mathematics

MDPI

Article

An Adoptive Miner-Misuse Based Online Anomaly Detection Approach in the Power System: An Optimum Reinforcement Learning Method

Abdulaziz Almalaq [1,*], Saleh Albadran [1] and Mohamed A. Mohamed [2,*]

[1] Department of Electrical Engineering, Engineering College, University of Ha'il, Ha'il 55476, Saudi Arabia
[2] Electrical Engineering Department, Faculty of Engineering, Minia University, Minia 61519, Egypt
* Correspondence: a.almalaq@uoh.edu.sa (A.A.); dr.mohamed.abdelaziz@mu.edu.eg (M.A.M.)

Abstract: Over the past few years, the Bitcoin-based financial trading system (BFTS) has created new challenges for the power system due to the high-risk consumption of mining devices. Briefly, such a problem would be a compelling incentive for cyber-attackers who intend to inflict significant infections on a power system. Simply put, an effort to phony up the consumption data of mining devices results in the furtherance of messing up the optimal energy management within the power system. Hence, this paper introduces a new cyber-attack named miner-misuse for power systems equipped by transaction tech. To overwhelm this dispute, this article also addresses an online coefficient anomaly detection approach with reliance on the reinforcement learning (RL) concept for the power system. On account of not being sufficiently aware of the system, we fulfilled the Observable Markov Decision Process (OMDP) idea in the RL mechanism in order to barricade the miner attack. The proposed method would be enhanced in an optimal and punctual way if the setting parameters were properly established in the learning procedure. So to speak, a hybrid mechanism of the optimization approach and learning structure will not only guarantee catching in the best and most far-sighted solution but also become the high converging time. To this end, this paper proposes an Intelligent Priority Selection (IPS) algorithm merging with the suggested RL method to become more punctual and optimum in the way of detecting miner attacks. Additionally, to conjure up the proposed detection approach's effectiveness, mathematical modeling of the energy consumption of the mining devices based on the hashing rate within BFTS is provided. The uncertain fluctuation related to the needed energy of miners makes energy management unpredictable and needs to be dealt with. Hence, the unscented transformation (UT) method can obtain a high chance of precisely modeling the uncertain parameters within the system. All in all, the F-score value and successful probability of attack inferred from results revealed that the proposed anomaly detection method has the ability to identify the miner attacks as real-time-short as possible compared to other approaches.

Keywords: reinforcement learning; FDI attack; mining device; power system; UT method; attack detection

MSC: 49-11

Citation: Almalaq, A.; Albadran, S.; Mohamed, M.A. An Adoptive Miner-Misuse Based Online Anomaly Detection Approach in the Power System: An Optimum Reinforcement Learning Method. *Mathematics* **2023**, *11*, 884. https://doi.org/10.3390/math11040884

Academic Editors: Gurami Tsitsiashvili and Alexander Bochkov

Received: 8 January 2023
Revised: 6 February 2023
Accepted: 7 February 2023
Published: 9 February 2023

1. Introduction

Nowadays, on account of widely utilizing the smart varied business tools for social welfare enhancement, there is a need to develop an environmentally friendly financing bed that can guarantee financial affairs in a risk-free and reliable way. Meanwhile, launching a financial trading network based on the platform of blockchain technology such as the Bitcoin systems needs to catch the remarkable energy consumption through the mining devices. Likewise, such technology being a scale-great load demand is technically known as another major difficulty for the power systems. This results in the electrical grid facing optimal energy scheduling for not being aware of the consumed power and seating within

the grid configuration. Recently, energy management based on blockchain became an attention center for researchers. However, the deployment of blockchain tech does not entrust a straightforward choice owing to the challenging obstacles in the way. Thus, the current study tried to settle the issue that blockchain technology may be successful in optimal energy management, whereas blockchain tech was introduced as an energy consumer for the power system. One aspect of the problem that has yet to be considered is the cyber-injurious intrusion in the form of making spurious mining devices and false data injection (FDI) into the defined financial trading platform, aiming to irregularly enlarge the load demands of the electrical grid [1]. This issue can emerge new inescapable challenges facing energy management that need to be dealt with. In the literature, various cyber-attack detection approaches are investigated. Looking over the above-mentioned considerations, the literature of the relevant works is addressed in the following.

1.1. Cyber-Attack Detection in the Power System

For the last few years, cyber-threats focused on the electrical grid's performance for the sake of depending on industrial technologies in every country. Hence, the cyber-attacks of FDI type are one of the most efficient ways to launch widespread sabotage in the varied fields of the cyber-physical system, e.g., the control system [2–5], energy markets [6], electrical grid [7,8], and so forth, in the sense that FDI attacks may bring irrecoverable injuries involving the electrical grid versus the modest cyber cost involved by the hackers. Indeed, FDI attacks tend to manipulate sensing and measuring info with the aim of improperly reading into the monitoring system [9]. Broadly speaking, most attackers, known as denial of service (DOS) attacks [10–12], aim to mislead the data estimation process to get into the all-embracing destruction, e.g., the wide-area blackouts in the power systems. The DOS attacks are able to decline the chance of having access to the system for meter devices [13–15], whereas, jamming attacks manipulate the process of data measurement by means of applying noise [16]. On the basis of blockchain tech, it is utilized in many applications, such as communication structures, smart contracts, the Internet of Things, electronic voting, and so on. Despite these all applications, blockchain is vulnerable to various cyber-attacks classified under the following: (1) attacks related to the used mathematical techniques (stale and orphaned blocks), (2) attacks of peer-to-peer structure (51% attack, distributed DoS attack), and (3) attacks of the application context [17]. In the literature, many reports have expressed the increase in data security risks for blockchain tech. For instance, a hacker in June 2016 could manage to steal USD 50 million from "The DAO" [18]. As mentioned already, on account of the intricate looping structure of the electrical network, any subversion or failure in the grid's vulnerable points lead to cascading damages over the grid for a short period of time. Keeping these in mind, it seems that timely anomaly detection is critical once the grid continually suffers cyber-attacks. In this case, quickly performing the detection method is of crucial significance in assessing the cyber-attack detection methods. In a more proper way, when a variation occurs in the sensing data at an unknown hour, the goal is to trace the data variation as soon as possible with emphasis on lowering the false alarm for the measuring devices that is available over time. Another matter of great concern that needs urgent attention is that the detection speed and detection accuracy are in conflict with each other once launching a cyber-attack. Conversely, it is very significant to find a way to reconcile these two conflicting views. The reinforcement learning (RL) concept is defined with the aim of controlling stochastic environments. For this reason, the observable Markov decision process (OMDP) problem may be solvable by inspiration from the RL concept [19,20]. Indeed, learning the underlying OMDP problems can be achieved by a model-based RL algorithm [21–25].

1.2. The Mining Devices

Satoshi Nakamoto proposed a Bitcoin-based cash trading system guaranteed by blockchain tech in 2008 [26]. A blockchain network is defined based on the ledgers maintained by the wide nodes in the form of a distributed structure [27]. Blockchain has the

needed ability to prove transaction security and trust in wide usages, e.g., smart healthcare, real estate, the logistics industry, and so forth. The consensus algorithm shows that the proof of work (PoW) is the most widely and commonly used in the blockchain and is able to check the bilateral trust among nodes within the network [28]. Indeed, the PoW mechanism brings high security with the help of working the mining devices while consuming lots of power during the hashing computation. For instance, the literature indicates that the bitcoin-based financial trading system (BFTS) takes an electrical consumption of about 1–32 TWh per year, whilst the power consumption of Denmark is estimated to be up to 32 TWh. This means that BFTS, notwithstanding providing a safe mutual fiscal bed, can turn into a serious menace to the electrical grid as a large-scale load demand. Hence, in [29], the authors tried to indicate a new consensus algorithm called proof of stake (PoS) for application in blockchain tech. All in all, new investigations are working on and starting to eliminate this concern.

1.3. Motivation and Contributions

In view of the fact, it is possible that the data monitoring center is overshadowed by cyber-attacks aimed at making counterfeit data instead of documenting it. Thus, it seems that the monitoring data need to be checked before being employed by the decision-maker. As noted, despite that the blockchain technology could guarantee data security for the power system through distributed data management, it can be vulnerable and can be a target for cyber attackers thanks to the presence of high-consumption miners. Accordingly, having a close look at the data detection mechanism to counter it is a must. This paper aimed at, first, introducing a new attack based on the consensus structure of blockchain and proposing an adaptive detection mechanism to deal with it. A cyber nascence attack called miner misuse was introduced here for energy systems utilizing the blockchain security structure. Indeed, the adversary's intent of deliberately increasing the miner's energy is problematic for the energy systems secured by blockchain tech. So to speak, a cyber attacker can often lead astray the monitoring operator by way of false data injection related to the miner energy. This work is possible in making misprision decisions about energy management. As such, to counter the aforementioned concern, the monitoring data related to the miners' energy consumption must be put into a checking way. Meanwhile, this paper captured a miner data detection method inspired by the reinforcement learning concept to raise the chance of trustiness for energy systems. Inasmuch as the learning mechanisms are in essence unable to catch the optimal solution, they need to be vouched by the optimization approaches. To this end, this paper followed another goal based on introducing an Intelligent Priority Selection (IPS) optimization algorithm melded with the learning mechanism. Hence, to differentiate this paper compared to others, Table 1 briefly indicates the main differences.

Table 1. Categorization of the different approaches.

	Block Chain	Cyber Attack	RL Mechanism	Optimization Method	Uncertainty
[17]	✓	✓			
[30,31]	✓	✓			✓
[32–34]			✓	✓	
[35–37]				✓	
Proposed model	✓	✓	✓	✓	✓

Therefore, the underlying contributions and characteristics of this paper concerning previous publications are categorized as follows:

- Introducing and mathematically modeling a new cyber nascence attack based on the miner misuse of the blockchain mechanism whose aim is data security within the smart energy system.
- Evolving an effective and adoptive IPS-RL-based online anomaly detection method to pinpoint the unknown malicious intrusions in the power system based on BFTSs called the miner misuse.
- Presenting the UT-based uncertainty approach to conjure up the high-risk energy consumption of the mining devices and check up on their effects on energy management.

The rest of the paper is provided such that the mathematical formulations of the studied model are described in Section 2. Section 3 expresses the proposed detection method. The solution process is represented in Section 4. Section 5 explains the UT-based uncertainty method. The effectiveness of the studied model proposed in this paper is evaluated and validated in Section 6, and Section 7 provides the conclusion.

