience of the system. Confidentiality is, however, still important, as discussed, but will typically rely on indirect mechanisms for detection. DoS attacks in the application environment are not too dissimilar from those in typical IT networks. This means the mechanisms for detection and mitigations on availability attacks will have a similar approach. Attacks on system integrity by an attacker with knowledge about the system will enable the attacker to launch stealthy attacks which can evade detection. This means that IDSs aimed and integrity attacks will need to be robust and take more application-specific contexts into account.
Confidentiality

- User Data should remain private
- Breach could pose a physical security risk
- Difficult to detect directly

Availability

- CPSs are time-critical systems
- Even small delays could be consequential
- Easiest to orchestrate but also easiest to detect

Integrity

- System relies on accurate data to make decisions
- Fabrication, modification or even replay attacks could be disastrous
- Attacker with knowledge of the system can evade detection

Figure 4. Summary of the three key network security objectives.

5. Microgrid IDPS

There have been a variety of different IDPSs proposed for Microgrids, in this section a few of them will be discussed. It has been mentioned previously, that a purely rule-based approach to intrusion detection would be inadequate in identifying previously unseen attacks. This is especially the case with rapid technological advancements that results in systems being upgraded with the latest hardware and software over time. The authors in [55], however, contend that the rule-based approach can be improved by combining it in stages with a deep learning approach in Microgrid applications. They propose using an unsupervised deep belief network (DBN) as the first stage. This feeds into rule-based second stage that determines the output via thresholding. The algorithm was proven to be effective at detecting both DoS and data injection attacks. It also outperformed more popular machine learning algorithms such as the convolutional neural network (CNN).

The authors in [56] also primarily focus on data injection attacks, but limit their analysis to the advanced metering infrastructure (AMI). As discussed previously, the prosumer model of the microgrid makes the AMI particularly vulnerable because of its location at the endpoint of the system. The proposed system uses the lower upper bound estimation (LUBE) method for feedforward neural networks to detect anomalies in the application environment. The algorithm determines the minimum and maximum values for data points within the system based on normal operating behaviour. Anything falling outside of these bounds is classified as an anomaly. Due to the complex nonlinear data of the application environment, the algorithm is reinforced with a modified version of the symbiotic organisms search (SOS) algorithm. This is conducted to improve the accuracy of the system. The proposed scheme was benchmarked against the conventional LUBE algorithm and one that was reinforced using the conventional SOS algorithm. It was found to outperform both across a variety of different performance metrics.

Data integrity attacks in the AMI were also considered by the authors in [57] using a similar approach. Instead of a feed-forward neural network, however, they proposed an improvement to the system by instead using a generative adversarial network (GAN) which is a deep learning algorithm. The proposed algorithm was also based on the LUBE method. To account for the complexity of the data, the authors instead propose reinforcing the scheme using a modified version of the teaching-learning-based optimization (TLBO) algorithm. As with the scheme proposed in [56], this algorithm was trained on normal system data to determine what normal behaviour is (upper and lower bound). Anything that deviates from this determined norm is classified as an intrusion. The algorithm was evaluated against a range of data injection attacks with varying severity and was shown to
be more effective than conventional LUBE algorithm and one that was reinforced using the conventional TLBO algorithm. The algorithm was also shown to outperform the support vector machine (SVM) algorithm in the application environment.

The authors in [58] identify these false data injection attacks as the most common type of attack in the application environment. It is thus no surprise that the bulk of the literature on this topic focuses on data integrity attacks. They distinguish between two types of data injection attacks. The first type have a large short term impact of the system but are easier to detect. The second type are stealthier but have longer term consequences if they go undetected. The authors propose the use of a deep learning based auto-encoder to detect both types of attacks in DC microgrids. The preprocessing stage makes use of a combination of the wavelet transform and singular value decomposition to extract the features required by the auto-encoders. In the next stage an ensemble of several deep auto-encoders are used and the result is determined using weighted voting scheme. The proposed algorithm was compared to a deep neural network and outperformed the algorithm in both classification accuracy and average detection time.

In [59], it is argued that traditional state vector estimation (SVE) algorithms are ineffective at detecting stealthy data integrity attacks. The authors demonstrate this by comparing a standard SVE algorithm with a variety of different machine learning algorithms. The algorithms were tested on data from a man-in-the-middle attack using the stealthy measurement as a reference (MaR) method. The results of the experiment are shown in Table 3 below and show that the machine learning algorithms significantly outperform the SVE algorithm. The table also illustrates that the choice of performance metric could affect the perceived performance of the algorithms. The authors use the accuracy as the main comparative metric but an analysis of the F1 score shows that the outcome would have been slightly different. This would have been even more pronounced had the dataset used reflected the real-world unbalanced data problem discussed previously. The metrics used to evaluate the algorithms are thus an important consideration and should not be neglected. A consideration of the application environment and the structure of the data will be paramount in determining which metrics would most accurately reflect the performance of the algorithms.

