Deep Reinforcement Learning Based Mimicry Defense System for IoT Message Transmission

Zhihao Wang, Dingde Jiang(✉), Jianguang Chen, and Wei Yang

University of Electronic Science and Technology of China, Chengdu 611731, China

Abstract. With the development of 5G and Internet of Everything, IoT has become an essential network infrastructure. The connection between massive devices brings huge convenience and effectiveness, also introducing more security threats and vulnerabilities that compromise the security, privacy and trust problem of the IoT data, devices and users or service providers. Traditional security approaches are mostly based on the analysis of attack characteristics, seeking vulnerabilities, or patching systems. Independent from prior knowledge or specific defense method, the mimic defense can realize a built-in security system through heterogeneity, redundancy, and dynamic. In this paper, to address the security problem of the IoT communication protocol MQTT, a DRL-based mimicry defense system for IoT message transmission is proposed. We conduct mimic transformation on the MQTT broker, with functionally equivalent but structural dissimilar variants. To refine the determining accuracy of basic mimic ruling mechanism, namely majority voting, an intelligent ruling mechanism based on deep Q network is proposed. Finally, the simulation results demonstrate the security and effectiveness of the proposed scheme.

Keywords: Mimicry defense · Deep reinforcement learning · IoT

1 Introduction

With the growth of Internet technology, the capability and capacity of network could accommodate massive devices to assess the network. Internet of Everything is also a typical characteristic of 5G. One of the essential infrastructures 5G is the Internet of Things (IoT). The connection between massive devices brings convenience and effectiveness. Every household electrical appliance, the wearable device can be connected by IoT, collecting data for further advanced data analysis and functionalities. High connectivity and massive devices also introduce more vulnerabilities, which severely compromise the security and privacy of the IoT users or service providers [1], especially the M2M (Machine to Machine) message transmission protocol, the MQTT (Message Queuing Telemetry Transport) protocol. The security of MQTT plays a crucial role in the security of IoT architecture, which faces various threats including replay attacks, MITM (Man-in-the-Middle Attack) [2], data confidentiality threats, authentication threats [3], etc. Therefore, many endeavors are paid to study novel approaches enhancing the security of

D. Jiang and H. Song (Eds.): SIMUtools 2021, LNICST 424, pp. 421–431, 2022.
https://doi.org/10.1007/978-3-030-97124-3_31

the MQTT protocol. A novel MQTT communication structure based on broker bridging is proposed to enhance the overall security of IoT system, along with the secure authentication and authorization scheme [4]. Liao et al. propose an improved attribute-based encryption scheme for MQTT, combined with chaos synchronization, which enhances the security of resource-constrained IoT devices [5].

With the extensive application of AI (Artificial Intelligence), many researchers employ machine learning techniques to address network security problems [6], including random forest, support vector machine, deep neural network, etc. Deep reinforcement learning integrates deep learning and reinforcement learning, which could learn from the interaction with the environment. The reward and punishment to the agent make it behave more like a human. Remarkable achievements in network security fields have been realized using DRL approaches [7]. Jiadai et al. propose an attack-tolerance scheme in Internet of Vehicles utilizing the DRL to defense the topology poisoning attack and enable the self-recovery capability of vehicular edge network [8]. Giovanni et al. propose a DRL-based framework for botnet detectors against the adversarial attack, which also could prevent several unforeseen evasion attacks [9].