2. The Modeling Structure of the Smart Grid Based on BFTS

2.1. The Basic Framework of the Power System

Recently, there is a considerable effort employing smart modern tools in the daily life of people in a way that social welfare would be on the rise. Developing the Bitcoin-based financial trading system (BFTS) can be a manifest case of the above considerations. For all these reasons, there is a tendency toward growing load demands in the power system, as shown in Figure 1. For instance, BFTSs need to employ mining devices for securing goals while taking the significantly consumed energy for hashing and blocking information into the well-known blockchain process. Developing BFTSs along with increasing the number of mining devices causes unavoidable challenges facing modern power systems to emerge. Ignoring the energy consumption of the mining device in energy scheduling management makes it impost the irrefutable limitations such as the high operation cost, line congestion, uncompromising power generation, and so forth on the power systems [38]. To realize the effects of the mining devices on energy management, a closer look at the energy operating framework of the electrical grids is needed. Hence, this section is dedicated to explaining how to meet the load demands by the generation units with an emphasis on satisfying the relevant constraints and the objective function followed by the power system. Keeping this in mind, the fossil fuel-based traditional generation units are operated by an energy control center aimed at following the relevant objective function with an emphasis on power limitations as defined by Equations (1)–(13). In this case, the generation scheduling goal described in (1) indicates that these generation units can catch in/out permits and the generated power in line with minimizing the operation cost and the relevant shot-down/start-up costs. Indeed, all generation units would be operated with an emphasis on the price/generation curve fulfilled by the electricity market, but this goal is satisfied by looking at the transmission grid limitations. Meanwhile, one of the significant issues that needs urgent attention concerns the power limitations related to power generation and ramping. These constraints are modeled by Equations (2)–(5) in the operating framework. It may be true that the most important factor for holding the grid power stability is regularly checking the generated/consumed power variation during the grid operation. Hence, Equation (2) defines balancing the swirling active power among the resources, the flowing power of lines, and load demands, including the public loads and the consumed power of the mining devices. Conversely, this elucidation can be accurately inferred for the reactive power as shown in Equation (3). Equations (4) and (5) are the limitations of the generated powers. This means that the generation units should adhere to keep their generated powers to a minimum/maximum level due to the technical conditions. Of course, each generation unit is allowed to ramp up the output power with the aim of covering the critical load fluctuations. Thus, we modeled the ramping power limitations of generation units based on Equations (6) and (7). It is a right assertion that the grid lines are connected in the form of a looping structure with the goal of coupling the generation units and loads [38]. With this description, there is a need to model in which line l the power

of bus n is injected/received to/from. In this regard, Equations (8) and (9) illustrate the active/reactive power injection based on the bus voltage and the technical properties of lines. Needless to say, one underlying concern for grid operators is the overloading of lines once raising the period of peak load. To counter this problem, the restriction of power flow is observed by constraints in Equations (12) and (13). Finally, to achieve voltage stability and optimal operation, it is necessary that the bus voltages and angles are restricted up to a permissible level based on Equations (10) and (11).

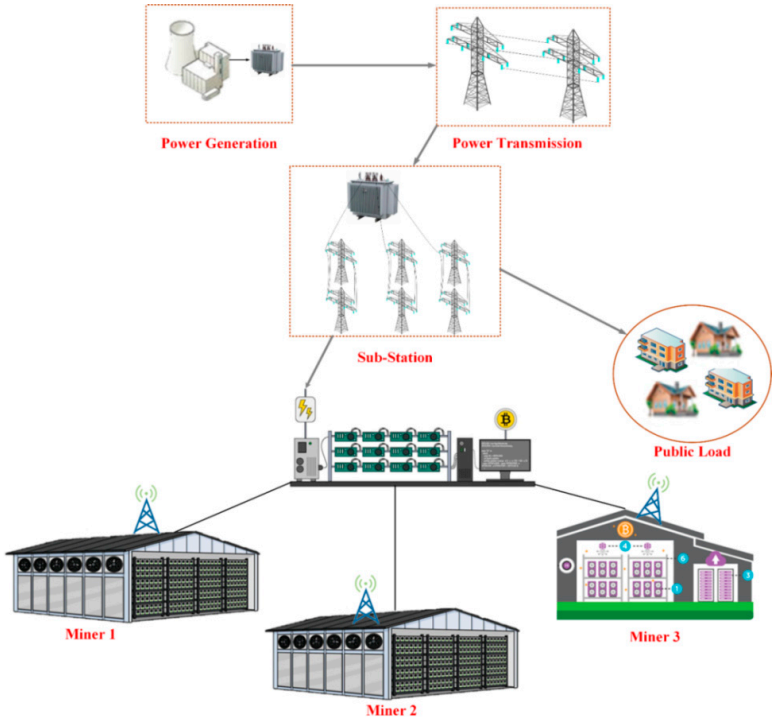

Figure 1. Illustration of the electrical grid based on the mining devices.

$$min\ TC^{grid} = \sum_{t}\sum_{g}\left[f^{c}\left(P_{t,g}\right) + S^{up}_{t,g} + D^{down}_{t,g}\right] \tag{1}$$

$$min\ TC^{grid} = \sum_{t}\sum_{g}\left[f^{c}\left(P_{t,g}\right) + S^{up}_{t,g} + D^{down}_{t,g}\right] \tag{2}$$

$$\sum_{g}\left(P_{t,g}\right) - \sum_{l}\left(P^{L}_{t,l}\right) = P^{L}_{n,t} + P^{mining} \qquad \forall t \in \Omega^{T}, \forall n \in \Omega^{n} \tag{3}$$

$$\sum_{g} Q_{t,g} + \sum_{l}\left(Q^{L}_{t,l}\right) = Q^{L}_{n,t} \qquad \forall t \in \Omega^{T}, \forall n \in \Omega^{n} \tag{4}$$

$$P^{min} z_{g,t} \le P_{t,g} \le P^{max} z_{g,t} \qquad \forall t \in \Omega^{T}, \forall g \in \Omega^{g} \tag{5}$$

$$Q^{min} z_{g,t} \le Q_{t,g} \le Q^{max} z_{g,t} \qquad \forall t \in \Omega^{T}, \forall g \in \Omega^{g} \tag{6}$$

$$P_{t,g} - P_{t-1,g} \le D^{+}_{G} z_{g,t-1} \qquad \forall t \in \Omega^{T}, \forall g \in \Omega^{g} \tag{7}$$

$$P_{t-1,g} - P_{t,g} \le D^{-}_{G} z_{g,t} \qquad \forall t \in \Omega^{T}, \forall g \in \Omega^{g} \tag{8}$$

$$P^{L}{}_{l} = (V_{n} - V_{m})R'_{l} - X'_{l}\eta_{n} \qquad \forall n \in \Omega^{n}, \forall m \in \Omega^{m}, \forall l \in \Omega^{l} \tag{9}$$

$$V^{min} \le V_{n,t} \le V^{max} \qquad \forall t \in \Omega^{T}, \forall n \in \Omega^{n} \tag{10}$$

$$\eta^{min}_{n} \le \eta_{n,t} \le \eta^{max}_{n} \qquad \forall t \in \Omega^{T}, \forall n \in \Omega^{n} \tag{11}$$

$$-P^{L\_max}_{l} \le P^{L}_{l,t} \le P^{L\_max}_{l} \qquad \forall t \in \Omega^{T}, \forall l \in \Omega^{l} \tag{12}$$

$$-Q_l^{L\_max} \leq Q_{t,l}^L \leq Q_l^{L\_max} \qquad \forall t \in \Omega^T, \forall l \in \Omega^l \tag{13}$$

2.2. The Mathematical Definition of the Mining Device

As recently mentioned, supplying the power involved in BFTSs, thanks to the mining devices, is a significant concern in energy management that needs to be dealt with. Indeed, mining devices are made up of a set of central processing units (CPUs), all of which have a duty in line with finding a solution for the targeted hashing puzzle. Broadly speaking, the needed power in this process is practically referred to as the computing power of these CPUs for cyclically solving the specified puzzle. For effective realization, this section aims to mathematically develop the mining device's behavior. To this end, the CPU's energy consumption was calculated based on the frequency and voltage associated with Equation (14) in the first place.

$$P^{mining} = Su^2 f \tag{14}$$

where:

- P^{mining}: The consumed power.
- S: A constant value related to the computing device.
- u: Voltage.
- f: Frequency.

In (14), the operating frequency is a variable taking into account the cycle number required for the hashing process per second. Meanwhile, assume the frequency and power consumption are in a direct way; as a result, Equation (14) with a slight change can substitute into Equation (15):

$$P^{mining} = qf \rightarrow q = Su^2 \tag{15}$$

By having considered $f = hw$, the aforementioned formulation is updated as:

$$P^{mining} = hq\omega \tag{16}$$

Additionally, the share in percentage related to CPU i for a mining device is achieved by the following:

$$\eta = \frac{P_i^{mining}}{\sum_{j \in N} P_j^{mining}} \tag{17}$$

Looking over the process, each CPU i takes a remarkable chance of solving the targeted puzzle per second. By adopting the Poisson distribution to the mining devices, the successful likelihood of CPU i can be obtained based on Equation (18):

$$\underset{i}{\mathrm{P}}(\mathrm{r}) = \eta e^{-\lambda \beta t r_i} \tag{18}$$

In (18), obtaining a successful CPU in the solution is subject to be altered, with an emphasis on the number of transactions inserted into the data block and constant ψ. It is worth pointing out that mining devices are encouraged to catch gains as tax in the form of public service if rapidly puzzling out compared to others. Nutshell, the final benefit arising from each mining device based on the power consumption is earned as follows:

$$bif_{B_i}\left(P_i^{mining}\right) = \frac{TP_i^{mining} e^{-\lambda \beta t r_i}}{\sum_{j \in N} P_j^{mining}} - y_i P_i^{mining} \tag{19}$$

3. The Miner-Misuse-Based Detection Approach

In the previous section, it was expressed how the mining device's performance is in line with the energy scheduling for the electrical grids. By having a look at the high-risk energy consumption related to these mining devices, one obtains a practical realization of

badly overshadowing energy management if the abnormally growing trend in employing mining devices is ignored. This concern can be turned into a high chance of imposing cyber-demolitions within the power system structure. Indeed, an adroit hacker in an effective effort can misuse this opportunity and can launch a bogus mining device called miner-misuse as the load demand in the electrical grid. For this reason, it seems that developing the BFTS-based electrical grid structure relying on a misuse detection approach to prevent any attack launch is a must. Hence, this section is dedicated to presenting the proposed RL-based online detection method aimed at meeting the miner-attack alarum.

