frontiers | Frontiers in Energy Research

TYPE Original Research
PUBLISHED 30 January 2025
DOI 10.3389/fenrg.2024.1494164

A deep reinforcement learning-based approach for cyber resilient demand response optimization

Ayush Sinha[1]*, Ranjana Vyas[1], Feras Alasali[2], William Holderbaum[3] and O. P. Vyas[1]

[1]Department of IT, Indian Institute of Information Technology, Allahabad, India, [2]Department of Electrical Engineering, Faculty of Engineering, The Hashemite University, Zarqa, Jordan, [3]School of Science, Engineering Environment, University of Salford, Salford, United Kingdom

OPEN ACCESS

EDITED BY
M. Premkumar,
Dayananda Sagar College of Engineering, India

REVIEWED BY
Ziming Yan,
Nanyang Technological University, Singapore
Lefeng Cheng,
Guangzhou University, China

*CORRESPONDENCE
Ayush Sinha,
✉ pro.ayush@iiita.ac.in

RECEIVED 10 September 2024
ACCEPTED 30 December 2024
PUBLISHED 30 January 2025

CITATION
Sinha A, Vyas R, Alasali F, Holderbaum W and Vyas OP (2025) A deep reinforcement learning-based approach for cyber resilient demand response optimization.
Front. Energy Res. 12:1494164.
doi: 10.3389/fenrg.2024.1494164

The contemporary smart grid infrastructure, characterized by its bidirectional communication capabilities between prosumers and utility organizations, has revolutionized the efficient execution of fine-grain computational tasks. Ensuring the uninterrupted delivery of power, even in the face of unforeseen contingencies, stands as a paramount concern for utility companies. Peak load forecasting, load balancing, and robust cyberattack detection and prevention mechanisms are integral components in achieving grid reliability. This research endeavors to advance peak load forecasting strategies and demand response optimization at the microgrid level, thereby enhancing grid reliability through the application of Deep Reinforcement Learning (DRL) techniques. Additionally, it investigates the ongoing threat of false data injection attacks. By synergizing these two critical investigations and implementing a novel framework and defense mechanism, this paper proposes a comprehensive approach to fortify the smart grid's reliability and security. The envisioned framework not only refines demand response (DR) optimization but also bolsters the grid's resilience in the face of the everevolving cyber threat landscape. The research outcomes showcase the practicality and effectiveness of the proposed framework, substantiated through extensive experimentation conducted on IEEE-3, IEEE-9, IEEE-14, and IEEE-33 bus systems.

KEYWORDS

smart grid architecture, load forecasting, demand response, load profiling, smart grid resilience, FDI attack

1 Introduction

The conventional design of the power network has advanced in sophisticated ways since its unique inception when a central framework regulated energy creation and distribution. The advent of innovations for Internet communication in this domain brought a shift toward a more interconnected, intelligent, and dynamic nature of the grid model, known as the Smart Grid (SG). Its fundamental advantage is two-way data communication, through which information can be exchanged between the client (i.e., a smart meter) and the power company, thus making it appropriate to play out a sophisticated power consumption metering (Mohassel et al., 2014). This allows the user to partake in programs that decrease power use when energy costs rise and allow the user him/her to sell the power produced at home (e.g., utilizing solar energy installations). This technology can also be leveraged by the power company to enhance the supply and demand of electricity by managing power generation and distribution in real-time, enabling power operators and administrators to anticipate periods of high demand and prevent scenarios of blackouts.

For this, the data collection is done through sophisticated advanced metering infrastructure (AMI) in aggregation with meter data management systems (MDMS). The data collection needs information technology-enabled industrial equipment. From one viewpoint, the power company is utilizing the supervisory control and data acquisition (SCADA) frameworks to deploy machines that continuously sense the energy generation and demand of numerous consumers. This incorporates, for instance, the programmable logic controllers (PLCs) and remote terminal units (RTUs) that are available in the substations spread over the wide area network (WAN) of the smart grid. From another viewpoint, support for the MDMS techniques involves interconnecting these modern resources with outside networks (e.g., the Web) and technical advances (e.g., distributed computing and the cloud) to go through additional information investigation and support demand response (DR).

The growing connection of SCADA systems that used to work separately has increased the number of online security risks, in this case, (Upadhyay and Sampalli, 2020). The main reason behind complex attacks are more likely to target multiple nodes in the control network over a long period of time. The presence of these attacks can harm the smart grid infrastructure and risk the accessibility of utility machines, which converts into scenarios responsible for holding the power supply and is likely to introduce power outages in the network (Romanenko et al., 2020). In a similar aspect, security measures should likewise be inducted to save the accessibility of the power supply in situations like high demand (that may likewise be incited on purpose), thus staying away from blackouts (Lopez et al., 2018).

In connection with the situation mentioned above, the reliability and security of the SG infrastructure are critical phenomena. It can be investigated by analyzing the resilience of SG (Singh and Govindarasu, 2020). The authors in (Clark and Zonouz, 2019) stated that the resilience of the SG focuses on (I) assurance for the full corrective measures of the core functionalities of the SG despite continuous ill-disposed mischievous activities and attacks. As a boundary condition, some non-core functionalities may be affected for the time being. (II) Ensured recovery of the crucial activity of the influenced sub-functionalities inside a predefined cost limit called the resilience limit. So, to analyze the stability of SG in terms of safety and security, it is imperative to not only study cyber security and DR but also explore the interdependence between them and how they contribute towards the resilience measure of SG. For this reason, the protection of SG infrastructure from such undesired actions with mischievous intentions is an emerging research area (Cybersecurity, 2018), for the government (UsEnergy. U)- (InEnergy), and international agencies like the National Institute of Standards and Technology (NIST) (Cybersecurity, 2018) and the European Union Agency for Cybersecurity (ENISA) (EuGovernment).

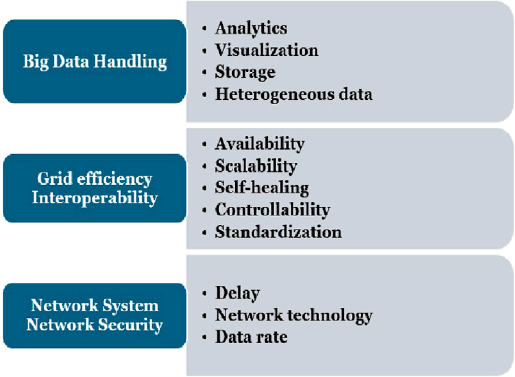

FIGURE 1
Smart grid reliability criteria.

There is a lack of academic research on smart grid reliability as it concerns to users, and further investigation on this subject is necessary (Balali et al., 2023) and (Bohra and Anvari-Moghaddam, 2022). To understand the criteria for judging the Reliability and Security Aspect of SG, the authors have proposed in (Mashal et al., 2023) as three main factors: (1) The "Network System" criterion is all about the needs and standards of the communication network system. (2) The "Big Data" criterion shows the features and traits of handling large amounts of data. (3) The "Grid efficiency" measure checks how well the smart grid works as presented in Figure 1.

In Mashal et al., (2023), authors proposed the problem of evaluating the reliability of smart grids as a Multiple Criteria Decision Making (MCDM) problem in order to investigate the elements that influence it. With the help of expert opinion and MCDA approach, the authors proposed a overall rank for criteria and subcriteria as mentioned in Figure 1. The below figure presents the rang and weightage of subcriteria, Figure 2. From this figure, it is evident that for Big Data handling criteria, Privacy and Analytics are two important sub-criteria, for Grid Efficiency point of view Interoperability, Availability and Self-Healing are an important sub criteria and finally for Smart Grid Network System aspect, Network Cyber Security and Delay are highly ranked sub-criteria. The present work is mostly based on the resilience mechanism defined by the NIST report (Ross et al., 2019). The main contributions of the paper are as follows:

1) Propose a framework for SG with demand response optimization and cyber events handling mechanism.
2) As mentioned above in Figure 2, the seven sub-criteria are explored through this work which are as Privacy, Analytics, Interoperability, Availability, Self-Healing, Cyber Security and Delay.
3) Present the design and execution of a detection mechanism for cyber events, thus ensuring the security of SG.
4) Address the wellbeing of the critical SG assets by carrying out a DR balance mechanism that allows a crucial energy supply for the whole SG, expediting the expectation of future utilization patterns.
5) The proposed framework demonstrate how the ideas of DR and cyber security with resilience are intrinsically related.

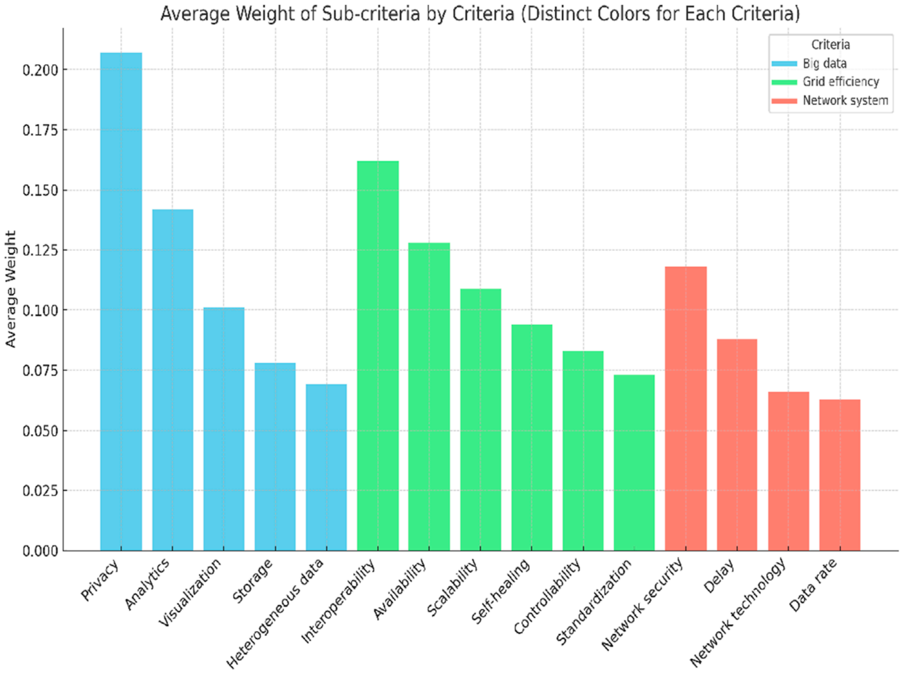

FIGURE 2
Global weights and ranking for sub-criteria -smart grid reliability.

The rest of the paper's structural flow follows: Section 2 describes the background and related work. It also describes the DR mechanism and the optimal method to ensure its smooth functioning, along with the relationship between DR and SG resilience. It also depicts the necessity of cyber security and its post-attack handling scenarios. Further, Section 3 presents the proposed framework as a solution for the problem formulated. To validate the effectiveness of the proposed framework, Section 4 describes the data used for the experiment, and further, Section 5 shows the usability of the DR and cyber security for the safety and security of SG from the resilience perspective. Finally, Section 6 is for conclusive remarks and future scope of the present work.

2 Background work

As per the NIST, the definition of the SG is the power delivery infrastructure based on integrating and amalgamating different smart computing and communication technologies with intelligent services. The ENISA also considers SG as an intelligent energy infrastructure with two-way communication capability for consumers and producers with smart components like Advance Metering Infrastructure (AMI). Throughout the world, the arrangement and activity of power infrastructure foundations are, by and large, dependent on security and sufficiency necessities. These principles permit the framework construction to withstand dangers to supply consumer requests with a great and negligible disruption throughout a period. Due to environmental change, the number and seriousness of natural disasters like tempests, droughts, and floods have been observed in many countries. In 2012, the northeastern territories of the USA were impacted by a hurricane that annihilated around 100,000 electrical wires. Around 7 million people were affected by a power cut as a result of this event. As per the authors (Panteli and Mancarella, 2015), the impact of severe weather events is expected to increase due to higher greenhouse gas concentrations. Such contingencies emphasize the urgency and importance of making the power grid smarter and more intelligent enough to withstand these catastrophic circumstances that also impact social life.