Traditional network defense techniques are mostly passive, developing targeted defense approaches based on the characteristics of specific attacks. If the network attack does not have specificity or is an unknown type, the traditional defense techniques will lose effectiveness. Besides, single-system architecture is easily affected by the single point of failure problem. Therefore, to secure the system from built-in or the attacked system perspective, active cyber defense approaches are proposed. The active cyber defense is to launch early warning before the attack is implemented, using big data analysis and AI technology. Typical active cyber defense approaches include data encryption, access control, intrusion detection [10], moving target defense [11], etc. The mimic defense is another type of active cyber defense approach, which is realized through DHR (Dynamic Heterogeneous Redundancy) structure, to address the unknown threats caused by the vulnerability, backdoor with uncharted characteristics [12]. Our previous work also focuses on the mimic transformation of smart grid system [13]. Several variants with distinct internal structures but with identical external functionalities are equipped in the mimic defense system to increase the uncertainty of internal structure [14]. The mimic defense system provides external uniform interfaces for system users, which makes the internal structure invisible to them. It has been demonstrated that the cost of attacking all variants in a short time is tremendous and scarcely possible, due to the various structure and implementation of the variants. Besides, the peculiar mimic ruling mechanism is capable to sense the abnormal variants with inconsistent output vectors, to further detect the non-cooperative attacks [15]. After detecting the abnormal variants, the negative feedback mechanism will force the abnormal variants to shut down and reconstruct. Backup variants will succeed the operation of invalid variants, with state synchronization. Based on the above analysis, the mimic defense system has attack tolerance and fault tolerance capability, which means even part of variants art attacked, the system is still secure and credible [16]. Besides, the uncertainty of the internal structure of mimic system can endow it with invisibility characteristic, which effectively confuses the attackers. Our previous work includes routing and distribution [17, 18, 21], security and networking [19, 20].

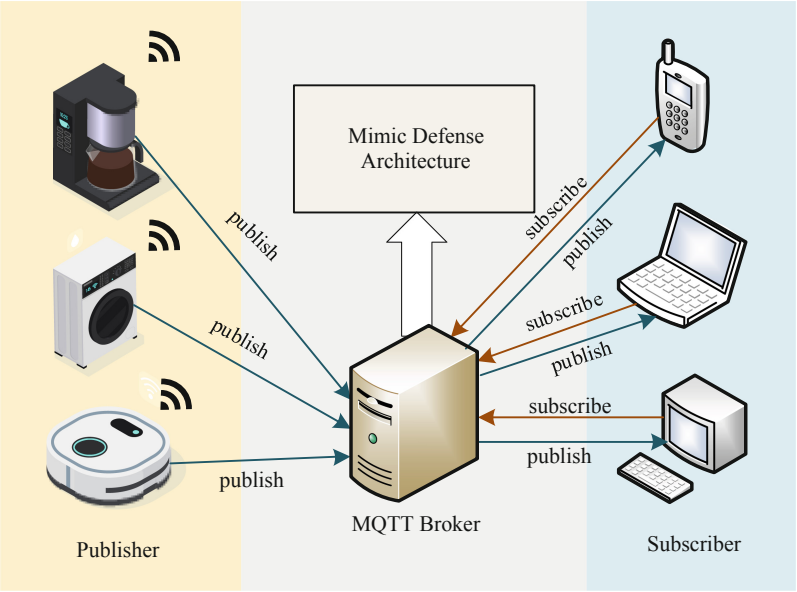

Fig. 1. MQTT structure

In this paper, to address the security problems of IoT system, a DRL-based mimic defense system for IoT message transmission is proposed to address the vulnerability of the single static architecture. By introducing the dynamic, redundancy, and heterogeneity characteristics into traditional IoT message transmission system, namely the MQTT protocol, the communication broker can maintain normal operation even there are network attacks. The mimic ruling mechanism improves the security and stability by validating the credibility of each variant and restructuring the problematic variants. Besides, to improve the effectiveness and accuracy of the ruling mechanism, the DRL approach is employed to realize intelligent mimic ruling, further improving the reliability of the system. A Deep Q Network (DQN) model is utilized to assess the credibility of the message from the variant. Finally, the simulation results demonstrate the security enhancement brought to the IoT communication broker.

2 System Architecture

2.1 MQTT Protocol

MQTT protocol is the most widely used communication protocol of IoT, a lightweight protocol based on the publish-subscribe model. MQTT works on the TCP/IP protocol, thereby it is feasible to extract the packet header to identify the packet type, which is employed in this paper to distinguish the correctness. It requires little code memory space, and a small amount of bandwidth, which makes it suitable for communication in resource-limited, low-bandwidth, high-latency, and unreliable networks. Besides, the QoS support of MQTT is flexible and applicable to a variety of scenarios. When the QoS

field in MQTT header is 0, the message will be sent once at most, which may cause the message missing. When the QoS field in MQTT header is 1, the message will be sent at least once to ensure arriving at the server, which may lead to duplicate. When the QoS field in MQTT header is 2, it will ensure the transmission will be conducted only once to guarantee that duplicate messages are not delivered to destination. There are three kinds of roles in MQTT, namely publisher, broker and subscriber. The publisher and subscriber are client, but the broker is realized on server, denoted by Fig. 1. Therefore, regarding the limited resource of the clients and the computing capability and space requirements of mimic structure, we implement the mimic on broker, realizing heterogeneous, redundant and dynamic variants.