3.1. The Definition of the Proposed RL Concept-Based Detection Approach

Recently, the learning machine in the field of cyber security illustrates has been utilized effectively in various industrial applications. Needless to say, learning machine problems are mainly defined in different forms of the learning concept, i.e., (1) supervised learning, (2) reinforcement learning (RL), and (3) unsupervised learning. In the literature, the RL method is introduced as an efficient and reliable security policy due to the learnable agent's employment in cooperating with the environment. Indeed, it was proven that the learning process could be more improvable with the chance of having environmental data. So to speak, the stepping learnedness concerning the penalty/reward approach makes the agent more reliable for realizing the environment as well as possible. Thus, as the agent receives the environmental data, it gives the corresponding retroaction in line with the relevant goal as well. Afterward, obtaining feedback on the agent's performance can be the best and most far-sighted solution for allotting the penalty/reward coefficients. An agent takes rewards on account of offering acceptable output within the environment. In contrast to gaining a reward, an agent would be fined due to misstating the fact of the environment. This work can help the agent recognize the environment where it is put into. To realize how the RL method works in action, Figure 2 shows briefly the RL approach's performance [32].

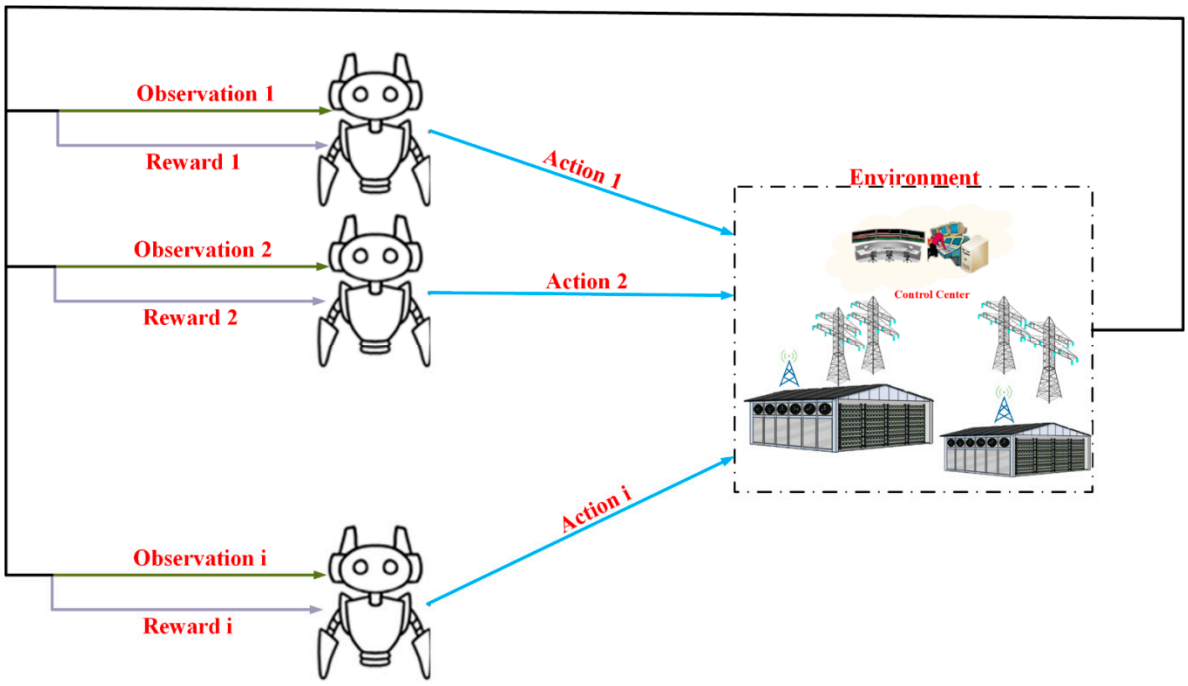

Figure 2. The reinforcement learning structure.

As can be seen, the launching of the RL method is defined by two spaces: (1) environment and (2) agent. Additionally, in order to declare anomalies in the environment, the RL method is carried out in reliance during the learning phase and the online detection phase. The learning phase needs to be described in the first place. In this step, the agent learns which environment observation each appropriate action belongs to. the agent takes a

reward from the environment as a result of the action selection, which guides the agent on how to modify its selection process versus the environmental observations. This learning phase goes toward enabling the agent to pick out the best action under an unknown environment based on uncertain states. Generally, in this step, the RL elements' performance is respectively served below:

The agent: (1) receives the environmental observation; (2) chooses the equivalent action; (3) takes a reward for the environment.

The environment: (1) indicates the observation; (2) examinates the relevant action; (3) releases a reward in line with the agent's action.

As mentioned, such an RL method is aimed at merely learning the agent by means of the unknown environmental observations that it needs to delineate in detail. This concept is inspired by the OMDP idea in which the environment is invisibly defined. Having said this, the elements made up of an OMDP problem are marked out based on the concealed states (s), environmental observations (o), the transition probability of the agent from a state to another one (T), the reward (r), and the agent's action (a). In such a problem, firstly, the agent deploys into the unknown environment and receives the relevant observations, and then it selects a related but not optimal action. By doing so, the environment obtains a reward related to the agent's action and the current state at t.

Now, after defining an OPMP problem, there is a need to express how such a problem can be mapped for the attack detection method. Hence, Figure 3 briefly illustrates the OMDP-based detection problem framework that is proposed in this paper. In this way, assume that a hacker could obtain access to data and launch an anomaly in reliance on the unknown method in the system at $t = k$. Due to the nameless attack strategy, let us suggest the environmental conditions in varied states based on 'prior-attack', 'later-attack', and 'terminal'. In each state, the agent can choose one of the 'continue' and 'stop' actions with an emphasis on the environmental observation. If the agent has the accurate election considering the current state, it receives reward = 0 from the environment and moves toward the 'terminal' state, which means the best choice. If so, the agent takes reward = 1 or reward = b concerning the current state due to the evil detection that needs amendment. Keeping these in mind, the goal followed by the agent in the approach is functioned by Equation (20):

$$minR^{pen} = E^{\kappa}\left[(re_t|t_s < k_t) + \sum_{t=\kappa}^{\infty} b|t_s > k_t\right] \tag{20}$$

where t_s and R^{pen} indicate the stopping time and the relevant reward emitted by the environment. It is important to say that the reward is fixed based on the detection time once an attack is launched on the system at $t = k$. For future realization, the learning process is explained in detail. Assume the agent's state is the 'prior-attack' state in the first place. Meanwhile, if the agent selected the 'continue' action, it would take *reward = b* owing to the ineptness in describing the anomaly at $t_s > k$. Conversely, *reward = 1* will be emitted for the agent if it badly adopts the 'stop' action when it is exposed to the 'later-attack' state at $t_s > k$. Therefore, in the learning phase, for each action-observation pair, it is suggested to learn a *P (o, a)* value by using the RL algorithm. All in all, the learning phase trend is briefly stepped by:

(1) Choose the arbitrary $P\ (o^t, a^t)$;
(2) Check the stopping time and alarm based on the above explanations;
(3) Collect the measurement X_t;
(4) Obtain the observation signal (o^{t+1}) and choose the action a^{t+1} based on the ε-greedy policy for which it is determined by the minimal $P\ (o^{t+1}, a^{t+1})$;
(5) Update the $P\ (o^t, a^t)$ value based on $P\ (o^t, a^t) = P\ (o^t, a^t) + \alpha\ (r + P\ (o^{t+1}, a^{t+1}) - P\ (o^t, a^t)$.

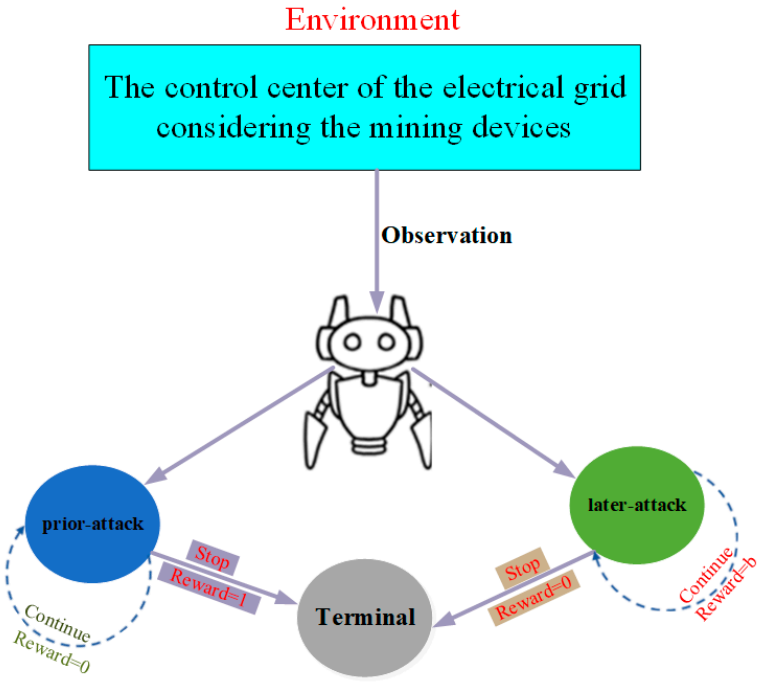

Figure 3. The anomaly detection method.

The previous phase, the agent needs the ability to detect every possible anomaly if it occurs in the system. In the detection phase, assume the agent observes signal o by relying on the 'prior-attack' state and chooses the 'continue' action. In the next step, the agent repeats steps 3–5 of the learning phase and opts for the best corresponding action (a). Here, there is a key point that needs an explanation. In step 5 of the learning phase, the updating process related to the p-value is dependent on the coefficient α, which is effective for being optimal in the learning phase. Hence, this paper suggests a coefficient Intelligent Priority Selection (IPS)-based optimization algorithm to obtain the appropriate coefficient α.