Along with natural disasters, cyber security is also one of the major concerns for SG's safe and smooth operation. In (Panteli and Mancarella, 2015), resilience is defined as "*the ability of a system to withstand, absorb, and rapidly recover from an external, high-impact, low-probability devastating event, like an extreme weather event or a cyber attack*". A resilient infrastructure can restore and recover from such a damaging situation within an acceptable time frame. Many researchers have defined the resilience concept from a critical infrastructure perspective. For example, authors in Mousavizadeh et al. (2018) defined resilience as the ability to recover and restore the system against extreme catastrophic events. The definition has covered both active and passive concepts. One of the important factors to consider while determining the reliability and stability of SG is the consideration of insider attack scenarios, as explained by (Singh et al., 2021). It is shown that if an attacker already has access to the system as an insider, this access can be used to launch attacks that are more difficult to detect and prevent. Another work by Cheng and Yu (2019) shows how the AI 2.0, driven by data, will speed up the growth of smart energy and electric power systems (Smart EEPS). In this version of AI, machine learning (ML) is a key method that analyzes large amounts of real and simulated

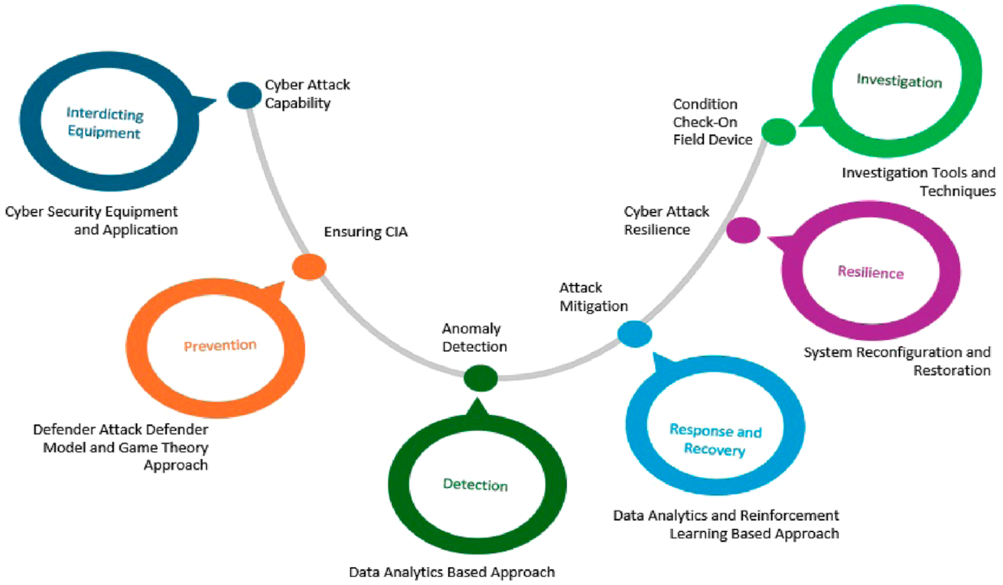
Interdicting Equipment
Cyber Attack Capability
Cyber Security Equipment and Application
Ensuring CIA
Prevention
Defender Attack Defender Model and Game Theory Approach
Anomaly Detection
Detection
Data Analytics Based Approach
Attack Mitigation
Response and Recovery
Data Analytics and Reinforcement Learning Based Approach
Cyber Attack Resilience
Resilience
System Reconfiguration and Restoration
Condition Check-On Field Device
Investigation
Investigation Tools and Techniques

FIGURE 3
Attack resilient smart grid functioning.

data to predict outcomes, make judgments, and help people make better decisions.

The SG infrastructure's intelligent communication and decision-making system components make its resilience more dependent on the underlying distribution network. For example, in Li et al. (2017), in normal mode, the SG may not have any Micro Grid (MG) formation. However, after disturbing events or partial blackouts, the same SG has one or more MG formations based on the resilience measures taken by the operator. With the proper utilization of DR in the MG system, load scheduling can be achieved efficiently by detecting anomalies in the system. Another research work in Fleschutz et al. (2021) highlights the price-based demand response (PBDR) system that is attributed to good economic and environmental aspects. With the analysis of carbon emissions based on the PBDR system, the authors have established that the PBDR system is good for economic and environmental aspects. The aim of another work, Chen et al. (2015), is to suggest a new method for managing power distribution systems during outages. The suggested strategy entails building numerous microgrids that are live-connected to the radial distribution system and powered by distributed generators (DG). This allows for the restoration of critical loads in a timely manner, thus making the SG more resilient (Figure 3).

In addition to the role of these advancements in day-to-day operator activity, they give greater adaptability to power grid utility in the extreme possibility conditions in which electrical lines are harmed or association with the upstream SG network is disturbed. This issue has constrained network operators to make an inescapable arrangement for the resilient operation of the SG in extreme conditions like technical issues, natural disasters, and man-made issues that cause irrecoverable losses. Hence, the occurrence of severe contingency conditions is a prominent issue. Consequently, advancing an appropriate procedure to decrease the adverse consequences of this issue on the SG network has become vital. Up to now, considerable research has been done in the context of both normal and contingency situations of SG. Therefore, the present section represents the resilience measures of SG as per the NIST (Ross et al., 2019) framework, which is as follows:

1) Adaptive response
2) Segmentation
3) Redundancy
4) Diversity
5) Deception

For more detailed explanation, the present section has been divided in to two segments as SG Resilience in context to Cyber Resilience and Optimized Demand Response, respectively.

2.1 SG resilience: cyber resilience

NIST has published a report with a special focus on cyber resilience, "*Developing Cyber Resilient Systems: A Systems Security Engineering Approach*", (Ross et al., 2019). It includes different verticals: adaptive response, segmentation, redundancy, diversity, and deception. Each of these (Figure 4) has standard procedures and best practices with the objective of making attack-resilient systems. A more detailed explanation of the NIST resilient mechanism is as follows:

1) Adaptive response: This method entails a prompt and suitable response to a cyberattack by changing specific system elements to change their functionality or adjust the resource allocation. The system must continue to function while these changes are implemented.
2) Segmentation: It prioritizes activities and resources based on their importance and reliability to identify and secure the most

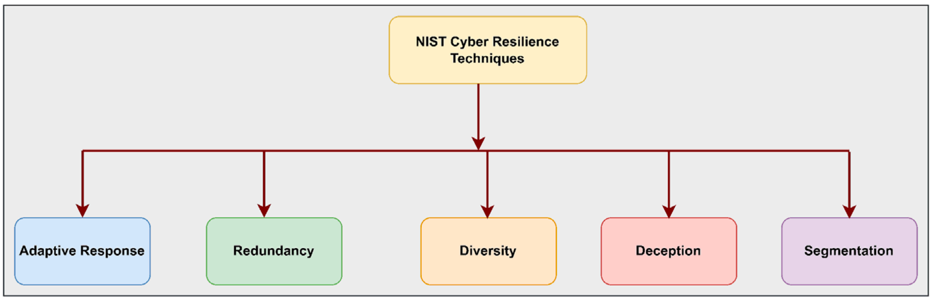

FIGURE 4
NIST Cyber Resilience technique (Ross et al., 2019)

attractive or susceptible ones. Segmentation can happen either manually or automatically while the system is running.

3) Redundancy: It embraces the existence of numerous, secure instances of critical components, including hardware, data, and functions (referred to as "replicas"), eliminates single points of failure, and enables the system to continue functioning even after a successful cyber attack. The retention of additional, alternate communication resources is another definition of redundancy. At this point, replicas must stay in sync.
4) Diversity: This strategy uses heterogeneity in terms of architecture, design, or technology to make it more difficult for attackers to take advantage of widespread vulnerabilities.
5) Deception: It is carried out by concealing crucial resources, knowingly disseminating false information, or leading attackers in the wrong direction to ripoffs of the genuine system components. Even when they have gained in, it may be able to stop them from seriously harming the system.

Another factor important for the resilience of SG, as analyzed in Adepu et al. (2020), is weaknesses in both the network infrastructure and the processes that control the smart grid. It was explained that these overlooked, common vulnerabilities can be effectively used to attack smart grids. This means that they showed that even vulnerabilities that are known to exist can be used to attack smart grids if they are not properly addressed. The authors in Adepu et al. (2020), specifically mentioned that distribution systems with multiple energy sources are particularly vulnerable to attack. This is because these systems are more complex and therefore have more potential vulnerabilities. Another important contribution towards the cyber resilience of SG has been mentioned in a special publication of NIST (Ross et al., 2019) and pointed out that SG are vulnerable to cyber attacks due to their use of heterogeneous communication technologies and their distributed nature. Further work on a similar line is being done by Syrmakesis et al. (2022) and shows that while preventing or detecting cyber attacks is a well-studied field of research, making SG more resilient against such threats is a challenging task. The article, (Babar et al., 2020), presents the implementation of a safe demand-side management system in the smart grid. This system utilizes machine learning and IoT techniques to accurately identify dishonest entities within the grid. Authors in Tebekaemi and Wijesekera (2019), introduced the secure overlay communication model as a means to distribute the operation and control of smart grids. This model includes a technique for detecting attacks that modify data. In another work as e Sousa et al. (2022), the authors addressed the identification of load-altering attacks, which have the potential to disrupt network stability, by employing linear matrix inequality optimization techniques. The study conducted by Srivastava and Parida (2022) focuses on identifying and isolating potential problems in AC microgrids using a machine-learning technique. The ramifications of injecting false data on the functioning of intelligent power distribution networks have been examined in the study conducted by Cao et al. (2022). A brief reason of deep reinforcement learning (DRL) suitability for false data injection (FDI) attack detection compared to traditional methods are as:

1) Adapts to Changing Conditions: Smart grids are constantly changing, with different power demands, weather conditions, and operational challenges. Traditional detection methods often rely on fixed thresholds or static patterns, which can struggle to keep up with these changes. DRL, on the other hand, learns and adapts in real-time. It can adjust to new situations as they happen, making it better suited for environments like smart grids where things are always shifting. This adaptability makes DRL more resilient to new, unexpected types of attacks.
2) Tracks Attacks Over Time FDI attacks usually do not happen all at once; attackers often inject false data in small doses over time to gradually influence the system. DRL is great at handling these kinds of "sequential" tasks. It considers the long-term impact of each data point, so it is able to detect patterns or small, ongoing changes that might indicate an attack in progress. Traditional methods might miss these subtle, accumulating signs of trouble since they often analyze data in isolated snapshots.
3) Learns Complex Patterns Automatically The data in smart grids can be pretty complex—things like voltage, power flows, and load data are all interrelated, and it is hard to manually define features to capture every subtle pattern or anomaly. DRL uses deep neural networks to automatically learn relevant features from this data. It is like having a model that can see subtle signs of unusual behavior without needing an expert to pre-program every possible pattern. This ability to learn what matters directly from the data gives DRL an edge in spotting the tricky, hidden signs of an FDI attack.
4) Fast, Real-Time Detection Traditional detection methods can sometimes be slower or less responsive. They may need time

to process batches of data or use fixed rules that might not respond fast enough. DRL models are designed to make quick decisions. They're built for real-time detection, so as soon as they spot something suspicious, they can flag it. This is especially important in large smart grids, where attacks need to be detected and stopped quickly to prevent damage.

5) Learns to Recognize Harmful Attacks Not all FDI attacks are equally harmful. Some might cause minor disruptions, while others could lead to serious issues like blackouts. DRL can be trained to recognize the attacks that are most dangerous. By using a reward system where it "learns" to avoid actions that lead to instability, DRL models become better at prioritizing serious threats over harmless anomalies. This way, it can focus on detecting the attacks that really matter.
6) Handles New, Evolving Attack Strategies Traditional methods often rely on known patterns or signatures, making them good at detecting familiar attack types but less effective against new or modified ones. DRL, however, can generalize better to new types of attacks. By training on a wide range of scenarios in a simulated environment, it learns what normal and abnormal behavior look like, even if it has not seen a specific attack before. This makes it more robust and able to catch innovative or evolving attack strategies.
7) Can Keep Improving Over Time Traditional detection methods are usually static—once trained, they do not change unless they're retrained on new data, which can be a lengthy process. DRL can be set up to keep learning continuously, adapting as new data comes in. This ongoing learning process helps it stay effective as the grid evolves, whether due to seasonal changes, new infrastructure, or changing consumer behaviors.

2.2 SG resilience: optimized demand response

Because of the new advancements in the modernization of the power grid framework, distributed energy resources (DERs) are presently essential in providing DR interest in various conditions. It needs cross-functional arrangements that speed up the coordination of DERs and help the organization administrator optimize SG operations. Notwithstanding the DERs, there are more appealing and moderate choices that make the present SG frameworks more intelligent than traditional networks. One of the popular options is distribution network reconfiguration (DNR). Despite having been presented a while ago, the concept of DNR is now considered to be a versatile solution in the process of modernizing SG frameworks (Arasteh et al., 2018). The DNR is characterized as the way toward changing the situation with regularly open/shut switches of the distribution network to arrive at an arrangement that enhances the objective while fulfilling all functional planning constraints of the SG without discarding any SG infrastructure network node(s) (Paterakis et al., 2015). Many researchers have handled the contingency situation, which arises due to disturbances in SG, with a different mechanism. For example, authors in Gholami et al. (2016) proposed using fuel in plug-in electric vehicles (PEV) as an alternative resource to combat partial blackout situations. Due to the important part that Distributed Generation (DG) and energy storage systems (ESS) play in the power system, many studies have been conducted to find ways to include these DERs in the SG framework under different circumstances. Authors in Nikkhah and Rabiee (2018) proposed a constraint on voltage stability for effectively managing wind power as an alternate energy resource. Further in another work Pilz et al. (2020), authors have investigated the impact of false data injection attacks on smart grids and designed a security game to help utility companies choose the best strategies. They finally stated that the security game can help utility companies choose the most appropriate monitoring and defense strategies so that false data injection attacks have only a limited, if any, impact on smart energy scheduling. The taxonomy of existing research for the Study of Resilience with Respect to DERs and DNR is presented in Table 1.