Table 3. SVE vs. machine learning algorithms [59].

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Accuracy</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random Forest</td>
<td>0.97</td>
<td>0.95</td>
</tr>
<tr>
<td>Ada Boost</td>
<td>0.96</td>
<td>0.93</td>
</tr>
<tr>
<td>KNN</td>
<td>0.96</td>
<td>0.93</td>
</tr>
<tr>
<td>Decision Tree</td>
<td>0.95</td>
<td>0.93</td>
</tr>
<tr>
<td>SVC polynomial</td>
<td>0.95</td>
<td>0.92</td>
</tr>
<tr>
<td>MLP</td>
<td>0.95</td>
<td>0.92</td>
</tr>
<tr>
<td>SCV RBF</td>
<td>0.94</td>
<td>0.90</td>
</tr>
<tr>
<td>Naïve Bayes</td>
<td>0.83</td>
<td>0.75</td>
</tr>
<tr>
<td>SVC Linear</td>
<td>0.8</td>
<td>0.79</td>
</tr>
<tr>
<td>SVE</td>
<td>0.53</td>
<td>0.35</td>
</tr>
</tbody>
</table>

Data integrity attacks can take on a variety of different forms and can also result in DoS attacks if carried out correctly. These types of data integrity attacks, including stealthy attacks whether the attacker attempts to evade detection, are considered in [60]. The authors proposed a time-sequenced intrusion detection framework that is based on machine learning for an inverter management system in a microgrid wind farm. Similar to the model proposed in [57], this scheme attempts to learn the normal system behaviour in order to identify when a system is exhibiting anomalous behaviour. A cyber-attack model demonstrating the potential devastating effects of the data integrity attacks in the system at this level is also developed and used to evaluate the proposed scheme. The scheme was benchmarked against the auto-encoder and a cluster-based technique and was found to
outperform both across all attacks. Unsurprisingly, the DoS attack was the easiest to detect while the replay attack had the lowest detection rate for all of the models.

These replay attacks thus require special attention as they are a specialised form of data integrity attack. An attacker who has access to authentic system packets would be able to use these strategically to either gain access to the system or to manipulate the system into achieve a desired goal. The authors in [61] propose using the hash-based message authentication code (HMAC) using the message digest algorithm (MD5) as the cryptographic hash function to prevent replay attacks in isolated smart grids. The system periodontally generates a random signal to be added to the message before applying the HMAC algorithm. In this way, once the time period elapses, the generated hash code would no longer be valid for the message. This, however, means that replay attacks are possible during the period of validity so the strength of this algorithm rests with the length of the period. This also shows that preventative mechanism alone would not be adequate in protecting the system against replay attacks.

The use of a random signal can also be used to improve the detection capabilities of proposed intrusion detection algorithms. The authors in [62] argue that a random signal can be used similar to how a watermark is used to prevent the unauthorised distribution of multimedia applications. The two main requirements for the application of this watermark are that (1) it should not affect the normal operation of the system and (2) an attacker should not be able to identify the watermark in the message. The proposed scheme was evaluated under steady-state conditions where an IDS was not capable of detecting replay attacks. It was shown that the introduction of the watermark improved the detection capabilities of the scheme. It was additionally shown that this could be achieved while using a watermark signal that is undetectable by the attacker as it would be masked by the system noise.

The schemes discussed so far are general purpose algorithms that distinguish between normal and anomalous data. This means that the schemes do not specifically classify the type of attack, even though they were tested for effectiveness at detecting the attack of focus. The authors in [63] argue that it is important to identify the class of attack in order to maintain the resilience requirement of microgrids. This is because once an attack is detected, the system is required to have measures in place to recover from this attack. Knowledge of the class of attack would thus allow more effective response mechanisms. They propose a framework for intrusion detection focussed on the microgrid central controller using CNN to both detect attacks and identify the class of attacks. The latter is more challenging than the former owing to the lack of available training data for each class of attack in the application environment. However, by separating the detection and classification components of the system, an attack will always be detected even if the classification of the attack is not possible. This means that the system will still be capable of detecting previously unseen attacks.

The ability to recover from an attack is particularly important for DoS attacks because availability is the most critical of the three network security objectives in the application environment. As mentioned previously, attacks on system availability, particularly in the network layer, are easier to detect than data integrity attacks. They are, however, more challenging to defend against making this a very active research field. More recently, machine learning has been proposed to mitigate the impact of DoS in microgrids. The authors in [64] propose a CNN-based multi-agent deep reinforcement learning algorithm for secondary frequency control and state of charge balancing in battery energy storage systems under DoS attacks in microgrids. The system uses an event driven approach that uses the signal-to-interference-plus-noise ratio (SINR) to detect DoS attacks and trigger the proposed mitigation strategy. The SINR was found to be a useful parameter in identifying DoS attacks as a higher SINR is directly proportional to packet loss in the network. The proposed scheme was shown to be effective at reducing the SINR and consequently the proportion of lost packets during real-time DoS attacks in the network.