3.2. Intelligent Priority Selection Based Optimization Algorithm

The learning methods, especially the RL method, are unable to find the optimal solution. By keeping this in mind, this part tries to cover a concern of the learning phase in the RL-based attack detection algorithm with the help of an effective optimization method. Indeed, one main goal of this paper is to find a hybrid optimum detection method. In the literature, varied approaches have been proposed to obtain an efficient strategy for optimization issues through mathematical solving methods and artificial intelligence. However, the fairly low authenticity and the high time-passing and time-consumption are significant concerns involved in these methods. Hence, the proposed algorithm aimed to cover them. In the statistics, computing out of n from set N is shown below:

$$\begin{pmatrix} N \\ n \end{pmatrix} = \frac{N!}{(n!).(N-n)!} \tag{21}$$

This results from above show that the problem space for finding an optimal solution may be so much that it leads to a high time passage. This method guarantees the solution to be more reliable. To overcome this problem, the proposed IPS approach has the ability to mitigate the time-consumption as much as possible. This method was inspired by the smart random permutation concept to find out the best solution based on the candidate's answers [39]. Hence, the proposed method follows the following steps:

Step1: Let us assume P is defined as the possible candidate for solving a problem. Additionally, the initial solutions are randomly chosen and stored in matrix K. The rest of the solutions (P-K) are saved in matrix W. The probable solution sets result from substituting

the members of matrix W into matrix K, named KT as indicated in (25). Next, it is necessary to calculate the best solution of the objective function arising from the outcomes of the matrix KT indicated as $F^{best}_{W_1,K'_n}$ in (26). Equation (27) shows the sorted matrix related to the objective function calculation. Of course, the optimal solution is achieved when opting for the best answer among candidates [39].

$$P = [p_1, \ldots, p_N] \tag{22}$$

$$K = [k_1, \ldots, k_n] \tag{23}$$

$$W = [w_1, \ldots, w_m] \tag{24}$$

$$KT = \begin{bmatrix} \begin{matrix} {}^{k}/_{1} = k'_1 & & & \\ \uparrow & & & \\ w_1 & k_2 & \ldots & k_n \\ & {}^{k}/_{2} = k'_2 & & \\ & \uparrow & & \\ k_1 & w_1 & \ldots & k_n \\ & & & {}^{k}/_{n} = k'_n \\ & & & \uparrow \\ k_1 & k_2 & \ldots & w_1 \\ & & \downarrow & \\ & & F(H_{w_1}) = F^{best}_{w_{1,k''_1}} & \end{matrix} \Bigg\} H_{w_1} \ldots \quad \begin{matrix} {}^{k}/_{1} = k'_1 & & & \\ \uparrow & & & \\ w_m & k_2 & \ldots & k_n \\ & & \vdots & \\ & & & {}^{k}/_{n} = k'_n \\ & & & \uparrow \\ k_1 & k_2 & \ldots & w_m \\ & & \downarrow & \\ & & F(H_{w_m}) = F^{best}_{w_{m,k''_m}} & \end{matrix} \Bigg\} H_{w_m} \\ H_{w_i} = [H_{w_1}, H_{w_2}, \ldots, H_{w_m}] \quad \forall i \in \Omega^i \\ k''_M = [k'_1, k'_2, \ldots, k'_n] \quad \forall M \in \Omega^M \end{bmatrix}, \tag{25}$$

On the basis of the objective function matrix (27), the relevant elements of the matrix W are arranged and saved in the matrix W'_j as illustrated in (28). This process is reiterated for set K''_j (29). Finally, the first level of the matrix F^{best_sort} is considered the best answer (see (30)).

$$F^{best}_m = \begin{bmatrix} F^{best}_{w_{1,k''_1}} \\ \vdots \\ F^{best}_{w_{m,k''_m}} \end{bmatrix} \qquad \forall m \in \Omega^m \tag{26}$$

$$F^{best\_sort} = \begin{bmatrix} F^{best}_{w'_1 \to k''_1} \\ \vdots \\ F^{best}_{w'_m \to k''_m} \end{bmatrix} \tag{27}$$

$$w'_j = [w'_1, \ldots, w'_m] \forall m \in \Omega^m \tag{28}$$

$$k''_j = [k''_1, \ldots, k''_m] \forall m \in \Omega^m \tag{29}$$

$$F = F^{best}_{w'_1 \to k''_1} \tag{30}$$

Step2: In this step, the matrix KT is updated by the new matrix KT^{new}_r. In this way, updating the W_j matrix is conducted by (31) based on w'_{j+1} in the next nitration. Then, the new matrix $K1^{new}_j$ is computed by eliminating the w'_j from the matrix K_j (32). Merging matrixes KT^{new}_r and w'_j is in the next stage based on Equation (33), in which ψ_r shows the coming together of the matrices and r is 1 to $m - j$. Value m is the length of matrix W. For each element of the matrix ψ_r, the relevant objective function is calculated, and the best

answer ($F1^{Best}$) is obtained. As a result, the corresponding component of the matrix ψ_r related to $F1^{Best}$ is inserted in matrix K, as illustrated in (34)–(36).

$$W_j = w'_{j+1} \qquad \forall j \in \Omega^j \tag{31}$$

$$K1_j^{new} = \left\{ x \middle| x \in K_j, x \neq k''_j, x \neq w'_j \right\} \qquad \forall j \in \Omega^j \tag{32}$$

$$\psi_r = KT_r^{new} \cup w'_j \qquad \forall j \in \Omega^j, \forall r \in \Omega^r = [1, 2, \ldots, m-j] \tag{33}$$

$$F1_r = f(\psi_r) \tag{34}$$

$$F_j = F1^{Best} \qquad \forall j \in \Omega^j \tag{35}$$

$$K_j = \psi^{Best} \qquad \forall j \in \Omega^j \tag{36}$$

Step3: In this step, the most optimal solution is obtained through the last answer based on the objective function.

$$F^{best\_total} = F^{Best} \tag{37}$$

The step-by-step structure of the proposed IPS algorithm is shown in Figure 4 [40].

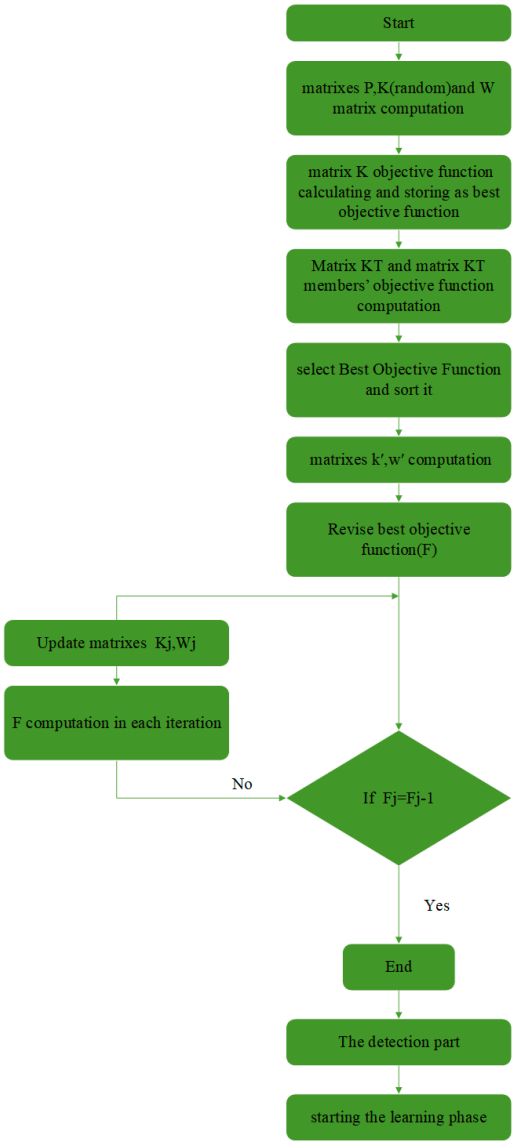

Figure 4. The framework of the IPS algorithm.

4. Stochastic Quantization Approach

Taking a close look at uncertainty effects on energy management can help to make the energy scheduling more faithful, more controllable, and so forth in the power system. To this end, this section takes inspiration from the UT approach-based uncertainty concept for modeling the relevant parameters such as the load demands and the consumed power of the mining devices in BFTSs. To recapitulate briefly, the UT method was designed as such that the points randomly generated are turned from the probability density function into the discreet distribution function [41]. Assume that the uncertain outcome of vector Q is obtained by the nonlinear function $T = \hat{f}(Q)$, in which Q comprises the number of parameters p with mean value z and covariance matrix A. The solving steps of the UT method for points $2p + 1$ are mainly explained as follows:

Step 1: In the first place, all points are calculated based on (38)–(40).

$$Q^0 = z \tag{38}$$

$$Q^k = z + \left(\sqrt{\frac{p}{1 - W^0} A_{aa}}\right)_k \quad k = 1, 2, \ldots, p \tag{39}$$

$$Q^{k+c} = z - \left(\sqrt{\frac{p}{1 - W^0} A_{aa}}\right)_k \quad k = 1, 2, \ldots, p \tag{40}$$

where A_{aa} is defined as the covariance matrix and $\overline{R} = z$.

Step 2: Computing the point Weight is achieved by (41):

$$W^k = \frac{1 - W^0}{2p} \quad k = 1, 2, \ldots, 2p \tag{41}$$

It is true that the sum of the weights should be equal to 1.

Step 3: By inserting these points into the nonlinear function $T^k = \hat{f}\left(Q^k\right)$, the stochastic outcome is computed by substituting these points into the function $\hat{f}\left(Q^k\right)$.

$$\overline{T} = \sum_{k=0}^{2p} W^k T^k \tag{42}$$

$$P_{TT} = \sum_{k=1}^{2p} W^k \left(T^k - \overline{T}\right)\left(T^k - \overline{T}\right)^R \tag{43}$$

5. Simulation Results

Exposing a discussion on widely developing BFTSs in modern social businesses and their effect on energy management in power systems are the underlying concerns in this paper. Indeed, the needed energy related to the mining devices to seek a puzzle in blockchain-based security platforms creates outstanding challenges in the spheres of generation scheduling and cyber-anomalies for the power systems. To be more precise, this weak spot can be an Achilles' heel for the cyber-hackers who put an end to energy management by faking out the mining devices in the face of energy consumption on the rise. Hence, this section is dedicated to indicating whether the proposed detection approach would be useful to cover this concern or not. To this end, we tried to carry out an IEEE 24-bus test system-based electrical grid [42,43] aimed at satisfying the public loads and needed power in the BFTSs. Next, we considered that an attacker spies out into the system to create the bogus mining device called miner-misuse for which the load demands would be spuriously increased. In this way, the proposed cyber-anomaly detection method's effectiveness in divulging the miner misuses was evaluated. Finally, regarding the high-risk energy consumption related to mining devices, we examined the uncertain effects on the

energy management in the last part. To briefly sum up, the relevant consequences are defined as follows:

Case I: Evaluating the energy management based on BFTSs;
Case II: Simulating the IPS-RL-based detection approach considering miner-misuses;
Case III: Modeling the uncertainty of mining devices and their effect on energy consumption.

6. Evaluating the Energy Management Based on BFTSs

As said already, BFTSs have the ability to safely trade off fiscal businesses while taking the great consumed energy insofar as is irrefutable in the generation schedule. For more realization, in the case study, we assumed the number of mining devices embedded in the BFTS was 24 devices, which are located in the varied areas of the grid. These devices provide the needed activity in response to users connected to the BFTS. Indeed, the mining devices assure the users the guarantee of true financial transactions among themselves by participating in a hard competitive scheme. Hence, the mining device owners may use varied computing tools to find the targeted puzzle that is victorious in this scheme. Thereby, every mining device has different energy consumption across a wide range. In light of this evidence, we carried out the studied case and illustrated the relevant simulation results in Figure 5 and Table 2. Figure 5 shows the consumed power for 24 mining devices, all of which were satisfied by the defined electrical grid in a 24 h horizon. As can be seen, the mining devices take the varied needed powers in the daytime due to the different activities in the way of finding the targeted puzzle's solution. Looking over the figure, it is inferred that device No.12 overtakes 338.9 kW in hour 22, which is the most consumption compared to the other devices. It is worth mentioning that the peak-energy consumption related to these devices is directly dependent on the hardship level of the targeted puzzle and may be at any time. Generally speaking, such fluctuations in the consumption profile exude outstanding challenges facing energy management. Hence, the results in Table 2 prove these elucidations. The total operation cost has an ascending change of almost 56.44% when considering the mining devices in energy management. This means that the effects of the presence of the BFTSs on the power system's performance are irrefutable.