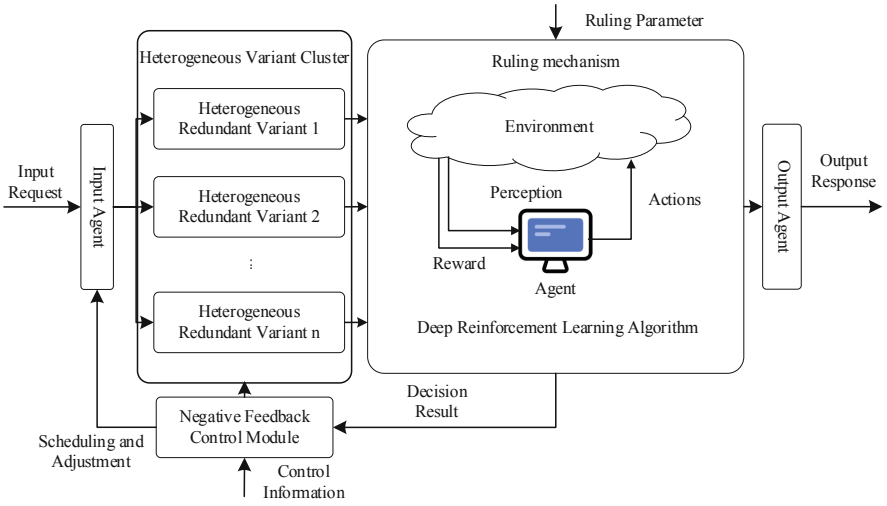

Fig. 2. Overall structure of mimicry defense system

2.2 Mimic Defense IoT System

To enhance the security of IoT system, based on the structure of MQTT communication protocol, a mimic defense architecture is equipped to the IoT system to realize a mimic MQTT broker. The overall structure of the mimic defense IoT system is shown in Fig. 2. Main components include input agent, heterogeneous variant cluster, ruling mechanism, negative feedback and output agent. First, once there are messages from the publisher, the input agent will reproduce the message to all variants that are functionally equivalent and have a uniform input interface. All the variants will process the message individually, extracting the packet header and matching the destination. After the variants yield outputs to the ruling mechanism, we utilize the DRL algorithm to determine the credibility of each message, through the decision-making of DRL model. A most credible message will be selected and transmitted to the output agent, which will further deliver it to corresponding subscribers. Meanwhile, the ruling result will also be duplicated and sent

to the negative feedback control module, where the variant with inconsistent output vector will be located and killed. Before the abnormal variant offline, a new variant will be constructed and the corresponding state will be synchronized to it, which will replace the running problematic variant. After the problematic variant is replaced, cleaning and reconstruction will be conducted on it to prevent further damage to the system. Attacks with unknown features, random faults, vulnerabilities can be normalized into an inconsistency output vector problem, which can be handled by the cleaning and reconstruction mechanism. All the data flow or information flow in the mimic IoT system is unidirectional, which significantly avoids network intrusion.

3 DRL Ruling Mechanism

3.1 Deep Reinforcement Learning

In this paper, the DRL model we employed is Deep Q Network (DQN), a typical value-based DRL method. DQN combines the basic Q-learning and deep learning, fitting the value function with DNN. Besides, the experience replay mechanism is employed, to sample the history record for training the Q network. Target value mechanism is also utilized to train Q network. In basic Q-function, the action taken by the agent is always fixed, causing the exploration-exploitation dilemma. The agent will continuously choose the specific action with good reward rather than exploring the action space. Hence, we introduce the Epsilon Greedy (ε-greedy) strategy to force the agent to explore unknown actions with a small probability, denoted as (1):