2.3 Research gap and motivation

Price attacks and energy theft or its parameter (voltage, current, and phase angle) manipulation are the two main types of cyberattacks that try to mess up SG-related DR strategies. For instance, attackers can mess up power distribution systems by posting fake energy prices that are less than the real ones through the Internet or social networks (Tang et al., 2019; Tang et al., 2018; Tang et al., 2019b). People who get false information about low electricity prices will probably use more as a smart reaction to the chance, which will likely cause a sudden (partial) load increase in the power system. Then, the quick rise in demand may lead to a peak load or even an overload on the power grid. Energy theft attacks are another type of attack. In these attacks, one or more customers in the power system are the thieves who try to make money by changing the data sent to the utility companies about either generation or usage (Amin et al., 2015; Esmalifalak et al., 2014). Attacks like these could make the people who do them money, but the energy companies would lose money. We look at a new kind of framework that combines false pricing attacks and energy parameter manipulation attacks in the context of SG security and reliability.

3 Attack resilient smart grid reference architecture

For the power grid to work reliably, there needs to be a full and all-encompassing cybersecurity framework that includes attack prevention, attribution (forensics), detection, prevention, restoration, and resilience for the smart grid. This framework needs to cover the physical, application, information, and infrastructure domains. Computerized reasoning and AI-based intrusion avoidance and recognition frameworks are the best strategies for cyber event identification, classification, and lessening its effect in SG. These arrangements fabricate a keen, adaptable, secure, resilient, and versatile cyber-physical smart grid infrastructure (Zeadally et al., 2020). The existing research summary for the architecture attributed to the automatic protection of SG is presented in Table 2. Table 2 encapsulates the coverage of different aspects like fault detection, network reconfiguration, demand response, stability and robustness, and regulatory policy which are integral constituents of the SG architecture.

Through this work, we proposed a resilient framework, considering the microgrid concept as the backbone of the SG

TABLE 1 Taxonomy of previous research for the study of resilience with respect to DERs and DNR.

References number	DERs consideration			DERs allocation			DR	Contingency
	DG	ESS	PEV	DG	ESS	PEV		
Ahmadi et al. (2019)	No	No	No	No	No	No	No	No
Nikkhah et al. (2020)	Yes	No	Yes	No	No	Yes	No	No
Home-Ortiz and Mantovani (2020a)	Yes	No	Yes	No	No	Yes	No	No
Gao et al. (2020)	No	No	Yes	No	No	Yes	Yes	No
Home-Ortiz and Mantovani (2020b)	Yes	No	Yes	No	No	Yes	No	Yes
Han et al. (2020)	Yes	No	Yes	Yes	No	Yes	No	No
Nikkhah et al. (2021)	No	No	Yes	No	No	Yes	No	No
Nikkhah et al. (2021)	No	No	Yes	No	No	Yes	No	No
Gholami et al. (2016)	Yes	Yes	Yes	No	No	No	No	Yes
Nick et al. (2017)	Yes	Yes	No	No	Yes	No	No	No
Nikkhah and Rabiee (2018)	Yes	No	No	Yes	No	No	No	No
Awad et al. (2015)	No	Yes	No	No	yes	No	No	No
Vahidinasab (2014)	Yes	No	No	Yes	No	No	No	No
Rabiee et al. (2018)	No	No	No	No	No	No	No	No
Ding et al. (2017)	No	No	No	No	No	No	No	No
Lin and Bie (2018)	No	No	No	No	No	No	Yes	Yes
Sharifi et al. (2017)	Yes	No	No	No	No	No	No	Yes
Aghaei et al. (2016)	Yes	No	No	No	No	No	Yes	Yes

infrastructure. A microgrid is a limited-scale and self-dependent power distribution framework. It comprises RES and ESS and is equipped with facilitated control techniques. Loads inside a microgrid can be upheld by its neighborhood distributed generators persistently, which facilitates the MG to be detached from its upstream or parent node microgrid during blackout events or contingencies (Zhang et al., 2019; Wang et al., 2015). These features contribute to maximizing the resilience of SG. A resilient SG framework must be equipped to withstand, expect, and react to extreme or unprecedented events (Wang and Wang, 2015). Though the self-adequate microgrid offers several benefits to SG, more emphasis must be placed on analyzing resilience from a DR and cybersecurity perspective. The proposed framework as presented in Figure 5 represents two components: DR optimization and Cyber event handling and both are explored through a data-driven approach.

The proposed framework's operational components are as:

1) Data Anomaly is detected and SG operational strategy component (having Cyber Event Detection and Demand Response management) is invoked.
2) The demand Response module is based on effective and accurate electrical load forecasting.
3) Based on the electrical load forecasting, the profit (DR) optimization is achieved for all stakeholders (producer and consumer). Our Previous work (Sinha et al., 2021) and (Holderbaum et al., 2023) supports the electrical load forecasting component with detailed experimentation and validation of the proposed DR mechanism.
4) Cyber Event detection is mainly for power fault detection and classification followed by FDI attach detection mechanism.
5) As part of our previous research work, we have done the power fault detection and classification (Sinha et al., 2022).
6) For the FDI attack detection component, we proposed a Reinforcement Learning-based detection mechanism.
7) Finally, the last component of the proposed framework is supported by our previous work on smart grid restoration mechanism (Sinha et al., 2020).

The next two subsections discuss these two aspects in detail.

TABLE 2 Previous Research for the Architecture attributed to Automatic Protection of Smart Grid.

References number	Fault detection	Network Reconfiguration	DR	Stability and robustness	Regulatory policy for RE
Bhattarai et al. (2015)	Y	N	N	N	N
Habib et al. (2017)	Y	N	N	N	N
Shih et al. (2017)	Y	N	N	N	N
Momesso et al. (2020)	Y	N	N	N	N
Ma et al. (2018)	Y	N	N	N	N
Liao et al. (2019)	Y	N	N	N	N
Tummasit et al. (2015)	Y	Y	N	N	N
Sampath Kumar et al. (2018)	N	Y	N	N	N
Muda and Jena (2017)	Y	Y	N	N	Y
Mahat et al. (2011)	Y	Y	N	N	Y
Ibrahim et al. (2016)	Y	Y	N	N	N
Nascimento et al. (2020)	Y	Y	N	N	N
Papaspiliotopoulos et al. (2015)	Y	Y	N	N	N
Tielens and Van Hertem (2016)	N	Y	N	N	Y
Arani and El-Saadany (2012)	N	Y	Y	N	Y
Alipoor et al. (2014)	N	Y	N	N	Y
Soni et al. (2013)	N	Y	N	N	Y
Zhang and Chi (2015)	N	N	Y	Y	N
Cohenpb and Charles (1985)	N	N	N	Y	N
Allesina and Tang (2012)	N	N	Y	Y	N
Gribble (2001)	N	N	Y	Y	N
Long et al. (2017)	N	N	Y	Y	N
Morstyn et al. (2018)	N	N	Y	Y	N
Liu et al. (2017)	N	N	Y	Y	N
Korjenic and Bednar (2011)	N	N	Y	Y	N
Szulecki et al. (2015)	N	N	Y	Y	N

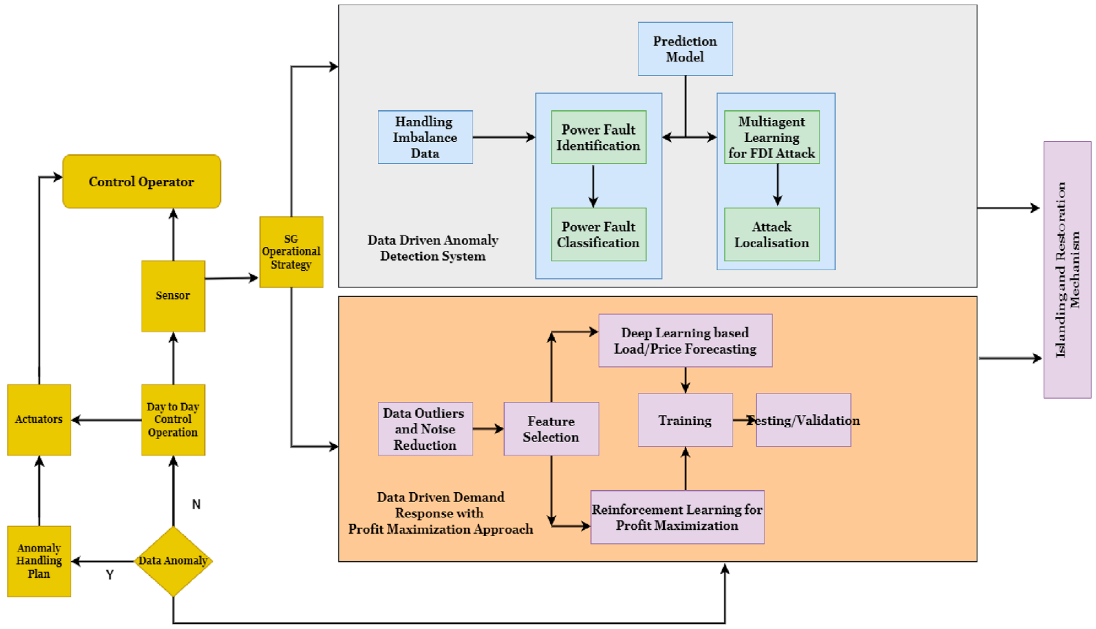
Prediction Model
Handling Imbalance Data
Power Fault Identification
Multiagent Learning for FDI Attack
Power Fault Classification
Attack Localisation
Data Driven Anomaly Detection System
Control Operator
SG Operational Strategy
Sensor
Islanding and Restoration Mechanism
Deep Learning based Load/Price Forecasting
Actuators
Day to Day Control Operation
Data Outliers and Noise Reduction
Feature Selection
Training
Testing/Validation
N
Data Driven Demand Response with Profit Maximization Approach
Reinforcement Learning for Profit Maximization
Anomaly Handling Plan
Data Anomaly
Y

FIGURE 5
Attack resilient smart grid architecture.

3.1 Resilience for smart grid demand response

Because of the techno-monetary problems of the extension of existing distribution infrastructure, DERs could be a successful way for electricity delivery to customers with minimized active power loss and load shedding. However, conventional Distributed Network Reconfiguration (DNR) models neglect to adjust to the imperatives and constraints introduced by new SG network advances. Considering the reasons mentioned above, the adaptation of an extensive coordinated model in which an optimal activity model for DR is vital, which is more likely to bring in the resilient operation of the grid infrastructure (Table 1). In the proposed framework, as presented in Figure 5, the resilience aspect with context to DR has been divided into two parts: (i) efficient and accurate load forecasting and (ii) optimization of profits among multiple stakeholders for distributed microgrid infrastructure. A detailed explanation of both aspects is as follows:

3.1.1 Efficient and accurate load forecasting

In our previous work, VAR-CNN-LSTM (Sinha et al., 2021), and (Holderbaum et al., 2023) we proposed a model based on Deep Learning (DL) technique to accurately forecast the next 6-hour electrical load. In load forecasting, historical data is considered time series data. It has linear and non-linear components (Equation 1).

$$d_t = N_t + L_t + \epsilon \tag{1}$$

where L_t is a linear component at time t, N_t is a component which is a non-linear component at time t and ϵ is the error component. A hybrid model called VAR-CNN-LSTM is proposed in this work. To handle the linear component the Vector Auto Regression (VAR) is used. The mathematical notation of the time series with A typical Auto Regression with order ‘p’ can be formulated as (Equation 2).

$$Y_t = \alpha + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \cdots + \beta_p Y_{t-p} + \varepsilon \tag{2}$$

where α is a constant denoting the intercept, $\beta_1, \beta_2, \ldots, \beta_p$ are lag coefficients. After making time series data stationary, the VAR model will do the forecasting task and the residual of this is fed as an input to the deep learning part that is CNN-LSTM. The resultant vector from k^{th} convolutional layer is formulated as (Equation 3).

$$y_{ij}^{l} = \sigma\left(b_j^l + \sum_{m=1}^{M} w_{m,j}^{l} x_{i+m-1,j}^{0}\right) \tag{3}$$

b_j^l represents bias for j^{th} feature map, y_{ij}^l is calculated by input x_{ij}^0 from previous layer, σ denotes the Rectified Linear Unit (ReLU) (Nair and Hinton, 2010) like activation function and w is the kernel. After the convolutional operation, the LSTM is used at a lower layer as it stores the temporal information well in advance from the features extracted by CNN layer.

3.1.2 Load profiling at microgrid level

DR scenarios for power distribution are getting attention as energy demand continues to grow. Their importance is set to grow consistently throughout the years before the Smart Grid (SG) foundation. DR programs attempt to support prosumers to use uninterrupted supply and decrease their consumption usage during peak hours, which would eventually support microgrid administrator changing of DR and draw profit by selling the amount of generated power to the SG. Though various research works think of utilizing DR systems, most of them revolve around a model based on single-agent electricity costs as a variable independent of climate. However, we recognize an urgency to analyze and support learning toward working with a multi-agent model that can enhance the DR process when power prices are administered through their respective demands. Our methodology is centered around utilizing price as a sign that will influence the adjustment of demand and subsequently optimize the DR reaction.