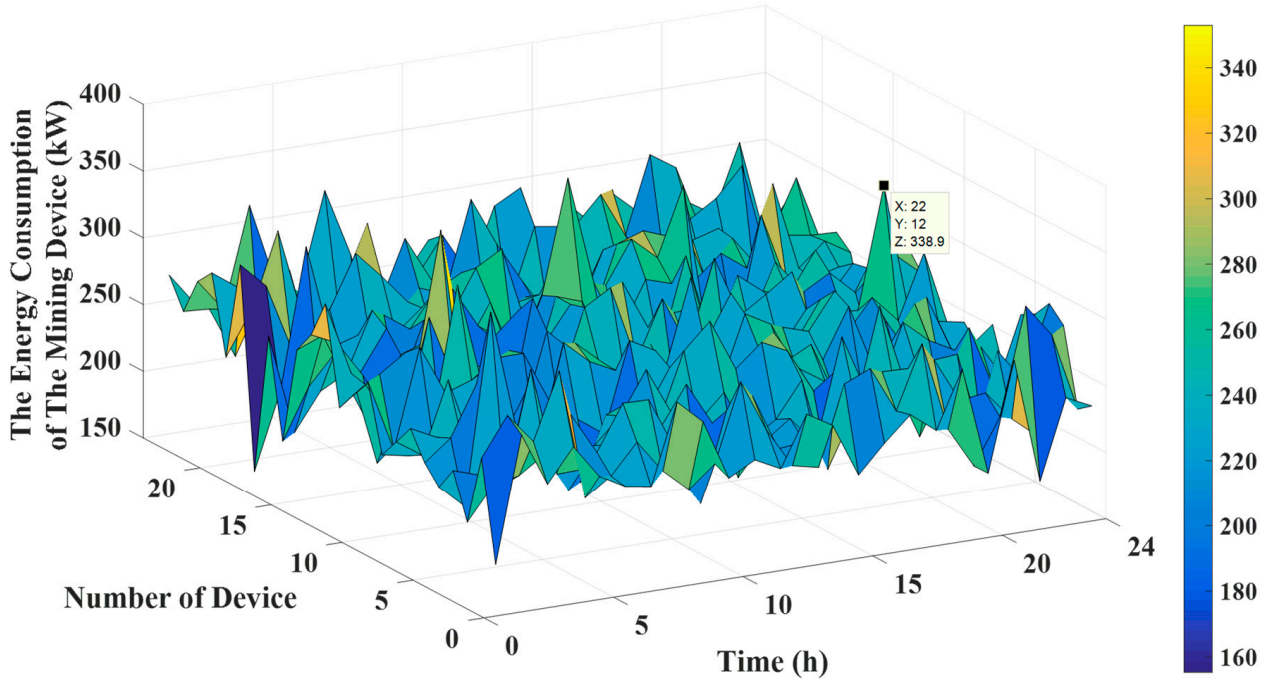

Figure 5. The energy consumption of the mining devices.

Table 2. Comparative results for the total cost.

Number of Case	Total Operation Cost (¢)
Considering the mining devices	$1.0793772038 \times 10^{10}$
Ignoring the mining devices	4.70×10^{9}

7. Simulating the IPS-RL Based Detection Approach Considering Miner-Misuses

As said already, in such a miner-based electrical grid, due to the high energy consumption of the miners, these devices can be known as a provocative factor for cyber-hackers who are able to bring to naught the energy scheduling process in power systems. Indeed, in this type of attack, intruders may carry out many bogus mining devices through deceitful data infusion into the control center, for which the energy consumption related to the miners takes an illusory tendency toward the ascending way. This assault is named the miner misuse in power systems based on BFTSs, and it needs to be dealt with. Hence, this section is dedicated to presenting how the proposed detection method would be worthwhile for tracking down the miner misuses in power systems based on BFTSs. To this end, we equipped the electrical grid with a security platform based on the proposed IPS-RL detection method. Meanwhile, we assume some hidden intrusions aimed at launching the miner misuse into the system have occurred. The absorbing consequences related to the proposed detection method's effectiveness in response to this attack are illustrated in Figures 6 and 7 and Table 3. In this way, Figure 6 demonstrates the results of monitoring the real energy consumption of the miner, which is compared to that manner being under attack. As can be seen, a cyber-hacker tried to launch the miner-attack at t = 80 s and could spuriously raise the energy consumption to a remarkable range of almost 25%. This means that the system stayed in the 'later-attack' state. In such a case, the IPS-RL platform was performed based on the defined phase 2 (see Section 3), and the agent unerringly chose the 'stop' action for which it could discover this anomaly at t = 88 s. The time-passing is calculated by the time delay between the start time and the end or the stopping time, which is 8 s. This shows that the proposed detection method not only is discrete and alarms anomalies, but it has the minimum stopping time. Besides that, we checked and evaluated this approach versus another attack at t = 123 s, as shown in Figure 6. In this situation, it could also detect the hidden intrusion for a lesser time delay of 5 s better than the other one. This advancement is obtained because the agent is learned by the prior anomaly and could provide superior performance. For more realization, Figure 7 indicates the energy consumption deviations for all attacks launched as well.

Table 3. Comparison among different trials.

%Number of Trials	Trials (1000)	Trials (1500)
False/Positive	%0	%0.09
True/Positive	%93.69	%91
False/Negative	%3.78	%4.84
True/Negative	%2.53	%4.07

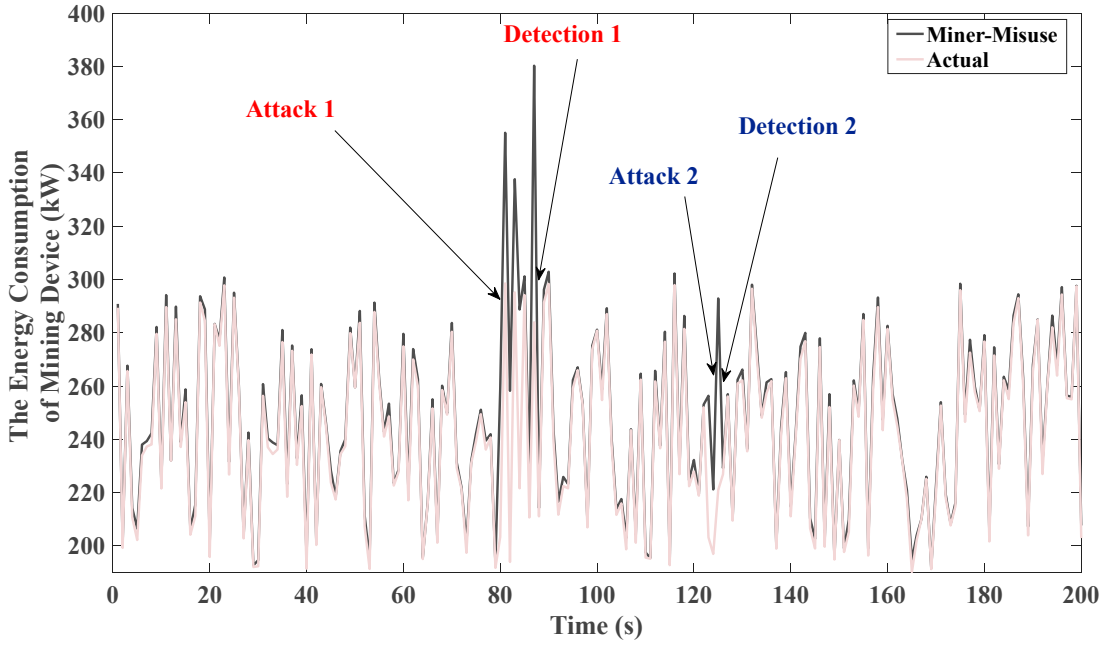

Figure 6. Illustration of the energy consumption under miner misuses.

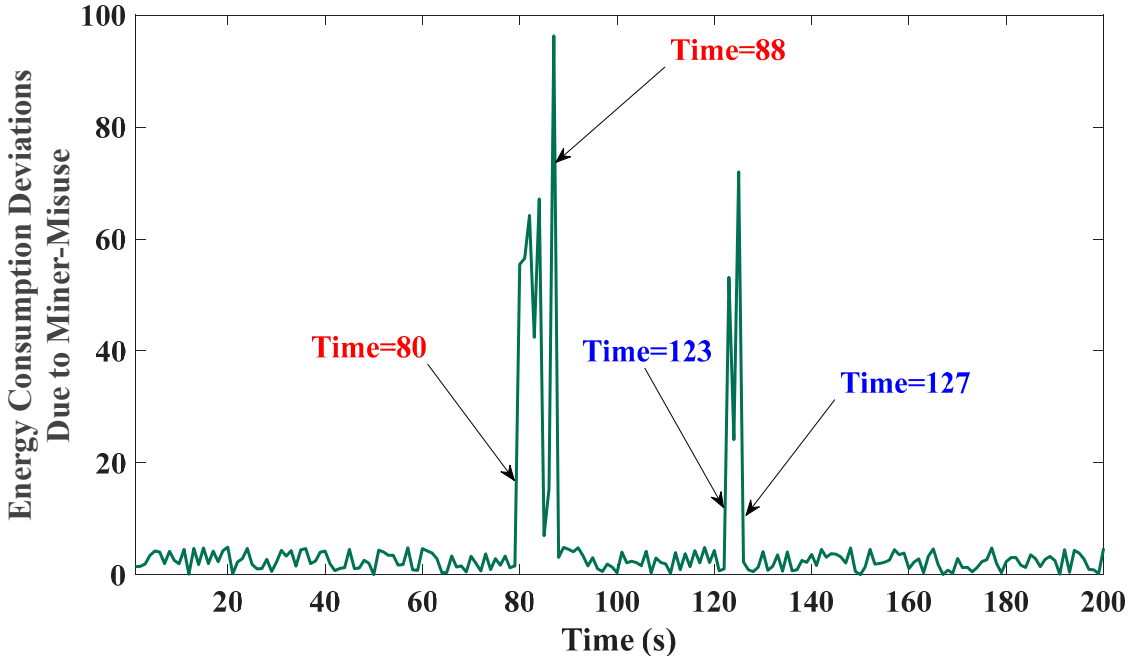

Figure 7. Illustration of the energy consumption deviations under miner misuses.