$$a = \begin{cases} \arg\max_a Q(s, a), \textit{with probability } 1 - \varepsilon \\ \textit{random}, \quad \textit{otherwise} \end{cases} \tag{1}$$

where Q means the Q-function, and s is the current state, and a represents the action. In DQN, two Q network is exploited, namely the Q and $\hat{Q}$, which are the same at first. In each training episode, the agent interacts with the environment through obtaining a state s and selecting action (ε-greedy) a. Then the environment will feedback a reward r to agent along with a new state $s^{'}$. Hence, a quadruple is obtained, as $(s, a, r, s^{'})$, which will be further stored in the experience replay buffer. To train the Q and $\hat{Q}$, a batch of records from experience replay buffer will be extracted to calculate the target value with $\hat{Q}$, defined as (2):

$$y = r + \max_a \hat{Q}(s^{'}, a) \tag{2}$$

where action a is to maximize the $\hat{Q}$. Then we will update parameters in Q to make $Q(s, a)$ approach to target y as much as possible. Besides, an interval is set to synchronize $\hat{Q}$ with Q. The parameter updating method of Q network is the same as DNN. After a certain degree of exploration and training, the agent can take reasonable actions according to the given state of the environment.

Table 1. DRL-based Mimic Ruling Mechanism

Algorithm 1. DRL-based Mimic Ruling Mechanism

Input: Training Message Dataset Msg_T, Real Message group Msg_R.
Output: Mimic Ruling Result.

Conduct Min-Max normalization on $Msg_T.X$
Train the DQN:
  for each message m in Msg_T:
    Input $m.X$ to DQN as a state s
    DQN take action a
    if $a = m.y$ **then** reward $r = 1$
    else $r = -1$
    end if
    Store the record and update network parameter
  end for
Employ trained DQN for ruling:
  Initialize result list R
  for each message g in Msg_R
    Normalize $g.X$ with parameter in step 1
    Input $g.X$ to DQN as a state s
    DQN take action a
    $R = R \cup a$
  end for
  return $Msg_R[\arg\max(R)]$

3.2 DRL-Based Ruling Mechanism

The traditional mimic ruling mechanism is based on the majority voting, which could adapt to most scenarios. However, once the attackers have strong enough attack intensity and time to manipulate more than half of the variants in mimic system, the voting ruling will be invalid and untrustworthy. Therefore, to refine the accuracy of the ruling mechanism and further enhance the system security, a DRL-based mimic ruling mechanism is proposed in this paper. The decision-making ability is employed to distinguish the normal message from the outputs of heterogeneous variants, through analogizing the variants as the agent. First, a set of training message samples is utilized to train the DRL model, where the state is a feature of messages, and the action is to determine the correctness of the message. Once the agent takes action, the verification will be conducted to determine feedback, namely reward or punishment. The DRL model will update the parameters of deep neural network to take more correct decisions gradually. After training, the DRL model is employed to realize mimic ruling, verifying the correctness of the given messages. The messages judged to be normal will be taken as the ruling result. For faster training and convergence of deep neural network, the sample message data should be preprocessed, dimension removing and data normalization, which we employed in this paper is the Min-Max Normalization. Each message in the training message dataset contains the features and label, denoted as X and y. Overall procedure of the DRL-based

mimic ruling mechanism is depicted in Table 1, where *DQN* indicates the employed deep Q network model.

4 Simulation Result

4.1 Experiment Setup

In this section, several experiments are conducted to verify the security enhancement brought to single-structure IoT communication broker and the performance of the proposed DRL-based ruling mechanism. First, we establish a simple MQTT simulation scenario, including message subscribing and publishing, to generate a validation dataset. Main features of the obtained MQTT messages are consisted of ControlPacketType, Flags, RemainingLength, PacketIdentifierMSB, PacketIdentifierLSB, PayloadLength. And an artificial-annotated label is utilized to represent the validity of the message. Then the message dataset will be processed by a single MQTT broker that takes the original message as output, a three-variant mimic broker with voting ruling mechanism, and the proposed DRL-based mimic broker with DRL ruling algorithm. We exploit the evaluation metrics in the machine learning classification task to validate the performance of three systems in environments with different attack intensities, namely the precision, recall, f1-score and accuracy.