We suggested a way to use the Asynchronous Advantage Actor-Critic model to create the agent and a setting that uses VAR-CNN-LSTM (from our earlier work) to mimic the real-life situation (Mnih et al., 2016). In our A3C model, we have a master agent who is responsible for the decision-making based on the current state of the environment, and we have worker agents whose sole responsibility is to explore and update both policy and value networks asynchronously, which are common to all worker agents. The list of symbols used for the A3C algorithm is presented in Table 3.

TABLE 3 List of symbols used for A3C algorithm.

Symbol	Meaning
$S \in \mathcal{S}$	States
$R \in \mathcal{R}$	Rewards
$A \in \mathcal{A}$	Actions
S_t, A_t, R_t	State, action, and reward at time step t of one trajectory
G_t	Return; or discounted future reward; $G_t = \sum_{k=0}^{\infty} \gamma^k R_{t+k+1}$
γ	Discount factor; penalty to uncertainty of future rewards; $0 < \gamma \leq 1$
$P(s', r \mid s, a)$	Transition probability of getting to the next state s' from the current state s with action a and reward r
$V(s)$	State-value function measures the expected return of state s; $V_w(.)$ is a value function parameterized by w and θ_v is the parameter to the value function
$\mu(s)$	Deterministic policy; we can also label this as $\pi(s)$, but using a different letter gives $\mu(s)$ better distinction so that we can easily tell when the policy is stochastic or deterministic without further explanation. Either π or μ is what a reinforcement learning algorithm aims to learn
$\pi(a \mid s)$	Stochastic policy (agent behavior strategy); $\pi_\theta(.)$ is a policy parameterized by θ
$A(s,a)$	Advantage function, $A(s,a) = Q(s,a) - V(s)$; it can be considered as another version of Q-value with lower variance by taking the state-value off as the baseline
$Q(s,a)$	Action-value function is similar to $V(s)$, but it assesses the expected return of a pair of state and action (s,a); $Q_w(.)$ is a action value function parameterized by w
$V^\pi(s)$	The value of state s when we follow a policy π; $V^\pi(s) = \mathbb{E}_{a\sim\pi}[G_t \mid S_t = s]$
$Q^\pi(s,a)$	Similar to $V^\pi(.)$, the value of (state, action) pair when we follow a policy π; $Q^\pi(s,a) = \mathbb{E}_{a\sim\pi}[G_t \mid S_t = s, A_t = a]$
n	varies from state to state and it's maximum value is t_{max}
θ	parameter to the policy

Working of Worker network: Worker agents, as in Algorithm 3 and 4, are created by the master agent that is responsible for the exploration and updation of the policy and value networks. The work of worker agents can be divided into the trajectory calculation and updation of the networks.

3.1.2.1 Calculation of trajectory

A trajectory is a path that the agent takes through a state, action and reward space. The length of the trajectory can vary and be set. Consider T to be the trajectory length that is set. It is assumed that every worker agent has a copy of the current state of the environment upon which they explore.

(1) First, the agent observes the current state of the environment at a given time t, S_t and this is given to the policy network to generate a probability distribution $\pi_{\theta A}(A_t \mid S_t)$.
(2) We create a categorical probability distribution with respect to the probability distribution function generated by the policy network that helps in sampling random action.
(3) Upon taking an action A_t, the agent observes the next state S_{t+1} and reward R_t.
(4) The agent stores (S_t;A_t; R_t) tuple and repeats the above steps till trajectory length T.

This process generates a trajectory for each worker agent, which is different for every agent because of the random actions chosen to explore the environment.

3.1.2.2 Updating policy and value networks

Before updating the policy and value networks, the worker agents calculate each tuple's advantage value, each state's target value, and the loss value of policy and value networks for considering the entire trajectory. The advantage value is calculated for each tuple present in the trajectory by using the n-step method for every tuple in the trajectory (Equation 4).

$$Ad(S_t, A_t \mid \theta, \theta_v) = \sum_{i=0}^{n-1} \gamma^i r_{t+i} + \gamma^n V(S_{t+n} \mid \theta_v) - V(S_t \mid \theta_v) \quad \forall t \in \{0, \ldots, T\} \tag{4}$$

Due to the iterative nature of our process, it is not feasible to utilize the cumulative rewards R(t) at each time step. In order to proceed, it is important to develop a Critic model that can effectively estimate the value function. The agent calculates the total reward G_t which can be viewed as a sum of flat partial returns as (Equation 5).

$$G_t = \sum_{k=0}^{\infty} \gamma^k R_{t+k+1} \tag{5}$$

For any two values within the interval $\gamma \in [0, 1)$, we can conceptualize the return G_0 as having partial termination in one step, resulting in a degree of (1 - γ) and yielding only the first reward, R_1. Additionally, it can be seen as partially terminating after two steps, with a degree of (1 - γ)$^*\gamma$, resulting in a return of $R_1 + R_2$, and so forth. Finally, before updating the policy and value networks which are represented by θ_A and θ_C, we calculate the loss of both policy and value networks.

Loss of policy network is calculated by the given (Equation 6).

$$L_{policy}(\theta_A) = \frac{\sum_{t=0}^{T} (G_t(S_t) - V(S_t))^2}{T} \tag{6}$$

The loss of value network is calculated by the given (Equation 7).

$$L_{value}(\theta_C) = \frac{\sum_{t=0}^{T} \left(-Ad(S_t, A_t)\log\left(\pi_{\theta_A}(A_t \mid S_t)\right)\right)}{T} \tag{7}$$

where, $Ad(S_t, A_t)$ is the advantage function and $\pi_{\theta A}(A_t \mid S_t)$ is the probability distribution of an action given a state at time t given by policy network.

The policy and value networks are updated in the following way, considering α_{θ_A} is the learning rate of the policy network (actor), α_{θ_C} is the learning rate of value network (critic). Policy network (Equation 8).

$$\theta_A = \theta_A + \alpha_{\theta_A} \nabla_{\theta_A} L_{policy}(\theta_A) \tag{8}$$

Value network or Critic network (Equation 9).

$$\theta_C = \theta_C + \alpha_{\theta_C} \nabla_{\theta_C} L_{value}(\theta_c) \tag{9}$$

After the exploration is done by the worker agents and the networks are updated, the master agent based on the current environment state takes the best suitable action to maximize the overall reward.

3.1.2.3 Action Space

The steps of the action function are defined as:

(i) Verify that the action is legal.
(ii) Send the history of environment states and calculate the next state (also including the current environment state) to the LSTM.
(iii) Compute the new price based on the effect of the action.
(iv) Set the just-calculated price as the price of the next state (the new price).
(v) Based on the demand and supply values of the *next state* and the new price, compute the non-normalized reward.
(vi) In the historical record of environment states, add the next state to it.
(vii) Return the value of the non-normalized reward and the next state to the agent.

3.1.2.4 Reward

Based on Algorithm 2, the reward function is formulated with the following goals as:

(i) Ensure that the demand is always more than supply in order to ensure that the producer makes a profit instead of paying back to the consumers to consume electricity.
(ii) Ensure that there is always a buffer present for demand. It will ensure that abrupt changes in demand or supply will not impact and decrease the producer's profitability by a huge amount.
(iii) In order to avoid a long-term reduction in demand, make sure that the price of electricity is not extremely high.

```
Initialize the policy parameter $\theta$ at random
Generate one trajectory on policy
 $\pi_\theta : S_1, a_1, R_2, S_2, a_2 \ldots, S_t$;
for $t = 1$ to $T$ do
  Estimate the return G
  Update policy parameters:
  $\theta \leftarrow \theta + \alpha\gamma^t G_t \nabla_\theta \log(\pi_\theta(a_t)|S_t)$
end
```

Algorithm 1. REINFORCE.

```
Initialize the hyperparameters maxAllowedPrice and
minAllowedPrice
correction ← 1
if ((demand − supply) < 0) or
 (newPrice < minAllowed) or
 (newPrice > maxAllowed):
correction = 0-abs (correction)
reward ← (mod(demand-supply)$^3$)*(abs(newPrice$^2$))*
correction
profit ← (demand − supply)*newPrice
return reward
```

Algorithm 2. Reward Function.

So, the formula to compute the reward value is as:

$$reward = |(demand - supply)^2| * |newPrice^3| * correction$$

However, the correction is based on the following points:

- If *newPrice* is within a certain limit or bounds.
- If *demand* − *supply* or *newPrice* is negative.

The value of x is set to 3 and y is set to 2 to satisfy the constraints mentioned above. The variable *newPrice* is the non-normalized value, as in Algorithm 5 of the price the action has given to the environment. For this function, the non-normalized values of demand and supply are used. Since we are using the demand and supply values of the same time step as that of the *newPrice*, we are therefore using the demand and supply values of the next time step, that is, the one that will be returned alongside the reward, to calculate the reward. Here, the *correction* is a factor that is simply used to ensure that the *newPrice* lies within limits or bounds defined earlier, eventually ensuring no exploding or vanishing price problems. This happens by recalculating and modifying the reward to punish the agent for getting too far out of bounds.

3.2 Resilience for smart grid cyber security

The second aspect of the proposed resilient framework (Figure 5) is cyber event handling with an emphasis on FDI attacks. For this, the work proposed a deep learning-based algorithm to detect the ongoing false data injection attack (FDI). The proposed method is divided into six sections:

```
Global parameters:-$\theta$, w
Initialise thread-specific parameters: $\theta'$ and $w'$
Initialize time step $t = 1$
while $T \leq T_{max}$ do
  Reset gradient: $d\theta = 0$ and $dw = 0$.
  Synchronize thread-specific parameters
with global
   ones: $\theta' = \theta$ and $w' = w$.
  $tstart = t$ and sample a starting state $S_t$.
  while $(st\ != TERMINAL)$ and $t - t_{start} \leq t_{max}$ do
   Pick the action $A_t \sim \pi_{\theta'}(A_t|S_t)$ and receive a
    new reward $R_t$ and a new state $S_{t+1}$.
   Update $t = t + 1$ and $T = T + 1$
  end
  Initialize the variable that holds the return
   estimation
     $R = \begin{cases} 0, & S_t = \text{TERMINAL} \\ V_{w'}(S_t), & \text{otherwise} \end{cases}$
  for $i = t - 1, \ldots, t_{start}$ do
   $R \leftarrow \gamma R + R_i$; here R is a MC measure of $G_i$.
   Accumulate gradients w.r.t.:
    $d\theta \leftarrow d\theta + \nabla_{\theta'} \log \pi_{\theta'}(A_i|S_i)(R - V_{w'}(S_i))$;
   Accumulate gradients w.r.t. w':
    $dw \leftarrow dw + 2(R - V_{w'}(S_i))\nabla_{w'}(R - V_{w'}(S_i))$.
  end
  Update asynchronously $\theta$ using $d\theta$, and w using $dw$
end
```

Algorithm 3. Asynchronous Advantage Actor-Critic (A3C) Offline + Online (Episodic).

3.2.1 Simulating cyber attacks

We simulate the attack by dividing the simulation into episodes. In each episode that spans 500 timesteps, the attack starts at a random time step (t). We train our reinforcement learning model on several episodes and use a predetermined reward system to compute the reward and take the appropriate action.

3.2.2 Predict agent action

In each episode, at each step, the neural network model receives the state value and predicts two values: the estimated reward for each action (stop or continue) for that given state. Using this prediction, we take action with the maximum reward and proceed to the next state. The reward system ensures that the model is punished for taking the wrong actions at the right time.

3.2.3 Goal of the reinforcement learning agent

We have two possible states that our system can be in:

1) Normal functioning (S_n)
2) Under FDI Attack (S_a)

We have two possible actions that the RL agent can take:

1) Continue the normal functioning of the grid (Do not stop the simulation) (A_c)
2) Stop the simulation (A_s)

```
Global parameters: -$\theta$, w
Initialise thread-specific parameters: - $\theta'$ and $w'$
Initialize time step $t = 1$
Initialize deques trajectoryReward,
trajectoryState
 trajectoryAction
while $(st\ != TERMINAL)$ and $t - t_{start} \leq t_{max}$ do
  Pick the action $a_t \sim \pi_{\theta'}(a_t|S_t)$ and receive a new
   reward $R_t$ and a new state $S_{t+1}$
  Update $t = t + 1$ and $T = T + 1$
  append state to trajectoryState
  append action to trajectoryAction
  append reward to trajectoryReward
end
while $True$ do
  Reset gradient: $d\theta = 0$ and $dw = 0$.
  Synchronize thread-specific parameters
with global
   ones: $\theta' = \theta$ and $w' = w$.
  $tstart = t$ and sample a starting state $st$.
  Pick the action $a_t \sim \pi_{\theta'}(A_t|S_t)$ and receive a new
   reward $R_t$ and a new state $S_{t+1}$.
  Update $t = t + 1$ and $T = T + 1$
  Pop the trajectoryState
  Pop the trajectoryAction
  Pop the trajectoryReward
  Append newState to trajectoryState
  Append newAction to trajectoryAction
  Append newReward to trajectoryReward
  Initialize the variable that holds the return
   estimation
     $R = \begin{cases} 0, & S_t = \text{TERMINAL} \\ V_{w'}(S_t)\text{otherwise} & \end{cases}$
  Calculate$V_{predicted}, V_{target}$ and Advantage
  Accumulate gradients w.r.t. :
   $d\theta \leftarrow d\theta + \nabla_{\theta'} \log \pi_{\theta'}(A_i|S_i)^*(Advantage(S_i)$;
  Accumulate gradients w.r.t. w':
   $dw \leftarrow dw + \nabla_{w'}(v_{predicted} - v_{target})^2$.
  Update asynchronously $\theta$ using $d\theta$, and w using
   $dw$.
end
```

Algorithm 4. Asynchronous Advantage Actor-Critic (A3C) Online mode Sliding Window.