Speaking of the proposed detection method's effectiveness, it seems that the obtained results are not enough, and more tests and evaluations of this proposed method under many trials are needed for better decisions in assessing the security platform. To this end, we separately tested our approach for 1000 trials and 1500 trials here. The relevant consequences are provided in Table 3. In order to better perceive the performance, we tried to illustrate the platform's behavior for the true/false decisions in varied forms of false/positive, true/positive, false/negative, and true/negative. It can be seen that our method could accurately make the true/positive arbitration for 93.69% of trials when launching many malicious attacks on the system. On the other hand, this method unfortunately detected and erroneously stopped the system for 3.78% of trials when the system was clear, while it was 2.53% to take the true decision in the same condition. The second

column of Table 2 indicates the same results for 1500 trials. Here, it is seen that the agent is successful and useful in the true decision for 91% of 1500 trials, which proves our detection method's effectiveness, as mentioned in the previous section. Another key factor that can make a profound impact on it being optimal and efficient in the RL-based detection method is setting the main parameters. Figure 8 shows the simulation result of the IPS-RL-based detection method considering the miner attack for the varied coefficients of α. Comparing the detection times obtained by the algorithm for α = 0.1, 0.5, 1, and 10, it proves that an optimal value α can remarkably be effective for mitigating the stopping time in the anomaly detection method's performance. As can be seen, the proposed algorithm based on α = 0.1 had a minimum time delay of 8 s, while it was obtained at 13 s for α = 10. This shows that the IPS optimization algorithm can optimally handle the learning procedure in the RL-based anomaly detection approach. It is worth saying that various optimization algorithms were suggested to catch the best solution in the literature. On the whole, we mainly found two underlying approaches that capture the optimal answer, including (1) mathematical methods and (2) meta-heuristic methods. On account of involving meta-heuristic methods in the local optimal solution for some studied cases, the first approach is referred to with an emphasis on accuracy. Meanwhile, we tried here to provide a mathematical optimization method named the IPS algorithm. To realize the effectiveness of the proposed algorithm, we plotted Figures 9 and 10 as the compared results of the convergence process for the various optimization methods. Our method experienced a shorter convergence process in comparison to other methods. Besides that, in order to assess the IPS optimization algorithm in the learning phase, we compared our detection method based on the IPS, PSO, and Bat optimization algorithms and indicated the F-score coefficient and the converging process for them. Figures 9–12 show the simulation results. It is important to say that all the above algorithms start with the same initial population to have an equitable comparison. According to this figure, the IPS algorithm not only first converged compared to the PSO and Bat algorithms, but it had a better F-score coefficient (indicating the ability to escape from the local optima). To validate the proposed method, this paper compared it to other detection approaches, i.e., the support vector machine (VSM) and reinforcement learning (RL). To this end, it is necessary that the F-score is calculated for trial numbers through the assessment of the precision versus and recall curves. Looking over Figure 12, it is inferred the IPS-RL method can be more sensitive to detecting the attack than the other models.

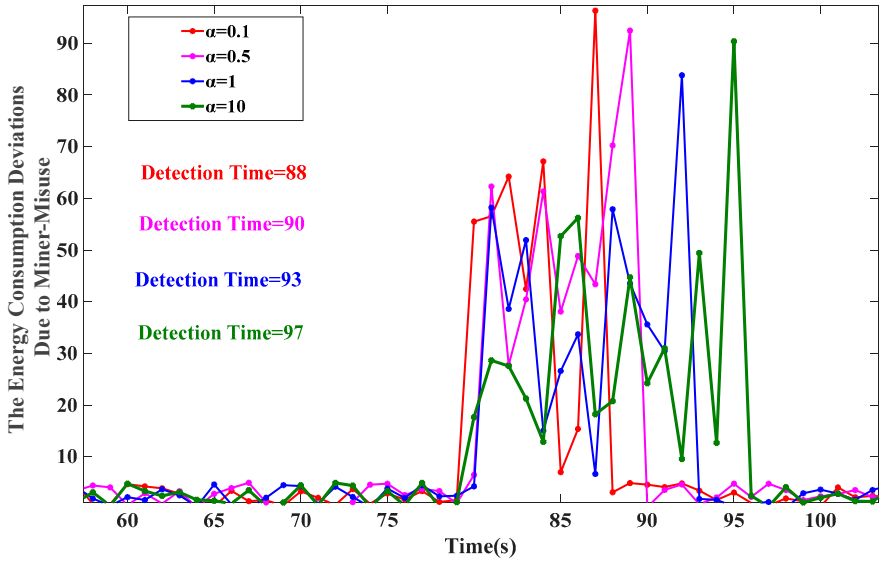

Figure 8. Illustration of the energy consumption deviations for different α.

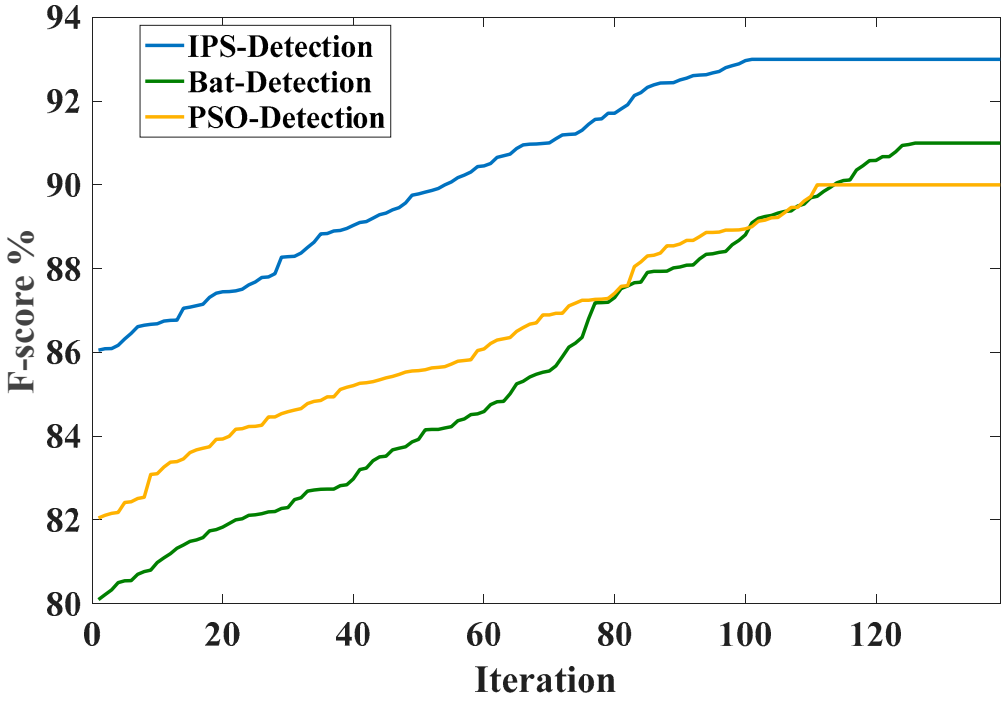

Figure 9. Evaluating the different optimization algorithms for F-score.

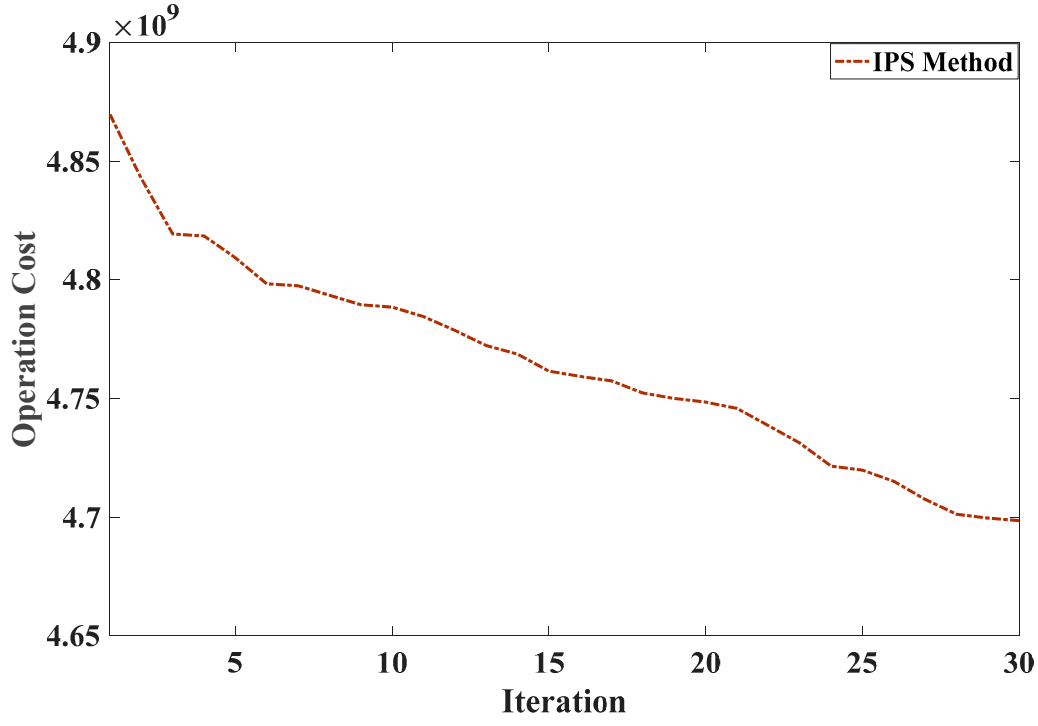

Figure 10. Convergence process of the IPS algorithm.