4.2 Security Enhancement

To validate the security enhancement of the proposed DRL-based mimic IoT message transmission system, we first compare the performance with a certain attack intensity, which means every variant in the system has the same probability of failure or attack. The principal goal of MQTT broker is to deliver the messages safely. Hence, we input the messages into the systems, and validate the output of the broker. The judgment principle is given a normal message, to compare the correctness of the output message. The comparison result is shown in Fig. 3. It is obvious that the recall value of abnormal message of all three approach is 1, which means a very extreme situation that all three systems are attacked. The precision value of a normal message represents a totally secure system with no variants attacked. Apart from the above two useless metrics, the rest metrics all demonstrate that the proposed DRL-based mimic broker exceeds the single broker and basic mimic broker. For the abnormal message, the precision indicates that the DRL-Mimic broker could prevent more normal messages from being misjudged as abnormal messages with part of variants attacked. The recall value of a normal message represents the capability of extracting normal message from all outputs of the variants, measuring the security performance. For single structure broker, once the broker is attacked, the output message will be manipulated easily. The basic mimic structure broker could distinguish normal output from three variants to some extent, but it will misjudge when there are more than half of the variants with abnormal output. The DRL-based mimic broker could effectively improve this problem by selecting the most credible variant and utilizing its output. The f1-score is a comprehensive measurement index of precision and recall, which also demonstrates the superiority of the proposed

scheme. Therefore, from the above analysis, compared with single-structure MQTT broker and basic mimic broker, the proposed DRL-based mimic defense IoT message transmission system has obvious security enhancement in the experiment scenario.

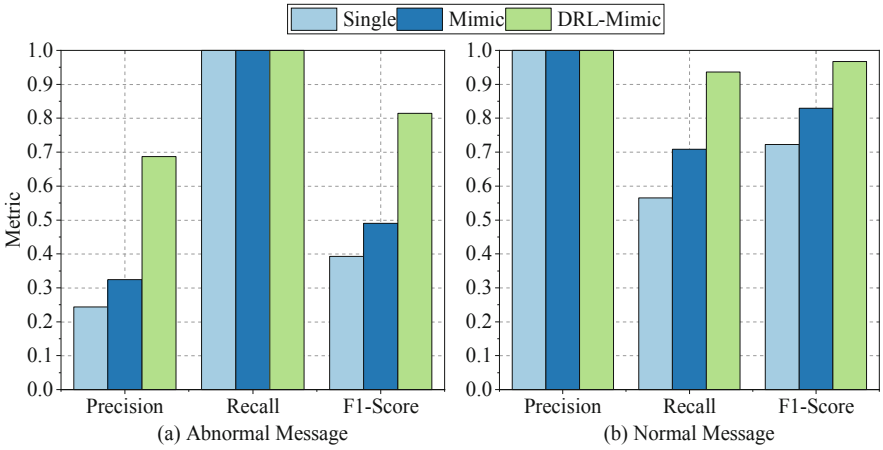

Fig. 3. Security enhancement validation.

4.3 Performance Analysis

To estimate the performance of three MQTT structures, we conduct the message transmission experiments with different attack intensity, which means every broker instance or variant is attacked with a certain probability and the transmitted message has a certain probability to be manipulated. Obviously, once the broker is attacked, the message in the single-structure MQTT broker could be easily manipulated. However, due to the redundancy of variant and the ruling mechanism, the mimic system has attack tolerance. Even part of variants is attacked, the mimic system still can function properly and self-recover. We take the accuracy and the recall value of normal message to validate the performance of three systems, the result of which is shown in Fig. 4. In (a), the accuracy metric is exploited to verify the effectiveness of the proposed scheme. It is obvious that the DRL-Mimic outperforms the other two approaches under all attack intensity conditions, which could maintain a relatively stable accuracy. The fluctuating of the other two schemes is because the calculating rule of accuracy is to validate whether the broker could output original results from publishers, the IoT devices. Therefore, before the attack intensity reaches 0.5 or 0.6, the ruling accuracy of single and mimic scheme drops at a nearly linear speed, and rises rapidly afterward. Because, once the attack intensity is higher than 0.5, the abnormal message takes the dominant position and accuracy will increase then. Different from the accuracy, the recall of normal message can truly reflect the security performance of the MQTT broker with different attack intensities, expressed as Fig. 4 (b). At the initial stage with no attack, all the systems are secure. With the increment of attack intensity, the recall value, indicating the ruling security performance