Our objective is to create an agent to identify the attacks as soon as they begin (not sooner, not later) in order to avert severe grid damage. We have four different possibilities as a result of our agents' actions. They are as follows:

1) The agent terminates the simulation before the attack occurs.
2) The agent terminates the simulation after the attack starts.
3) The agent does not halt the simulation after the attack starts.
4) The agent does not halt the simulation before the attack occurs.

```
Set hyper-parameters:
 $\zeta$,$totalNumActions$,$priceUpperBound$,$priceLowerBound$
 for ∀timestep $t$ do
   $action = a \in [0, totalNumActions - 1]$
   $maxChange = (priceUpperBound - priceLowerBound)/2$
   $correctingFactor = 2(maxChange^{1/\zeta})/totalNumActions$ $correctedAction = action - (totalNumActions/2)$
   $price_t = price_{t-1} + (correctingFactor * correctedAction)^{\zeta}$
 end
```

Algorithm 5. Update Price.

In the above four outcomes, only 2 and 4 are desired, whereas actions 1 and 3 are unintended.

3.2.4 Reward system

For each of the above four consequences of our agent's action, we reward it in such a way that we punish the unintended consequences and reward the intended ones. Suppose the current state is defined as S_t and the current action as A_t. We can define a possible reward policy as (Equations 10–13).

$$\text{Reward}_t = C_1,\ \text{if}\ S_t = S_a\,\text{and}\,A_t = A_s \tag{10}$$

$$\text{Reward}_t = -k_1 \times (t - t_0),\ \text{if}\ S_t = S_a\,\text{and}\,A_t = A_c \tag{11}$$

$$\text{Reward}_t = C_2,\ \text{if}\ S_t = S_n\,\text{and}\,A_t = A_c \tag{12}$$

$$\text{Reward}_t = -k_2 \times (k_3 - noise),\ \text{if}\ S_t = S_n textand A_t = A_s \tag{13}$$

Where C_1 and C_2 can be small positive values to ensure positive reward, k_1, k_2 and k_3 are constants that we can fine-tune states to improve performance. S_n is the state under normal functioning, and S_a is the state under FDI attack. At the same time, A_c denotes the action to continue the normal functioning of the grid (do not stop the simulation), and A_s is the action denoting to stop the simulation. The start (t_0) is the timestep when the attack begins. Equation 11 is the reward when the agent fails to stop the grid while the attack is happening. In this case, the reward is based on the time elapsed since the attack began (t_0). Equation 13 is the reward when the agent stops the grid when there is no attack (false positive). We want to punish this consequence, and thus the reward can have a huge negative value when this happens. We use the mean of the noise vector as the reward term at that state to incorporate it into the agent's learning process and impact its decision when a similar observation occurs at a future point in time. Hence, this reward policy should theoretically ensure that our agent learns to avoid unintended actions.

3.2.5 Attack detection learning Algorithm

The steps involved in the training process of the deep reinforcement learning algorithm are mentioned in Algorithm 6.

```
Input = model, targetModel, params
procedure TRAIN(model, targetModel)
$\epsilon \leftarrow 1, c \leftarrow 0, trainsteps \leftarrow 5000, t \leftarrow 0, Bu\,f\,ferSize \leftarrow 1000, BatchSize \leftarrow 0$
Initialize Replay Buffer as an empty list
While $t < trainsteps$ do
  $timesteps \leftarrow 500$
  $attackStarted \leftarrow False$
  $start \leftarrow$ random number between 0 and 150
  $state, noise \leftarrow grid.IEEE14bus(attack\_started)$
  $qval \leftarrow model.predict(state)$
  $action \leftarrow max(qval)$
  for i ← 0 to timesteps do
    if i = = start then
      $attack\_started \leftarrow True$
      $new\_state, new\_tot\_noise =$
       $grid.IEEE14bus(attack\_started)$
    if t ≤ observe then
      $new\_qval \leftarrow model.predict(new\_state)$
      $new\_action \leftarrow$ random number between 0 and 2
    else
      —
      $new\_qval \leftarrow model.predict(new\_state)$
      $new\_action \leftarrow max(new\_qval)$
    if attack has started and action is 0(Stop Simulation) then
      —
      $reward \leftarrow 500$
       attack has not started and action is 1(Continue Simulation)
      $reward \leftarrow 5$
       attack has not started and action is 0(Stop Simulation)
      $reward \leftarrow -100*(10 - noise)$
       attach has started and action is 0(Stop Simulation)
      $reward \leftarrow -2.5*(i - start)$
      $state = new\_state$
      $qval = new\_qval$
      $action = new\_action$
      $tot\_noise = new\_tot\_noise$
  if $\epsilon > 0$ and $t > observe$ then
  —
  $\epsilon \leftarrow \epsilon - (1/trainsteps)$
    if $t \geq 500$ or Current Size of Replay Buffer > buffer then
  Remove the oldest entry in the Replay Buffer
  Generate a mini batch from of BatchSize entries from the Replay Buffer
  $X, y \leftarrow$
$processMinibatch(minibatch)$ train the model using X and y generated
  Save the model after every 100 steps.
```

Algorithm 6. Attack Detection Learning Algorithm.

3.2.6 Correlation with NIST framework

- Identify: In the previous sections, we looked at a few possible attacks on SG systems. False Data Injection (FDI) attack is one of the most simple yet lethal attacks that can be done on smart grid networks. The FDI attack falls under two of the three possible categories of attacks we defined previously (confidentiality of data, integrity of data and commands, availability of information, and electricity). FDI attacks can tamper with the integrity of data and commands. The data that can be tampered with is the critical sensor information essential in making important decisions about electricity production and supply. FDI attacks can also cause problems with the availability of information and electricity. This happens indirectly when the tampered sensor information is used to make predictions or estimations about the amount of power needed to be generated or supplied. Having the wrong estimation can lead to unintended power surges or outages.
- Protect: This step involves ensuring that adequate safety and security measures are in place to stop the attack from happening in the first place. In this work, we discussed an additional software barrier running in real-time to protect against the FDI attack. It analyzes the sensor readings at regular intervals and ensures the detection of aberrations in the readings.
- Detect: This step involves identifying the attacks as they happen. This is essential since it enables us to respond to the threat and act accordingly. However, as technology advances, it becomes easier for attackers to bypass these detection mechanisms. Hence, having a detection mechanism exclusively for a particular attack makes it much harder for an attacker to bypass it. This work proposes a reinforcement learning-based detection mechanism for the FDI attack in particular. This method constantly tracks the sensor information and utilizes it to estimate the state of the SG system (under attack or normal functioning).
- Respond: This is the step where we take the necessary action to deal with the threat of an attack. The proposed framework takes the necessary steps as soon as it notices threats in order to lessen or lessen the severity of the damage that the attackers have caused. There are several ways to respond to an attack, ranging from blacklisting the IP address of the attack source by setting up new firewall rules to stopping the system from running for a few minutes while you respond to the threat. This ensures that the amount of damage caused is minimized. In this work, we proposed a reinforcement learning agent-based technique that takes the sensor readings at any point in time as the state of the environment and takes one of two actions: to stop the system from running further or to continue running. By stopping the system from running, we are buying time to respond to the threat posed by the attack and minimize both the physical and financial damage dealt in the process.
- Recover: The recovery phase ensures that all the services that were hindered during the attack are restored to their normal functionality. It also involves setting up necessary security measures to prevent attacks from happening further. Many strategies were proposed to deal with the recovery phase. In this work, we do not deal with the recovery phase of the NIST framework. We suggest possible approaches in the future work section of this work.

4 Dataset

4.1 For load profiling

For training the reinforcement learning-based agents, we tried to mimic the behavior of the power flow analysis of the smart grid. We have used the publicly available dataset called Independent Electricity System Operator (IESO). The IESO data is the collection of various reports released by Ontario's power grid operators. The report contains supply, demand, tariffs, and other relevant parameters. The dataset has values from the time interval 2010-01-01 to 2019-12-20, with data points taken at an interval of 5 minutes. We mainly required hourly electricity consumption by the consumers and the historical tariff price so we could use it to train our LSTM model to better simulate the future parameters of the smart grid, including demand, supply, and other parameters, based on the changes to the tariffs made by the agent.

While performing the correlation analysis, we removed the column with a negligible correlation with the price and demand value, which resulted in the reduction of the column from 47 to 13. This would make the training more accurate for the LSTM model and efficiently mimic the SG behavior in the agent's actions. Finally, there might be some cases where relevant supply data is unavailable, so we study the latest supply and demand values from IESO and compute the supply column values, keeping in mind that their correlation with other column variables is equal to that of the latest data. This step was essential for the reward computation due to our assumption that in the SG environment, most consumers also behave like prosumers, thus making their demand fluctuate. In this model behavior, the worst situation may arise when the power supply is greater than demand, thus making the agent (producer) pay back to the consumers. This will enhance the importance of the supply value column to make the model appropriate.

In addition, we normalized the dataset to speed up the learning process for the LSTM to simulate the environment and, indirectly, the learning for the policy and critic networks, leading to faster convergence. The environment and the agents all use normalized data, except for the reward function and the logging facilities, the former to give more fine-tuned reward signals to the agent and the latter for debugging purposes.

Dataset normalization is done to make training easier for NN. We use the following formula to normalize the dataset values (Equation 14).

$$value_{row,col} = \frac{value_{row,col} - min\left(valueS_{col}\right)}{max\left(valueS_{col}\right) - min\left(valueS_{col}\right)} \tag{14}$$

4.2 For cyber threat detection

The investigation and data generation was completed in a phased manner so that the readiness of data utilized in FDI

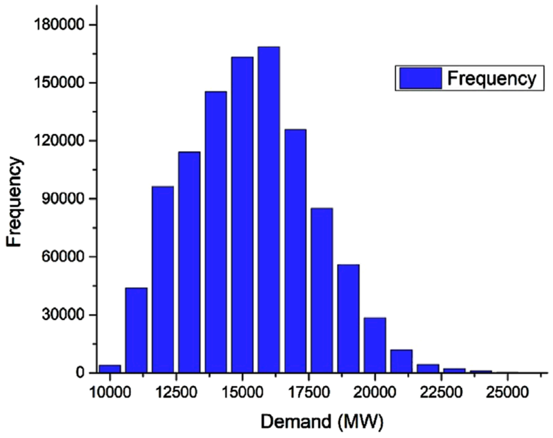

FIGURE 6
Distribution of the electricity Demand.

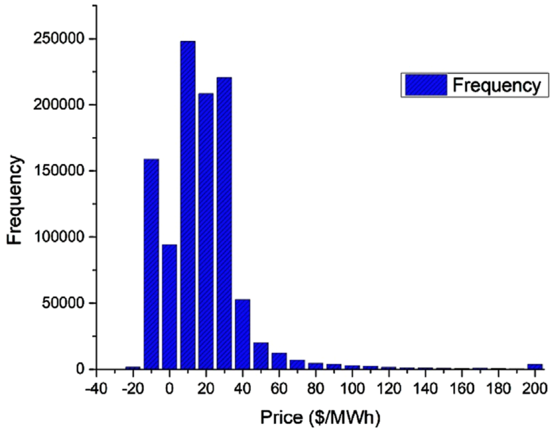

FIGURE 7
Distribution of the electricity Price.

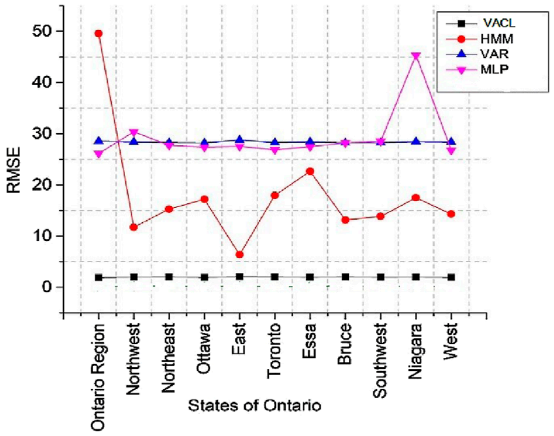

FIGURE 8
Forecast of the electricity Price.

attack detection on an IEEE 9-bus framework and a 14-bus system was carried out in MATLAB Simulink (Documentation, 2020) and MATPOWER (Zimmerman et al., 2011). For 9 Bus, we deployed six three-phase V-I measurement parts to recreate PMUs introduced in power frameworks. Likewise, for 14 bus, we deployed. 11 three-phase V-I measurement parts. For each PMU in the 9-bus framework, we recorded 18 distinctive electrical amounts, like the magnitude and those related to current and voltage. Similarly, for 14 bus, 28 such electrical estimations are recorded.