Another major consideration in the application of IDSs is that these systems are not 100% accurate and are thus prone to have relatively high false positive rates. The cost of
not detecting an attack in the application environment is far greater than falsely detecting
normal data as anomalous. The resilience requirement of microgrids, however, requires
automated mitigation strategies once an attack has been detected. This means that these
false positives could negatively affect the normal operation of the system. The authors
in [65] propose a secondary frequency regulator designed to improve the resilience of
Microgrids to DoS attacks. The scheme was proven to be effective even when used with
IDSs which have high false detection rates. It was also shown to be more robust than
schemes that do not consider the possibility of the IDS producing false positives.

The discussion so far has been limited to intrusion detection at a system level and
not at a device level. As stated in the previous section, in order to detect device level
intrusions, a host-based IDS would be required. The authors in [66] propose a scheme
to detect malicious code injection in microgrid inverters. They argue that these inverters
are particularly vulnerable because they are consumer electronics and also not designed
with security in mind. Firmware updates can also be installed remotely, which increases
the likelihood that an attacker would be able to inject malicious code under the guise
of a firmware update. The authors propose using custom built hardware performance
counters (HPC) to generate the features required by machine learning models to detect
malicious code. The system was tested against DoS and data integrity attacks and compared
three popular machine learning algorithms: neural networks, decision trees, and random
forests. The baseline system had mixed results but drastically improved when the data
were balanced and principle component analysis (PCA) was used for feature extraction.
The performances of the different machine learning algorithms varied depending on the
configuration of the input data. This shows that the preprocessing stage of machine learning
algorithms has a direct impact on the performance of the algorithms and should thus be
considered seriously.

The discussion in the previous paragraph highlights three important things for host-
based intrusion detection using machine learning: (1) the algorithms need HPC generated
features, (2) the preprocessing of these HPCs will be paramount in how these algorithms
perform, and (3) the performance of the machine learning algorithms differ as the conditions
change (i.e., there is no one general purpose algorithm that can outperform all others in
all situations). These conclusions were highlighted by the authors in [67], who proposed
a generic framework from hardware malware detectors using machine learning. They
found that the number of HPCs that are available at runtime differ depending on the device
and so proposed schemes should limit the number of HPCs used. Their proposed scheme
uses a standard four HPC features which they found most adequately represented the
performance of the system under normal operating conditions. In order to improve the
results, the second stage of the scheme used four additional features specific to each kind
of malware and an ensemble learning technique using adaptive boosting. This allowed the
algorithm to not only detect malicious code, but also to classify the type of malware that
was detected. Experiments showed that the proposed scheme outperforms state-of-the-art
malware detectors that use a larger number of HPC features, even when compared only to
the first stage of the scheme that only uses four features.

Discussion

Table 4 shows a summary of all of the algorithms discussed in this section. There
are a few key observations that will be discussed briefly here. Firstly, it can be seen that
the bulk of the proposed schemes are intended to detect data integrity attacks. Even the
algorithms that include an evaluation of DoS attacks focus on those that result from data
integrity attacks. This is unsurprising as networked DoS attacks are generally easy to
detect but difficult to defend against. The same detection methods used in convention IT
applications would also be applicable to industrial control networks so research in this
area primarily focuses on the mitigation of attacks once they have been detected. The
second observation is that there is a heavy emphasis on data pre-processing methods across
all of the proposed schemes. Most schemes additionally also made use of deep learning
algorithms and multi-stage/ensemble techniques to improve the results. This is due to the complex nonlinear data that is found in the application environment. The choice of algorithm and preprocessing techniques was thus paramount in the performance of the proposed schemes as there is no one algorithm that can outperform the rest in all situations. The last observation is that the host-based techniques favoured hardware-based solutions over software implementations. This is due the resource constraints of the devices which would make the software implementations infeasible due to the computational costs. The same consideration for the system-wide IDSs are also applicable in this case once the computational costs have been accounted for. A summary of this discussion is shown in Figure 5.

Table 4. Microgrid IDPS.