of the broker, is getting lower gradually. The Single and Mimic system drop almost linearly. However, the descending speed or trend of DRL-Mimic system is significantly slower than the other two, keeping above 90%, which means even the attack intensity is huge, the proposed DRL-Mimic system could effectively determine the normal variants. The intelligent ruling mechanism address the simplicity and inaccuracy problem of the traditional majority voting approach. When the attack intensity comes to 1.0, all the recall values drop to zero, the reason of which is that each instance will certainly be attacked and there is no normal message that can be output. Even the mimic defense structure could improve the security of MQTT broker to some extent, the protection ability will decrease rapidly as the attack intensity increases. In summary, compared to the single-structure MQTT broker and basic mimic defense structure, the proposed DRL-based mimic defense IoT communication scheme can effectively improve the security of message transmission, with stability under different attack intensity.

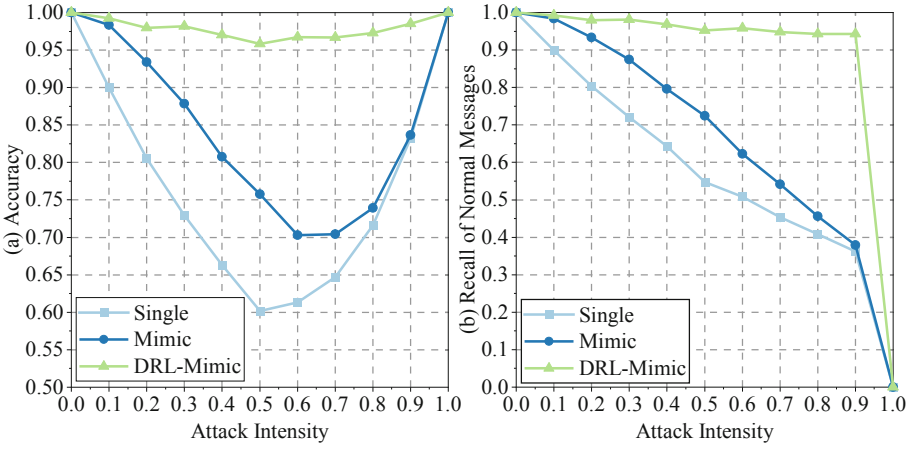

Fig. 4. Security improvement of mimic system with different algorithms.

5 Conclusion

As an essential infrastructure of 5G and future network, IoT is an important way to realize the Internet of Everything. MQTT is the most widespread transmission protocol in IoT, which is lightweight and space-saving. However, the massive devices and huge transmission operations bring tremendous vulnerabilities. To enhance the security of the IoT message transmission protocol, MQTT, a DRL-based mimic defense system is proposed in this paper. We employ equivalent but structural dissimilar variants to handle the same transmission task. The output of the variants will be distinguished and determined by the proposed DRL ruling mechanism, where the variant with inconsistent output will be reconstructed. The uncertain and dynamic internal structure will effectively mislead the attackers and enhance the security. The simulation results also reveal the security

improvement brought to the MQTT broker and the reliability and effectiveness of the proposed scheme.

Acknowledgements. This work was supported in part by the National Natural Science Foundation of China (No. 61571104), the Sichuan Science and Technology Program (No. 2018JY0539), the Key projects of the Sichuan Provincial Education Department (No. 18ZA0219), the Fundamental Research Funds for the Central Universities (No. ZYGX2017KYQD170), the CERNET Innovation Project (No. NGII20190111), the Fund Projects (Nos. 2020-JCJQ-ZD-016-11, 61403110405, 315075802, JZX6Y202001010161), and the Innovation Funding (No. 2018510007000134). The authors wish to thank the reviewers for their helpful comments.