5 Power system use case

5.1 Achieving resilience through demand response

5.1.1 Load forecasting

The load forecasting is done using model proposed in Sinha et al. (2021). The experiment has been carried out in IESO data (I. E. S. O. (IESO)) combined with Canada weather data. The data distribution is presented as in Figures 6, 7.

The model outcome on the IESO dataset is described in Figure 8. From the result, it is clear that the model has achieved good accuracy and outperforms well in comparison to other existing models like MLP, HMM, and VAR. For better clarity, we have tested the model on all eleven regions of Ontario, and the RMSE score is calculated and presented for all the regions simultaneously.

5.1.2 Load profiling at microgrid level

We have used Algorithms 1, 2, 3, 4, and 5 for optimal profit and demand response optimization in the microgrid environment. The online real-time training of the agent as in Figures 9–11, and Figure 12 elaborates the outcome of the online training of the agent with a method of model updates chosen as episodic (i.e., sliding window approach is not being used here). Instead of it, we adopted a head-start approach-based pre-trained network described earlier. Figure 9 depicts that the value of the mean profit gathered by the agent is relatively higher than the dataset profit. This would imply that the agent gradually becomes competent in maximizing profits while keeping the price column's value within acceptable bounds or limits.

Figure 10 shows that the critic loss constantly decreases, implying that the value network can predict the correct value of the states. When compared to offline mode, we can see that after 2000 episodes, the critic loss in online mode is of the order of 0.01, compared to 10^5 in offline mode.

Figure 11 shows that in online mode, policy loss becomes almost optimal after 2000 episodes, whereas it takes 5,000 episodes in offline mode for it to become optimal. This implies a better convergence rate to optimal policy in online mode due to the use of pre-trained networks (headstart modifications).

It is clear from Figure 12 that the average advantage in online mode is much better than in offline mode. The algorithm also quickly converges (on the input data), implying that in online mode,

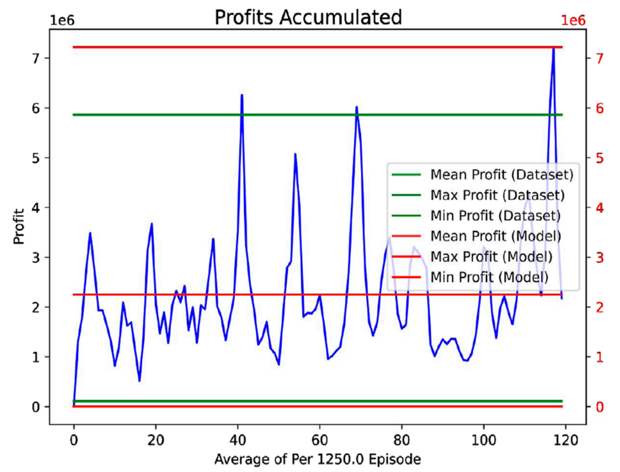

FIGURE 9
Profit accumulated plot with the mean, max and min profits values of models and original datasets.

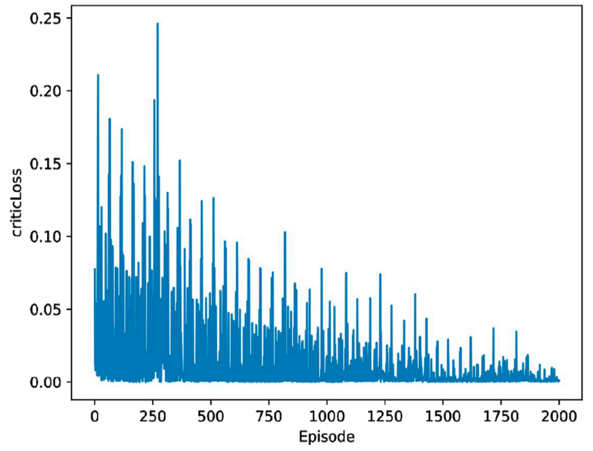

FIGURE 10
Critic loss curve based on our RL model.

the agent is taking comparatively better actions than its offline counterpart and is quick to identify the most optimal actions to be taken in a state.

So, it may be inferred that Deep reinforcement learning (DRL) is highly effective for load profiling and demand response optimization because it adapts in real time to changing demand patterns, predicts future needs, and makes sequential decisions to balance load. Unlike traditional methods, DRL can personalize load management for different users, automate demand-side responses, and handle the complexities of renewable energy integration. By continuously learning from real-time data, DRL enables proactive peak shaving, cost reduction, and enhanced grid stability, making the smart grid more resilient, efficient, and responsive.

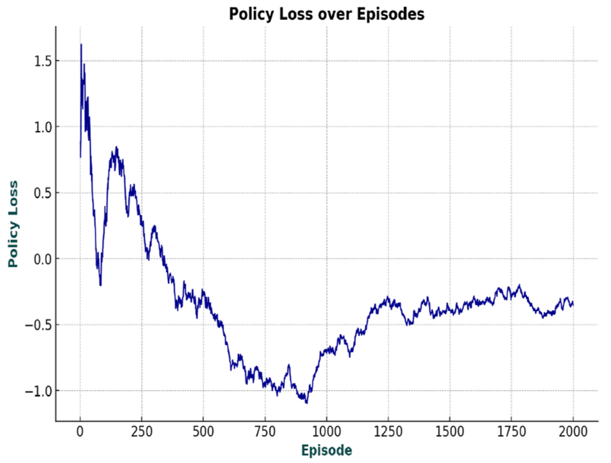

FIGURE 11
Policy loss curve based on our RL model.

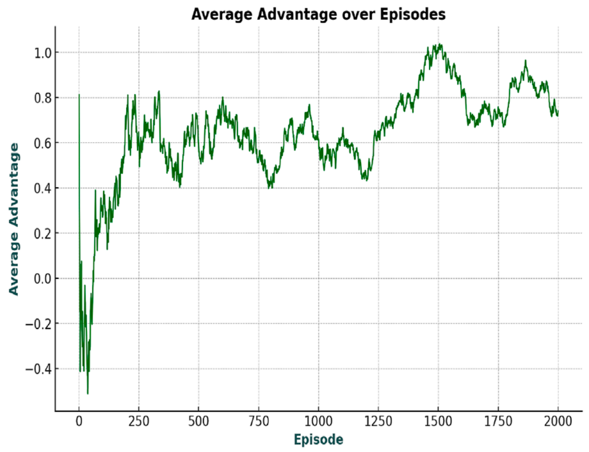

FIGURE 12
Average advantage.

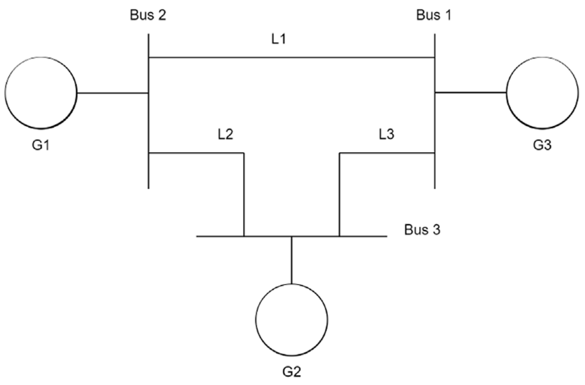

FIGURE 13
3 bus grid.

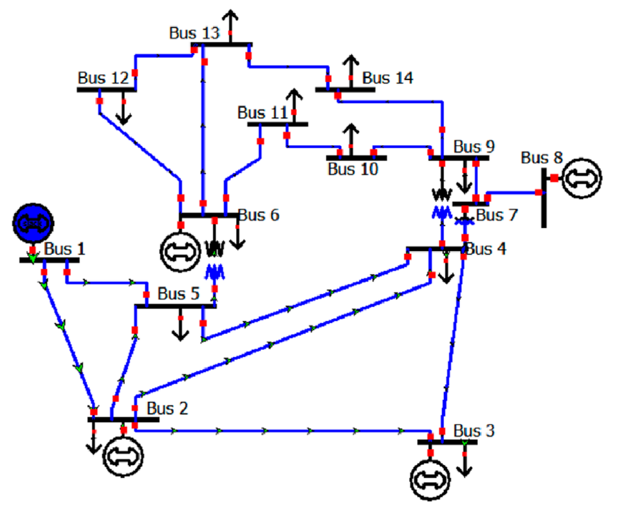

FIGURE 14
IEEE 14 bus grid network.

5.2 Achieving resiliency through multiagent detection mechanism

This section deals with the experimental outcome as achieved using the proposed Algorithm 6 for the in-progress FDI attack. The experiment was carried out on 3-bus (Abur and Exposito, 2004), IEEE-9 bus, IEEE-14 bus, and IEEE-30 bus systems (Figures 13, 14; Table 4). The assessment will be conducted on three exemplary bus grids, in conjunction with several standard IEEE grids, including the IEEE 9 bus, IEEE 14 bus, and IEEE 32 bus systems. The existing system state vector, comprising voltage magnitudes and phase angles, is ascertained through the utilization of State Estimation functions inherent to the PANDAPOWER Python library (Thurner et al., 2018). Simulation endeavors involve the initialization of network configurations to their default values for the simple 3-bus grid, IEEE 9 bus, IEEE 14 bus, and IEEE 33 bus systems, each of which is subjected to a prescribed number of steps during each episode. The inception of a Fault Detection and Isolation (FDI) attack is introduced at a randomly selected point within an episode and persists for an indeterminate duration. The principal objective entails training our model to promptly terminate the episode upon the commencement of the FDI attack, with temporal precision. To check the convergence of the proposed reinforcement-based learning algorithm, we plot the value of the loss function as the training of the model progresses. The plot is shown in Figure 15.

The evaluation metrics for the proposed RL agent are as follows:

1) Perfect Calls Percentage
2) False Alarm Rate
3) Good Calls Percentage
4) Late Calls rate
5) Detection Failure Percentage

1) Perfect Calls Percentage: The number of attacks that our RL agent can detect as soon as they start. Let the number of episodes where the attack starts time step is the same as the attack detection time step be N_p and the total number of episodes be T. This can be computed by the equation:

$$Perfect\ Calls\ Percentage = \frac{N_p}{T} * 100$$

2) Good Calls Percentage: Let us define good calls as detecting attacks before a certain number of time steps after they start. We can call this threshold the Good Calls Percentage. Let the number of episodes where the attack detection is within time steps after it starts to be N_g and the total number of episodes be T. This can be computed using the equation:

$$Good\ Calls\ Percentage = \frac{N_g}{T} * 100$$

3) Delayed Calls Rate: Let us define Delayed calls as detecting the attack anytime after the attack begins. Let start be the time step when the attack has begun, t be the time step when the agent detects and the total number of episodes be T. For all the episodes that start, this can be computed using the equation:

$$Delayed\ Calls\ Rate = \frac{\sum|t-start|}{T}$$

4) False Alarm Rate: False Alarms are the calls that occur before the attack begins. We would want to avoid these as much as possible to avoid disruptions in power supplies. Let start be the time step when the attack has begun, t be the time step when the agent detects it, and the total number of episodes be T. For all the episodes that start, this can be computed using the equation:

$$False\ Alarms\ Rate = \frac{\sum|t-start|}{T}$$

5) Detection Failure Percentage: When our agent is unable to detect that the attack occurred by the end of the episode, we call this a detection failure. Let the number of episodes this happens to be called N_df and the total episodes be T. This can be computed using the equation:

$$Detection\ Failure\ Percentage = \frac{N_{df}}{T} * 100$$

We have used evaluation metrics to estimate the performance of the proposed RL algorithm for the model trained for simple −3 bus, IEEE-9, IEEE-14, and IEEE-33 bus systems. We have experimented using 100 episodes for all IEEE system grids. The threshold for good calls was set to 10 time steps after the attack began. Table 4 shows the results of the experiment done on an IEEE 9, 14, 30, and 3-bus grid system.

6 Conclusion

The Smart Grid framework is an emerging innovation that carries many advantages to administrators and users, even though it has a few downsides regarding safety and security, which may impact its deployment in real-time applications. Like other frameworks for critical infrastructure, the advancement of SG modern devices toward an exceptionally associated and distributed model paves a few issues for the reliability and safety of the integrated framework. The proposed framework investigates the reliance of DR on the smart grid and shows how the ideas of DR and cyber security with resilience are intrinsically related. The framework first gives

TABLE 4 Evaluation results of in progress FDI attack using proposed method.