<table>
<thead>
<tr>
<th>Paper</th>
<th>Method</th>
<th>Attack Detected</th>
</tr>
</thead>
<tbody>
<tr>
<td>Durairaj et al. [55]</td>
<td>Hybrid rule-based system using a deep belief network</td>
<td>Data injection and DoS</td>
</tr>
<tr>
<td>Kavousi-Fard et al. [56]</td>
<td>Feed forward neural network using LUBE method and modified SOS algorithm</td>
<td>Data injection attacks in AMI</td>
</tr>
<tr>
<td>Tang et al. [57]</td>
<td>Generative adversarial network using LUBE method and modified TLBO algorithm</td>
<td>Data injection attacks in AMI</td>
</tr>
<tr>
<td>Dehghani et al. [58]</td>
<td>Weighted voting ensemble deep auto encoder with wavelet transform and singular value decomposition</td>
<td>Data integrity attacks</td>
</tr>
<tr>
<td>Ma et al. [59]</td>
<td>Authors compare popular machine learning algorithms to traditional SVE algorithm</td>
<td>Stealthy data integrity attacks</td>
</tr>
<tr>
<td>Sadi et al. [60]</td>
<td>Time sequenced feed forward neural network</td>
<td>Stealthy data integrity attacks and DoS</td>
</tr>
<tr>
<td>Gallo et al. [62]</td>
<td>Water marking using random signals to improve detection of replay attacks in IDSs</td>
<td>Data integrity attacks</td>
</tr>
<tr>
<td>Xi et al. [63]</td>
<td>Framework for intrusion detection and attack classification using CNN</td>
<td>Data integrity and DOS</td>
</tr>
<tr>
<td>Chen et al. [64]</td>
<td>Event-triggered CNN-based multi-agent deep reinforcement learning</td>
<td>Detection and mitigation of DoS attacks</td>
</tr>
<tr>
<td>Liu et al. [65]</td>
<td>Secondary frequency regulator to improve resilience when IDSs have high false positive rates</td>
<td>Mitigation of DoS attacks</td>
</tr>
<tr>
<td>Kuruvila et al. [66]</td>
<td>Various machine learning models using PCA and custom built HPCs</td>
<td>Malicious code injection</td>
</tr>
<tr>
<td>Sayadi et al. [67]</td>
<td>a Framework for malicious code detection using a limited number of HPC features with various machine learning models</td>
<td>Malicious code injection</td>
</tr>
</tbody>
</table>

Data Integrity Attacks
- Most popular class of attack in the application environment
- Can lead to a compromising both confidentiality and availability
- Variety of machine learning algorithms proven to be effective at detecting this class of attack

Replay Attacks
- Special form of data integrity attack
- Requires additional data preprocessing to improve detection of existing algorithms
- Most difficult to detect using machine learning. Other methods more effective

Malicious Code Injection
- Special form of data integrity attack at a device level
- Lack of training data requires extensive data preprocessing to extract appropriate features
- Light-weight algorithms or hardware implementation preferred

Figure 5. Summary of Microgrid IDPS.
6. Conclusions

This paper considered the practical consideration of intrusion detection using machine learning in microgrid applications. The discussion focussed on a high level of abstraction in order to generalise the considerations across a different ICSs. As a case study, previous work focussing on water distribution systems was introduced and the lessons learned were generalised for the application of IDs to cyber–physical ICSs. The vulnerabilities specifically related to microgrid applications were then discussed with this in mind and the challenges for implementation were considered. Lastly, the state-of-the-art research was considered which supported the deductions from the previous sections. This is a growing and increasingly important field from both a research and an industrial perspective. Rapid advancements in the field spurred on by the fourth industrial revolution will require researchers to place a heavier emphasis on the practical considerations of proposed schemes for these to be viable in practice. Future work will thus need to consider frameworks for data preprocessing, model implementation, and algorithm evaluation in the application environment.

Author Contributions: Conceptualization, D.T.R. and G.P.H.; Funding acquisition, G.P.H.; Investigation, D.T.R.; Methodology, D.T.R.; Project administration, G.P.H. and A.M.A.-M.; Supervision, G.P.H. and A.M.A.-M.; Writing—original draft, D.T.R.; Writing—review and editing, G.P.H. and A.M.A.-M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

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. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

AMI Advanced Metering Infrastructure
CNN Convolutional Neural Network
DBN Deep Belief Network
DoS Denial of Service
GAN Generative Adversarial Network
HMI Human-Machine Interface
HPC Hardware Performance Counters
ICS Industrial Control System
IDPS Intrusion Detection and Prevention System
IDS Intrusion Detection System
IoT Internet of Things
LUBE Lower Upper Bound Estimation
MaR Measurement as a Reference
MCAS Manoeuvring Characteristics Augmentation System
PCA Principle Component Analysis
PLC Progammable Logic Controller
RTU Remote Terminal Unit
SCADA Supervisory Control And Data Acquisition
SINR Signal-to-Interference-plus-Noise Ratio
SOS Symbiotic Organisms Search
SVE  State Vector Estimation
SVM  Support Vector Machine
TLBO  Teaching-Learning-Based Optimization
WPAN  Wireless Personal Area Network

References


Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.