References

1. Butun, I., Österberg, P., Song, H.: Security of the internet of things: vulnerabilities, attacks, and countermeasures. IEEE Commun. Surv. Tutor. **22**, 616–644 (2020)
2. Chen, F., Huo, Y., Zhu, J., Fan, D.: A review on the study on MQTT security challenge. In: 2020 IEEE International Conference on Smart Cloud (SmartCloud), pp. 128–133 (2020)
3. Swamy, S.N., Jadhav, D., Kulkarni, N.: Security threats in the application layer in IoT applications. In: 2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), pp. 477–480 (2017)
4. Amoretti, M., Pecori, R., Protskaya, Y., Veltri, L., Zanichelli, F.: A scalable and secure publish/subscribe-based framework for industrial IoT. IEEE Trans. Ind. Inf. **17**, 3815–3825 (2021)
5. Liao, T.-L., Lin, H.-R., Wan, P.-Y., Yan, J.-J.: Improved attribute-based encryption using chaos synchronization and its application to MQTT security. Appl. Sci. **9**, 4454 (2019)
6. Chaabouni, N., Mosbah, M., Zemmari, A., Sauvignac, C., Faruki, P.: Network intrusion detection for IoT security based on learning techniques. IEEE Commun. Surv. Tutor. **21**, 2671–2701 (2019)
7. Luong, N.C., et al.: Applications of deep reinforcement learning in communications and networking: a survey. IEEE Commun. Surv. Tutor. **21**, 3133–3174 (2019)
8. Wang, J., Tan, Y., Liu, J., Zhang, Y.: Topology poisoning attack in SDN-enabled vehicular edge network. IEEE Internet Things J. **7**, 9563–9574 (2020)
9. Apruzzese, G., Andreolini, M., Marchetti, M., Venturi, A., Colajanni, M.: Deep reinforcement adversarial learning against botnet evasion attacks. IEEE Trans. Netw. Serv. Manag. **17**, 1975–1987 (2020)
10. Wang, Z., Jiang, D., Huo, L., Yang, W.: An efficient network intrusion detection approach based on deep learning. Wirel. Netw. (2021)
11. Wang, S., Shi, H., Hu, Q., Lin, B., Cheng, X.: Moving target defense for internet of things based on the zero-determinant theory. IEEE Internet Things J. **7**, 661–668 (2020)
12. Hu, H., Wu, J., Wang, Z., Cheng, G.: Mimic defense: a designed-in cybersecurity defense framework. IET Inf. Secur. **12**, 226–237 (2017)
13. Wang, Z., Jiang, D., Wang, F., Lv, Z., Nowak, R.: A polymorphic heterogeneous security architecture for edge-enabled smart grids. Sustain. Cities Soc. **67**, 102661 (2021)
14. Li, G., Wang, W., Gai, K., Tang, Y., Yang, B., Si, X.: A framework for mimic defense system in cyberspace. J. Signal Process. Syst. **93**, 169–185 (2021)
15. Wu, J.: Cyberspace Mimic Defense. Springer, Heidelberg (2020)
16. Wang, Y.-W., Wu, J.-X., Guo, Y.-F., Hu, H.-C., Liu, W.-Y., Cheng, G.-Z.: Scientific workflow execution system based on mimic defense in the cloud environment. Front. Inf. Technol. Electron. Eng. **19**(12), 1522–1536 (2018). https://doi.org/10.1631/FITEE.1800621

17. Jiang, D., et al.: AI-assisted energy-efficient and intelligent routing for reconfigurable wireless networks. IEEE Trans. Netw. Sci. Eng. (2021). https://doi.org/10.1109/TNSE.2021.3075428
18. Jiang, D., et al.: QoE-aware efficient content distribution scheme for satellite-terrestrial networks. IEEE Trans. Mob. Comput. (2021). https://doi.org/10.1109/TMC.2021.3074917
19. Wang, Z., et al.: A polymorphic heterogeneous security architecture for edge-enabled smart grid. Sustain. Cities Soc. (2020)
20. Jiang, D., et al.: A performance measurement and analysis method for software-defined networking of IoV. IEEE Trans. Intell. Transp. Syst. (2020). https://doi.org/10.1109/TITS.2020.3029076
21. Jiang, D., et al.: Energy-efficient heterogeneous networking for electric vehicles networks in smart future cities. IEEE Trans. Intell. Transp. Syst. (2020). https://doi.org/10.1109/TITS.2020.3029015