Bus type	Perfect calls percentage	Good calls percentage	Delayed calls rate	False alarms rate	Detection failure percentage
3 Bus Grid	81	96	0.72	1.84	0
IEEE 9 Bus	94	96	0.01	0.50	0
IEEE 14 Bus	88	99	0.50	0.00	0
IEEE 33 Bus	94	92	0.00	4.60	0

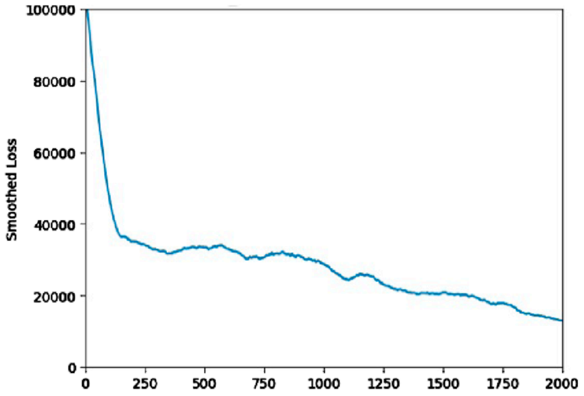

FIGURE 15
Loss curve.

the deep learning model for accurately estimating electrical load and price. Consequently, it proposes an optimized demand response strategy in a multi-micro-grid environment using a modified RL-based A3C algorithm (in offline and online modes). The results show that for DR optimization, online mode converges more quickly than offline mode, implying that in online mode, the agent is taking comparatively better actions than its offline counterpart and is quick to identify the most optimal actions to be taken in a state. Further, the framework explored the intricacies of in-progress cyber attacks, especially FDI attacks. It proposed a reinforcement learning-based algorithm for the same, and the experiment is carried out on IEEE-3, IEEE-9, IEEE-14, and IEEE-33 bus systems. It is shown with the help of a plot that the loss function minimizes as the model's training progresses. The evaluation metrics for the proposed RL agent for the in-progress FDI attack are Perfect Calls Percentage, False Alarm rate, Good Calls Percentage, Late Calls rate, and Detection Failure Percentage. Finally, the framework shows the interdependence of DR and cyber security and proposes a solution for reliable smart grid functioning.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: https://www.ieso.ca/power-data.

Author contributions

AS: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Software, Validation, Writing–original draft, Writing–review and editing. RV: Project administration, Supervision, Writing–review and editing. FA: Project administration, Supervision, Writing–review and editing. WH: Investigation, Methodology, Supervision, Writing–original draft, Writing–review and editing. OV: Conceptualization, Formal Analysis, Investigation, Methodology, Project administration, Resources, Supervision, Writing–review and editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. The work is partially funded by the Department of Science and Technology (DST), and C3i-Hub (Indian Institute of Technology, Kanpur), India for the Risk Averse Resilience Framework for Critical Infrastructure Security (RARCIS) project.

Acknowledgments

The English language of this manuscript was reviewed and improved using Grammarly, an AI-based writing assistant tool that helps enhance grammar, clarity, and style. (https://app.grammarly.com/).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Abur, A., and Expósito, A. G. (2004). *Power system state estimation: theory and implementation*. 1st ed. CRC press. doi:10.1201/9780203913673

Adepu, S., Kandasamy, N. K., Zhou, J., and Mathur, A. (2020). Attacks on smart grid: power supply interruption and malicious power generation. *Int. J. Inf. Secur.* 19, 189–211. doi:10.1007/s10207-019-00452-z

Aghaei, J., Alizadeh, M.-I., Siano, P., and Heidari, A. (2016). Contribution of emergency demand response programs in power system reliability. *Energy* 103, 688–696. doi:10.1016/j.energy.2016.03.031

Ahmadi, S.-A., Vahidinasab, V., Ghazizadeh, M. S., Mehran, K., Giaouris, D., and Taylor, P. (2019). Co-optimising distribution network adequacy and security by simultaneous utilisation of network reconfiguration and distributed energy resources. *IET Generation, Transm. and Distribution* 13 (20), 4747–4755. doi:10.1049/iet-gtd.2019.0824

Alipoor, J., Miura, Y., and Ise, T. (2014). Power system stabilization using virtual synchronous generator with alternating moment of inertia. *IEEE J. Emerg. Sel. Top. power Electron.* 3 (2), 451–458. doi:10.1109/jestpe.2014.2362530

Allesina, S., and Tang, S. (2012). Stability criteria for complex ecosystems. *Nature* 483 (7388), 205–208. doi:10.1038/nature10832

Amin, S., Schwartz, G. A., Cardenas, A. A., and Sastry, S. S. (2015). Game-theoretic models of electricity theft detection in smart utility networks: providing new capabilities with advanced metering infrastructure. *IEEE Control Syst. Mag.* 35 (1), 66–81. doi:10.1109/MCS.2014.2364711

Arani, M. F. M., and El-Saadany, E. F. (2012). Implementing virtual inertia in dfig-based wind power generation. *IEEE Trans. Power Syst.* 28 (2), 1373–1384. doi:10.1109/tpwrs.2012.2207972

Arasteh, H., Vahidinasab, V., Sepasian, M. S., and Aghaei, J. (2018). Stochastic system of systems architecture for adaptive expansion of smart distribution grids. *IEEE Trans. Industrial Inf.* 15 (1), 377–389. doi:10.1109/tii.2018.2808268

Awad, A. S., El-Fouly, T. H., and Salama, M. M. (2015). Optimal ess allocation for benefit maximization in distribution networks. *IEEE Trans. Smart Grid* 8 (4), 1668–1678. doi:10.1109/tsg.2015.2499264

Babar, M., Tariq, M. U., and Jan, M. A. (2020). Secure and resilient demand side management engine using machine learning for iot-enabled smart grid. *Sustain. Cities Soc.* 62, 102370. doi:10.1016/j.scs.2020.102370

Balali, A., Yunusa-Kaltungo, A., and Edwards, R. (2023). A systematic review of passive energy consumption optimisation strategy selection for buildings through multiple criteria decision-making techniques. *Renew. Sustain. Energy Rev.* 171, 113013. doi:10.1016/j.rser.2022.113013

Bhattarai, B. P., Bak-Jensen, B., Chaudhary, S., and Pillai, J. R. (2015). "An adaptive overcurrent protection in smart distribution grid," in *2015 IEEE eindhoven PowerTech* (IEEE), 1–6.

Bohra, S. S., and Anvari-Moghaddam, A. (2022). A comprehensive review on applications of multicriteria decision-making methods in power and energy systems. *Int. J. Energy Res.* 46 (4), 4088–4118. doi:10.1002/er.7517

Cao, G., Gu, W., Lou, G., Sheng, W., and Liu, K. (2022). Distributed synchronous detection for false data injection attack in cyber-physical microgrids. *Int. J. Electr. Power and Energy Syst.* 137, 107788. doi:10.1016/j.ijepes.2021.107788

Chen, C., Wang, J., Qiu, F., and Zhao, D. (2015). Resilient distribution system by microgrids formation after natural disasters. *IEEE Trans. smart grid* 7 (2), 958–966. doi:10.1109/tsg.2015.2429653

Cheng, L., and Yu, T. (2019). A new generation of ai: a review and perspective on machine learning technologies applied to smart energy and electric power systems. *Int. J. Energy Res.* 43 (6), 1928–1973. doi:10.1002/er.4333

Clark, A., and Zonouz, S. (2019). Cyber-physical resilience: definition and assessment metric. *IEEE Trans. Smart Grid* 10 (2), 1671–1684. doi:10.1109/tsg.2017.2776279

Cohenpb, J. E., and Charles, M. N. (1985). When will a large complex system be stable? *J. Theor. Biol.* 113, 153–156.

Cybersecurity, C. I. (2018). Framework for improving critical infrastructure cybersecurity. CSWP, vol. 4162018. Available at: https://nvlpubs. nist. gov/nistpubs/CSWP/NIST.

Ding, T., Lin, Y., Bie, Z., and Chen, C. (2017). A resilient microgrid formation strategy for load restoration considering master-slave distributed generators and topology reconfiguration. *Appl. energy* 199, 205–216. doi:10.1016/j.apenergy.2017.05.012

Documentation, S. (2020). Simulation and model-based design. Available at: https://www.mathworks.com/products/simulink.html.

Esmalifalak, M., Liu, L., Nguyen, N., Zheng, R., and Han, Z. (2014). Detecting stealthy false data injection using machine learning in smart grid. *IEEE Syst. J.* 11 (3), 1644–1652. doi:10.1109/jsyst.2014.2341597

e Sousa, Á., Messai, N., and Manamanni, N. (2022). Load-altering attack detection on smart grid using functional observers. *Int. J. Crit. Infrastructure Prot.* 37, 100518. doi:10.1016/j.ijcip.2022.100518

EuGovernment European Union agency for cybersecurity (enisa), smart grids. Available at: https://www.enisa.europa.eu/topics/critical-information-infrastructures-and-services/smart-grids?tab=details.

Fleschutz, M., Bohlayer, M., Braun, M., Henze, G., and Murphy, M. D. (2021). The effect of price-based demand response on carbon emissions in european electricity markets: the importance of adequate carbon prices. *Appl. Energy* 295, 117040. doi:10.1016/j.apenergy.2021.117040 Available at: https://www.sciencedirect.com/science/article/pii/S0306261921004992.

Gao, Y., Wang, W., Shi, J., and Yu, N. (2020). Batch-constrained reinforcement learning for dynamic distribution network reconfiguration. *IEEE Trans. Smart Grid* 11 (6), 5357–5369. doi:10.1109/tsg.2020.3005270

Gholami, A., Shekari, T., Aminifar, F., and Shahidehpour, M. (2016). Microgrid scheduling with uncertainty: the quest for resilience. *IEEE Trans. Smart Grid* 7 (6), 2849–2858. doi:10.1109/tsg.2016.2598802

Gribble, S. D.(2001). "Robustness in complex systems," in *Proceedings eighth workshop on hot topics in operating systems* (IEEE), 21–26.

Habib, H. F., Mohamed, A., El Hariri, M., and Mohammed, O. A. (2017). Utilizing supercapacitors for resiliency enhancements and adaptive microgrid protection against communication failures. *Electr. Power Syst. Res.* 145, 223–233. doi:10.1016/j.epsr.2016.12.027

Han, M., Chen, Y., Gao, S., Zhou, J., and Ren, C. (2020). "Coupled optimization of topology reconfiguration and voltage source converter control for enlarging load margin of ac/dc distribution network," in 2020 IEEE 4th conference on energy Internet and energy system integration (EI2) (IEEE), 633–637.

Holderbaum, W., Alasali, F., and Sinha, A. (2023). "Model predictive control," in *Energy forecasting and control methods for energy storage systems in distribution networks: predictive modelling and control techniques* (Springer), 129–148.

Home-Ortiz, J. M., and Mantovani, J. R. S. (2020a). "Enhancement of the resilience through microgrids formation and dg allocation with master-slave dg operation," in 2020 international conference on smart energy systems and technologies (SEST) (IEEE), 1–6.

Home-Ortiz, J. M., and Mantovani, J. R. S. (2020b). "Resilience enhancing through microgrids formation and distributed generation allocation," in *2020 IEEE PES innovative smart grid technologies europe (ISGT-Europe)* (IEEE), 995–999.

Ibrahim, A., El-Khattam, W., ElMesallamy, M., and Talaat, H. (2016). Adaptive protection coordination scheme for distribution network with distributed generation using abc. *J. Electr. Syst. Inf. Technol.* 3 (2), 320–332. doi:10.1016/j.jesit.2015.11.012

I. E. S. O. (IESO) Power data. Available at: https://www.ieso.ca/power-data.

InEnergy India smart grid forum. Available at: https://www.indiasmartgrid.org/.

Korjenic, A., and Bednar, T. (2011). "Impact of lifestyle on the energy demand of a single family house," in *Build. Simul., building simulation*, 4, Springer, 89–95. doi:10.1007/s12273-010-0013-4

Li, Z., Shahidehpour, M., Aminifar, F., Alabdulwahab, A., and Al-Turki, Y. (2017). Networked microgrids for enhancing the power system resilience. *Proc. IEEE* 105 (7), 1289–1310. doi:10.1109/jproc.2017.2685558

Liao, B., Cheng, J., and Ren, G. (2019). "Microgrid adaptive current instantaneous trip protection," in 2019 IEEE innovative smart grid technologies-asia (ISGT asia) (IEEE), 2074–2078.

Lin, Y., and Bie, Z. (2018). Tri-level optimal hardening plan for a resilient distribution system considering reconfiguration and dg islanding. *Appl. Energy* 210, 1266–1279. doi:10.1016/j.apenergy.2017.06.059

Liu, N., Yu, X., Wang, C., Li, C., Ma, L., and Lei, J. (2017). Energy-sharing model with price-based demand response for microgrids of peer-to-peer prosumers. *IEEE Trans. Power Syst.* 32 (5), 3569–3583. doi:10.1109/tpwrs.2017.2649558

Long, C., Wu, J., Zhang, C., Thomas, L., Cheng, M., and Jenkins, N. (2017). *Peer-to-peer energy trading in a community microgrid*. IEEE power and energy society general meeting. IEEE, 1–5.