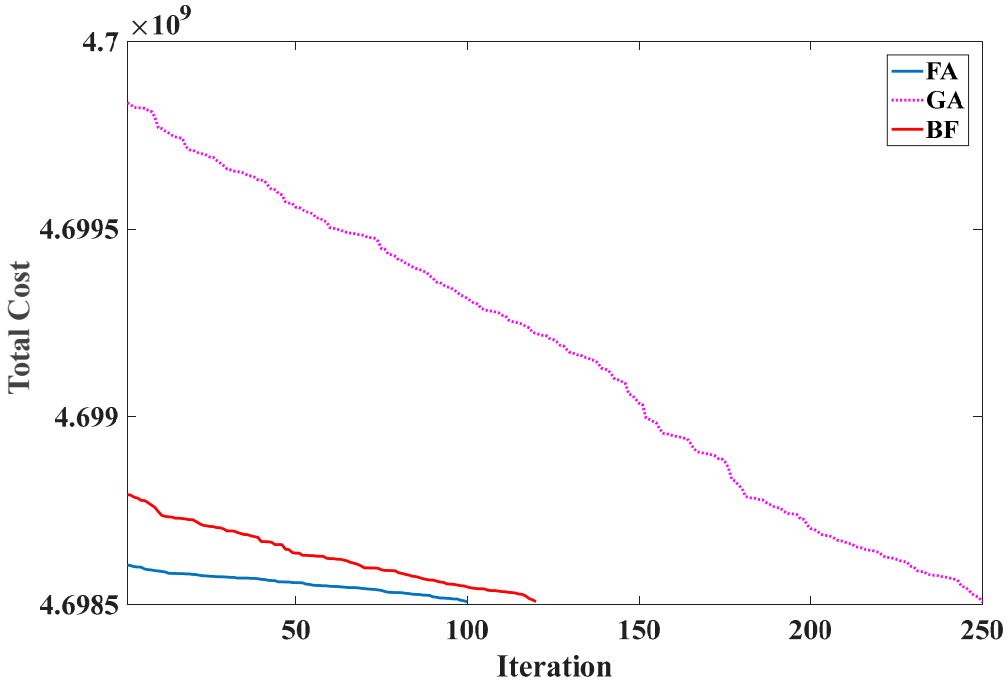

Figure 11. Comparison of algorithms for the convergence algorithm.

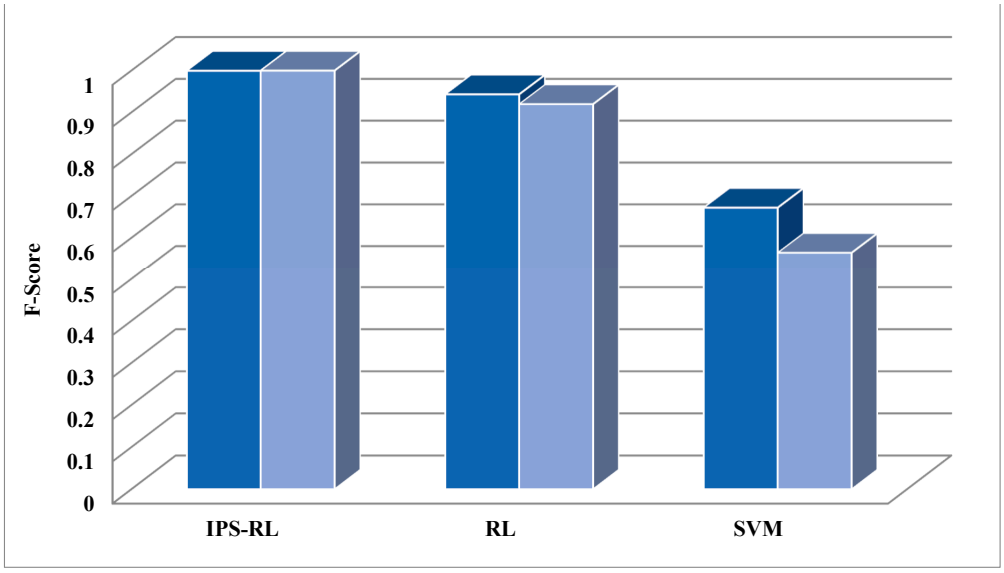

Figure 12. Illustration of the compared results for different detection approaches.

8. Modeling the Uncertainty of Mining Devices and Their Effect on the Energy Consumption

This section is dedicated to taking a close look at the effects of the uncertain energy consumption related to the mining devices on energy management within the electrical grid based on BFTSs. As stated already, the mining devices' behavior in the solving process of a targeted puzzle based on the blockchain tech causes them to make their energy consumption different and unstable. Hence, having considered the high energy consumption of these devices, modeling their uncertainty can remarkably increase the chance of having an optimal energy scheduling framework in the electrical grid. To clarify the truth of the matter, we modeled the uncertain parameters inspired by the UT concept and provided the simulation results in Figures 13 and 14. As can be seen in Figure 13, the average energy consumption of all mining devices takes a remarkable fluctuation

of almost 33% for all times when modeling the uncertain parameters fixed into the grid structure. Additionally these changes can affect the total operation cost, which is on the rise as indicated in Figure 14. To put it in a nutshell, the energy management in the presence of the mining devices in the system needs to model the uncertainty for providing an optimal and accurate evaluation.

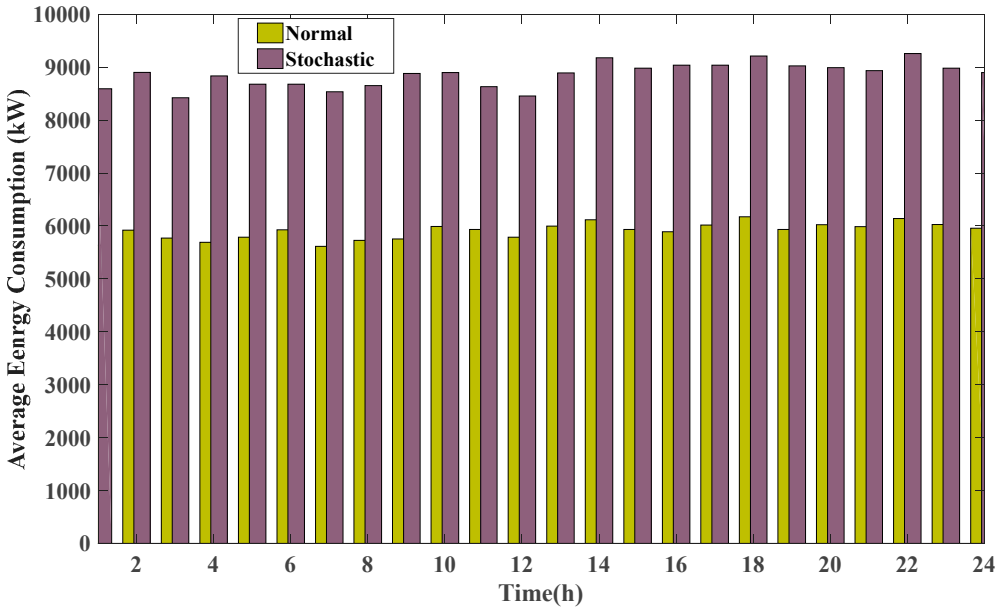

Figure 13. The average energy consumption of mining devices under uncertainty/normal conditions.

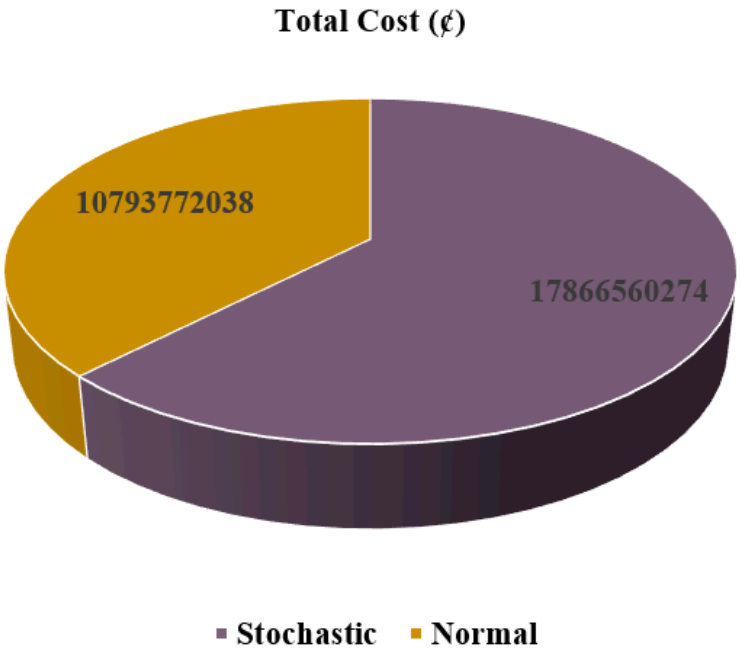

Figure 14. The total cost under uncertainty/normal conditions.

9. Conclusions

This article took an approach at an IPS-RL-based anomaly detection model to pinpoint the hidden malicious intrusions aimed to phony up the mining devices' behavior, which were named miner misuses within the power systems based on BFTSs. Indeed, in this type of anomaly, a cyber-hacker would try to make the bogus mining devices to create a spurious upward trend in the energy consumption, which is gobbled up by these devices to solve

a targeted puzzle within BFTSs. To prove the truth of the matter, we carried out an IEEE 24-bus test system-based electrical grid in furtherance of bearing out the public loads and consumed power in the BFTSs. After that, we assumed an intruder could make many bogus miners through deceitful data infusion into the decision-making center of the electrical grid for which the energy consumption was spuriously increased. Hence, consequences emanating from the proposed detection model in response to this anomaly were provided in the result section. They indicated that the proposed method could precisely pick out the true/positive detection for 93.69% of 1000 trials when launching malicious anomalies on the grid, while the agent learned by the proposed method made false decisions only for 3.78% of trials. This result was 91% and 4.84% for 1500 trials in the same condition. Besides that, looking over the results, the IPS optimization algorithm could obtain an F-score coefficient of 92%, which indicates that it is more useful to help in the optimal learning process based on the RL method compared to the other algorithms. Finally, the results show that energy management in the grid based on BFTSs would be more accurate and flexible if precisely modeling the uncertainty of energy consumption in the mining devices.

Author Contributions: Conceptualization, A.A., S.A. and M.A.M.; methodology, A.A., S.A. and M.A.M.; software, A.A., S.A. and M.A.M.; validation, A.A., S.A. and M.A.M.; formal analysis, A.A., S.A. and M.A.M.; investigation, A.A., S.A. and M.A.M.; data curation, A.A., S.A. and M.A.M.; writing—original draft preparation, A.A., S.A. and M.A.M.; writing—review and editing, A.A., S.A. and M.A.M.; visualization, A.A., S.A. and M.A.M.; supervision, A.A., S.A. and M.A.M.; project administration, A.A. and S.A.; funding acquisition, A.A. and S.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research has been funded by Deputy for Research & Innovation, Ministry of Education through Initiative of Institutional Funding at University of Ha'il, Saudi Arabia through project number IFP-22 099.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

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

References

1. Alnowibet, K.; Annuk, A.; Dampage, U.; Mohamed, M.A. Effective energy management via false data detection scheme for the interconnected smart energy hub–microgrid system under stochastic framework. *Sustainability* **2021**, *13*, 11836. [CrossRef]
2. Almalaq, A.; Albadran, S.; Alghadhban, A.; Jin, T.; Mohamed, M.A. An effective hybrid-energy framework for grid vulnerability alleviation under cyber-stealthy intrusions. *Mathematics* **2022**, *10*, 2510.
3. Zhang, J.X.; Yang, G.H. Low-complexity tracking control of strict-feedback systems with unknown control directions. *IEEE Trans. Autom. Control* **2019**, *64*, 5175–5182.
4. Zhang, X.; Lewen, D. Image enhancement based on rough set and fractional order differentiator. *Fractal Fract.* **2022**, *6*, 214.
5. Zhang, J.X.; Yang, G.H. Fault-tolerant output-constrained control of unknown Euler–Lagrange systems with prescribed tracking accuracy. *Automatica* **2020**, *111*, 108606. [CrossRef]
6. Almalaq, A.; Albadran, S.; Mohamed, M.A. Deep machine learning model-based cyber-attacks detection in smart power systems. *Mathematics* **2022**, *10*, 2574.
7. Alsokhiry, F.; Annuk, A.; Kabanen, T.; Mohamed, M.A. A Malware Attack Enabled an Online Energy Strategy for Dynamic Wireless EVs within Transportation Systems. *Mathematics* **2022**, *10*, 4691.
8. Soliman, M.S.; Belkhier, Y.; Ullah, N.; Achour, A.; Alharbi, Y.M.; Al Alahmadi, A.A.; Abeida, H.; Khraisat, Y.S.H. Supervisory energy management of a hybrid battery/PV/tidal/wind sources integrated in DC-microgrid energy storage system. *Energy Rep.* **2021**, *7*, 7728–7740. [CrossRef]
9. Rahman, M.A.; Mohsenian-Rad, H. False data injection attacks with incomplete information against smart power grids. In Proceedings of the 2012 IEEE Global Communications Conference (GLOBECOM), Anaheim, CA, USA, 3–7 December 2012; pp. 3153–3158.
10. Liang, G.; Zhao, J.; Luo, F.; Weller, S.R.; Dong, Z.Y. A review of false data injection attacks against modern power systems. *IEEE Trans. Smart Grid* **2017**, *8*, 1630–1638.
11. Wang, W.; Lu, Z. Cyber security in the smart grid: Survey and challenges. *Comput. Netw.* **2013**, *57*, 1344–1371.
12. Yan, Y.; Qian, Y.; Sharif, H.; Tipper, D. A survey on cyber-security for smart grid communications. *IEEE Commun. Surv. Tutor.* **2012**, *14*, 998–1010.

13. Kurt, M.N.; Yilmaz, Y.; Wang, X. Distributed quickest detection of cyber-attacks in smart grid. *IEEE Trans. Inf. Forensics Secur.* **2018**, *13*, 2015–2030.
14. Asri, S.; Pranggono, B. Impact of distributed denial-of-service attack on advanced metering infrastructure. *Wirel. Pers. Commun.* **2015**, *83*, 2211–2223.
15. Zhang, Y.; Wang, L.; Sun, W.; Green, R.C.; Alam, M. Distributed intrusion detection system in a multi-layer network architecture of smart grids. *IEEE Trans. Smart Grid* **2011**, *2*, 796–808.
16. Kurt, M.N.; Yilmaz, Y.; Wang, X. Real-time detection of hybrid and stealthy cyber-attacks in smart grid. *IEEE Trans. Inf. Forensics Secur.* **2019**, *14*, 498–513.
17. Saad, M.; Spaulding, J.; Njilla, L.; Kamhoua, C.; Shetty, S.; Nyang, D.; Mohaisen, D. Exploring the attack surface of blockchain: A comprehensive survey. *IEEE Commun. Surv. Tutor.* **2020**, *22*, 1977–2008.
18. Siegel, D. Understanding the DAO Attack. Available online: https://www.coindesk.com/understandingdao-hack-journalists/ (accessed on 13 September 2019).
19. Ross, S.; Pineau, J.; Chaib-draa, B.; Kreitmann, P. A Bayesian approach for learning and planning in partially observable Markov decision processes. *J. Mach. Learn. Res.* **2011**, *12*, 1729–1770.
20. Doshi-Velez, F.; Wingate, D.; Roy, N.; Tenenbaum, J. Nonparametric Bayesian policy priors for reinforcement learning. *Adv. Neural Inf. Process. Syst.* **2010**, *1*, 532–540.
21. Jaakkola, T.; Singh, S.P.; Jordan, M.I. Reinforcement learning algorithm for partially observable Markov decision problems. *Adv. Neural Inf. Process. Syst.* **1994**, *7*, 345–352.
22. Perkins, T.J. Reinforcement learning for POMDPs based on action values and stochastic optimization. In Proceedings of the Eighteenth National Conference on Artificial Intelligence and Fourteenth Conference on Innovative Applications of Artificial Intelligence, Edmonton, AB, Canada, 28 July 28–1 August 2002; pp. 199–204.
23. Loch, J.; Singh, S.P. Using eligibility traces to find the best memoryless policy in partially observable Markov decision processes. In Proceedings of the 15th International Conference on Machine Learning, (ICML), Madison, WI, USA, 24–27 July 1998; pp. 323–331.
24. Lanzi, P.L. Adaptive agents with reinforcement learning and internal memory. In Proceedings of the Sixth International Conference on the Simulation of Adaptive Behavior (SAB2000), Paris, France, 17–18 October 2005; MIT Press: Cambridge, MA, USA, 2000; pp. 333–342.
25. Peshkin, L.; Meuleau, N.; Kaelbling, L.P. Learning policies with external memory. In Proceedings of the 16th International Conference on Machine Learning, San Francisco, CA, USA, 27–30 June 1999; pp. 307–314.
26. Yli-Huumo, J.; Ko, D.; Choi, S.; Park, S.; Smolander, K. Where is current research on blockchain technology?—A systematic review. *PLoS ONE* **2016**, *11*, e0163477.
27. Yuan, Y.; Wang, F.-Y. Parallel blockchain: Concept, methods and issues. *Acta Autom. Sin.* **2017**, *43*, 1703–1712.
28. Eyal, I.; Sirer, E.G. Majority is not enough: Bitcoin mining is vulnerable. In Proceedings of the International Conference on Financial Cryptography and Data Security, Christ Church, Barbados, 3–7 March 2014; pp. 436–454.
29. Kiayias, A.; Russell, A.; David, B.; Oliynykov, R. Ouroboros: A Provably Secure Proof-of-Stake Blockchain Protocol. Cryptology ePrint Archive. *Tech. Rep. 2016/889*. 2016. Available online: https://eprint.iacr.org/2016/889 (accessed on 6 December 2022).
30. Aien, M.; Hajebrahimi, A.; Fotuhi-Firuzabad, M. A comprehensive review on uncertainty modeling techniques in power system studies. *Renew. Sustain. Energy Rev.* **2016**, *57*, 1077–1089.
31. Sheikh, M.; Aghaei, J.; Chabok, H.; Roustaei, M.; Niknam, T.; Kavousi-Fard, A.; Shafie-Khah, M.; Catalão, J.P.S. Synergies between transportation systems, energy hub and the grid in smart cities. *IEEE Trans. Intell. Transp. Syst.* **2021**, *23*, 7371–7385. [CrossRef]
32. Chen, X.; Qu, G.; Tang, Y.; Low, S.; Li, N. Reinforcement learning for selective key applications in power systems: Recent advances and future challenges. *IEEE Trans. Smart Grid.* **2022**, *13*, 2935–2958. [CrossRef]
33. Cao, D.; Hu, W.; Zhao, J.; Zhang, G.; Zhang, B.; Liu, Z.; Chen, Z.; Blaabjerg, F. Reinforcement learning and its applications in modern power and energy systems: A review. *J. Mod. Power Syst. Clean Energy* **2020**, *8*, 1029–1042. [CrossRef]
34. Ruan, G.; Zhong, H.; Zhang, G.; He, Y.; Wang, X.; Pu, T. Review of learning-assisted power system optimization. *CSEE J. Power Energy Syst.* **2020**, *2*, 221–231.
35. Mohamed, M.A.; Awwad, E.M.; El-Sherbeeny, A.M.; Nasr, E.A.; Ali, Z.M. Optimal scheduling of reconfigurable grids considering dynamic line rating constraint. *IET Gener. Transm. Distrib.* **2020**, *14*, 1862–1871. [CrossRef]
36. Al Alahmadi, A.A.; Belkhier, Y.; Ullah, N.; Abeida, H.; Soliman, M.S.; Khraisat YS, H.; Alharbi, Y.M. Hybrid wind/PV/battery energy management-based intelligent non-integer control for smart DC-microgrid of smart university. *IEEE Access* **2021**, *9*, 98948–98961.
37. Makhadmeh, S.N.; Khader, A.T.; Al-Betar, M.A.; Naim, S.; Abasi, A.K.; Alyasseri, Z.A.A. Optimization methods for power scheduling problems in smart home: Survey. *Renew. Sustain. Energy Rev.* **2019**, *115*, 109362.
38. Roustaei, M.; Letafat, A.; Sheikh, M.; Sadoughi, R.; Ardeshiri, M. A cost-effective voltage security constrained congestion management approach for transmission system operation improvement. *Electr. Power Syst. Res.* **2022**, *203*, 107674. [CrossRef]
39. Mohamed, M.A.; Hajjiah, A.; Alnowibet, K.A.; Alrasheedi, A.F.; Awwad, E.M.; Muyeen, S.M. A secured advanced management architecture in peer-to-peer energy trading for multi-microgrid in the stochastic environment. *IEEE Access* **2021**, *9*, 92083–92100. [CrossRef]
40. Mohamed, M.A.; Mirjalili, S.; Dampage, U.; Salmen, S.H.; Obaid, S.A.; Annuk, A. A cost-efficient-based cooperative allocation of mining devices and renewable resources enhancing blockchain architecture. *Sustainability* **2021**, *13*, 10382. [CrossRef]

41. Zou, H.; Tao, J.; Elsayed, S.K.; Elattar, E.E.; Almalaq, A.; Mohamed, M.A. Stochastic multi-carrier energy management in the smart islands using reinforcement learning and unscented transform. *Int. J. Electr. Power Energy Syst.* **2021**, *130*, 106988.
42. Min, L.; Alnowibet, K.A.; Alrasheedi, A.F.; Moazzen, F.; Awwad, E.M.; Mohamed, M.A. A stochastic machine learning based approach for observability enhancement of automated smart grids. *Sustain. Cities Soc.* **2021**, *72*, 103071.
43. Chabok, H.; Aghaei, J.; Sheikh, M.; Roustaei, M.; Zare, M.; Niknam, T.; Lehtonen, M.; Shafi-khah, M.; Catalão, J.P.S. Transmission-constrained optimal allocation of price-maker wind-storage units in electricity markets. *Appl. Energy* **2022**, *310*, 118542.