Lopez, J., Rubio, J. E., and Alcaraz, C. (2018). A resilient architecture for the smart grid. *IEEE Trans. Industrial Inf.* 14 (8), 3745–3753. doi:10.1109/tii.2018.2826226

Ma, J., Liu, J., Deng, Z., Wu, S., and Thorp, J. S. (2018). An adaptive directional current protection scheme for distribution network with dg integration based on fault steady-state component. *Int. J. Electr. Power and Energy Syst.* 102, 223–234. doi:10.1016/j.ijepes.2018.04.024

Mahat, P., Chen, Z., Bak-Jensen, B., and Bak, C. L. (2011). A simple adaptive overcurrent protection of distribution systems with distributed generation. *IEEE Trans. Smart Grid* 2 (3), 428–437. doi:10.1109/tsg.2011.2149550

Mashal, I., Khashan, O. A., Hijjawi, M., and Alshinwan, M. (2023). The determinants of reliable smart grid from experts' perspective. *Energy Inf.* 6 (1), 10. doi:10.1186/s42162-023-00266-3

Mnih, V., Badia, A. P., Mirza, M., Graves, A., Lillicrap, T., Harley, T., et al. (2016). "Asynchronous methods for deep reinforcement learning," in International conference on machine learning (PMLR), 1928–1937.

Mohassel, R. R., Fung, A., Mohammadi, F., and Raahemifar, K. (2014). A survey on advanced metering infrastructure. *Int. J. Electr. Power and Energy Syst.* 63, 473–484. doi:10.1016/j.ijepes.2014.06.025

Momesso, A. E., Bernardes, W. M. S., and Asada, E. N. (2020). Adaptive directional overcurrent protection considering stability constraint. *Electr. Power Syst. Res.* 181, 106190. doi:10.1016/j.epsr.2019.106190

Morstyn, T., Farrell, N., Darby, S. J., and McCulloch, M. D. (2018). Using peer-to-peer energy-trading platforms to incentivize prosumers to form federated power plants. *Nat. Energy* 3 (2), 94–101. doi:10.1038/s41560-017-0075-y

Mousavizadeh, S., Haghifam, M.-R., and Shariatkhah, M.-H. (2018). A linear two-stage method for resiliency analysis in distribution systems considering renewable energy and demand response resources. *Appl. energy* 211, 443–460. doi:10.1016/j.apenergy.2017.11.067

Muda, H., and Jena, P. (2017). Sequence currents based adaptive protection approach for dns with distributed energy resources. *IET Generation, Transm. and Distribution* 11 (1), 154–165. doi:10.1049/iet-gtd.2016.0727

Nair, V., and Hinton, G. E. (2010). Rectified linear units improve restricted Boltzmann machines. *Icml.*

Nascimento, J. P., Brito, N. S., and Souza, B. A. (2020). An adaptive overcurrent protection system applied to distribution systems. *Comput. and Electr. Eng.* 81, 106545. doi:10.1016/j.compeleceng.2019.106545

Nick, M., Cherkaoui, R., and Paolone, M. (2017). Optimal planning of distributed energy storage systems in active distribution networks embedding grid reconfiguration. *IEEE Trans. Power Syst.* 33 (2), 1577–1590. doi:10.1109/tpwrs.2017.2734942

Nikkhah, S., Allahham, A., Royapoor, M., Bialek, J. W., and Giaouris, D. (2021). "A community-based building-to-building strategy for multi-objective energy management of residential microgrids," in 2021 12th international renewable engineering conference (IREC) (IEEE), 1–6.

Nikkhah, S., and Rabiee, A. (2018). Voltage stability constrained multi-objective optimisation model for long-term expansion planning of large-scale wind farms. *IET Generation, Transm. and Distribution* 12 (3), 548–555. doi:10.1049/iet-gtd.2017.0763

Nikkhah, S., Rabiee, A., Mohseni-Bonab, S. M., and Kamwa, I. (2020). Risk averse energy management strategy in the presence of distributed energy resources considering distribution network reconfiguration: an information gap decision theory approach. *IET Renew. Power Gener.* 14 (2), 305–312. doi:10.1049/iet-rpg.2019.0472

Panteli, M., and Mancarella, P. (2015). The grid: stronger, bigger, smarter? presenting a conceptual framework of power system resilience. *IEEE Power Energy Mag.* 13 (3), 58–66. doi:10.1109/mpe.2015.2397334

Papaspiliotopoulos, V. A., Korres, G. N., Kleftakis, V. A., and Hatziargyriou, N. D. (2015). Hardware-in-the-loop design and optimal setting of adaptive protection schemes for distribution systems with distributed generation. *IEEE Trans. Power Deliv.* 32 (1), 393–400. doi:10.1109/tpwrd.2015.2509784

Paterakis, N. G., Mazza, A., Santos, S. F., Erdinç, O., Chicco, G., Bakirtzis, A. G., et al. (2015). Multi-objective reconfiguration of radial distribution systems using reliability indices. *IEEE Trans. Power Syst.* 31 (2), 1048–1062. doi:10.1109/tpwrs.2015.2425801

Pilz, M., Naeini, F. B., Grammont, K., Smagghe, C., Davis, M., Nebel, J.-C., et al. (2020). Security attacks on smart grid scheduling and their defences: a game-theoretic approach. *Int. J. Inf. Secur.* 19, 427–443. doi:10.1007/s10207-019-00460-z

Rabiee, A., Nikkhah, S., and Soroudi, A. (2018). Information gap decision theory to deal with long-term wind energy planning considering voltage stability. *Energy* 147, 451–463. doi:10.1016/j.energy.2018.01.061

Romanenko, A., Tanjimuddin, M., Raussi, P., Aro, M., Tikka, V., and Honkapuro, S. (2020). "Taxonomy of security threats in energy systems," in 2020 17th international conference on the European energy market (EEM) (IEEE), 1–7.

Ross, R., Pillitteri, V., Graubart, R., Bodeau, D., and McQuaid, R. (2019). Developing cyber resilient systems: a systems security engineering approach. *Natl. Inst. Stand. Technol. Tech. Rep.*

Sampath Kumar, D., Srinivasan, D., Sharma, A., and Reindl, T. (2018). Adaptive directional overcurrent relaying scheme for meshed distribution networks. *IET Generation, Transm. and Distribution* 12 (13), 3212–3220. doi:10.1049/iet-gtd.2017.1279

Sharifi, R., Fathi, S., and Vahidinasab, V. (2017). A review on demand-side tools in electricity market. *Renew. Sustain. Energy Rev.* 72, 565–572. doi:10.1016/j.rser.2017.01.020

Shih, M. Y., Conde, A., Leonowicz, Z., and Martirano, L. (2017). An adaptive overcurrent coordination scheme to improve relay sensitivity and overcome drawbacks due to distributed generation in smart grids. *IEEE Trans. industry Appl.* 53 (6), 5217–5228. doi:10.1109/tia.2017.2717880

Singh, J., Sinha, A., Goli, P., Subramanian, V., Shukla, S. K., and Vyas, O. P. (2021). Insider attack mitigation in a smart metering infrastructure using reputation score and blockchain technology. *Int. J. Inf. Secur.* 21, 527–546. doi:10.1007/s10207-021-00561-8

Singh, V. K., and Govindarasu, M. (2020). "A novel architecture for attack-resilient wide-area protection and control system in smart grid," in *2020 resilience week (RWS)* (IEEE), 41–47.

Sinha, A., Chakrabarti, S., and Vyas, O. (2020). "Distributed grid restoration based on graph theory," in *2020 IEEE international symposium on sustainable energy, signal processing and cyber security (iSSSC)*, 1–6.

Sinha, A., Dwivedi, S., Shukla, S. K., and Vyas, O. (2022). "Commissioning random matrix theory and synthetic minority oversampling technique for power system faults detection and classification," in International conference on neural information processing (Springer), 518–529.

Sinha, A., Tayal, R., Vyas, R., and Vyas, O. (2021). "Operational flexibility with statistical and deep learning model for electricity load forecasting," in *Accepted in lecture notes in electrical engineering (LNEE)*. Springer.

Soni, N., Doolla, S., and Chandorkar, M. C. (2013). Improvement of transient response in microgrids using virtual inertia. *IEEE Trans. power Deliv.* 28 (3), 1830–1838. doi:10.1109/tpwrd.2013.2264738

Srivastava, A., and Parida, S. (2022). Data driven approach for fault detection and Gaussian process regression based location prognosis in smart ac microgrid. *Electr. Power Syst. Res.* 208, 107889. doi:10.1016/j.epsr.2022.107889

Syrmakesis, A. D., Alcaraz, C., and Hatziargyriou, N. D. (2022). Classifying resilience approaches for protecting smart grids against cyber threats. *Int. J. Inf. Secur.* 21 (5), 1189–1210. doi:10.1007/s10207-022-00594-7

Szulecki, K., Ancygier, A., and Szwed, D. (2015) "Energy democratization? societal aspects of de-carbonization in the German and polish energy sectors," in *Societal aspects of de-carbonization in the German and polish energy sectors*.

Tang, D., Fang, Y., and Zio, E. (2019b). "A zero-sum markov defender-attacker game for modeling false pricing in smart grids and its solution by multi-agent reinforcement learning," in 29th European safety and reliability conference (ESREL2019).

Tang, D., Fang, Y., Zio, E., and Ramirez-Marquez, J. E. (2018). "Analysis of the vulnerability of smart grids to social network-based attacks," in 2018 3rd international conference on system reliability and safety (ICSRS) (IEEE), 130–134.

Tang, D., Fang, Y. P., Zio, E., and Ramirez-Marquez, J. E. (2019). Resilience of smart power grids to false pricing attacks in the social network. *IEEE Access* 7, 80491–80505. doi:10.1109/access.2019.2923578

Tebekaemi, E., and Wijesekera, D. (2019). Secure overlay communication and control model for decentralized autonomous control of smart micro-grids. *Sustain. Energy, Grids Netw.* 18, 100222. doi:10.1016/j.segan.2019.100222

Thurner, L., Scheidler, A., Schäfer, F., Menke, J.-H., Dollichon, J., Meier, F., et al. (2018). pandapower—an open-source python tool for convenient modeling, analysis, and optimization of electric power systems. *IEEE Trans. Power Syst.* 33 (6), 6510–6521. doi:10.1109/tpwrs.2018.2829021

Tielens, P., and Van Hertem, D. (2016). The relevance of inertia in power systems. *Renew. Sustain. Energy Rev.* 55, 999–1009. doi:10.1016/j.rser.2015.11.016

Tummasit, N., Premrudeepreechacharn, S., and Tantichayakorn, N. (2015). "Adaptive overcurrent protection considering critical clearing time for a microgrid system," in *2015 IEEE innovative smart grid technologies-asia (ISGT ASIA)* (IEEE), 1–6.

Upadhyay, D., and Sampalli, S. (2020). Scada (supervisory control and data acquisition) systems: vulnerability assessment and security recommendations. *Comput. and Secur.* 89, 101666. doi:10.1016/j.cose.2019.101666

UsEnergy. U.s. department of energy, cybersecurity. Available at: https://www.energy.gov/national-security-safety/cybersecurity.

Vahidinasab, V. (2014). Optimal distributed energy resources planning in a competitive electricity market: multiobjective optimization and probabilistic design. *Renew. energy* 66, 354–363. doi:10.1016/j.renene.2013.12.042

Wang, Z., Chen, B., Wang, J., and kim, J. (2015). Decentralized energy management system for networked microgrids in grid-connected and islanded modes. *IEEE Trans. Smart Grid* 7 (2), 1097–1105. doi:10.1109/tsg.2015.2427371

Wang, Z., and Wang, J. (2015). Self-healing resilient distribution systems based on sectionalization into microgrids. *IEEE Trans. Power Syst.* 30 (6), 3139–3149. doi:10.1109/tpwrs.2015.2389753

Zeadally, S., Adi, E., Baig, Z., and Khan, I. A. (2020). Harnessing artificial intelligence capabilities to improve cybersecurity. *Ieee Access* 8, 23 817–823 837. doi:10.1109/access.2020.2968045

Zhang, Q., Dehghanpour, K., Wang, Z., and Huang, Q. (2019). A learning-based power management method for networked microgrids under incomplete information. *IEEE Trans. Smart Grid* 11 (2), 1193–1204. doi:10.1109/tsg.2019.2933502

Zhang, X., and Chi, K. T. (2015). "Assessment of robustness of power systems from the perspective of complex networks," in 2015 IEEE international symposium on circuits and systems (ISCAS) (IEEE), 2684–2687.

Zimmerman, R. D., Murillo-Sánchez, C. E., and Thomas, R. J. (2011). Matpower: steady-state operations, planning, and analysis tools for power systems research and education. *IEEE Trans. Power Syst.* 26 (1), 12–19. doi:10.1109/tpwrs.2010.2051168