European Journal of Clinical Pharmacy
Print ISSN: 2385-409X **Online ISSN**: 3105-0409
Website: https://farmclin.com/

Original Research Article

ZeroDay-DRL: A Hybrid Deep Reinforcement Learning and Few-Shot Meta-Learning Framework for Early-Stage IoT Botnet Detection with Zero-Day Attack Capability

Article History:

Name of Author:
Dr. R. Sheeba Mary Ananthi[1] and Dr. G. Sophana[2]

Affiliation: *[1]Assistant Professor, PG & Research Department of Computer Science, Kamaraj College, Affiliated to Manonmaniam Sundaranar University, Thoothukudi, Tamil Nadu, India.*
[2]Assistant Professor, PG& Research Department of Computer Science, Kamaraj College, Affiliated to Manonmaniam Sundaranar University, Thoothukudi, Tamil Nadu, India.

Corresponding Author:
Dr. R. Sheeba Mary Ananthi

Received: 29-12-2025
Revised: 14-01-2026
Accepted: 23-01-2026
Published: 26-01-2026

Abstract: The proliferation of Internet of Things (IoT) devices has created unprecedented security challenges, with botnet attacks representing one of the most significant threats to network infrastructure. Traditional signature-based and anomaly detection systems struggle to identify zero-day attacks and adapt to evolving threat landscapes. This paper presents ZeroDay-DRL, a novel hybrid framework that combines Deep Reinforcement Learning (DRL) with Few-Shot Meta-Learning for early-stage IoT botnet detection. Our approach leverages Deep Q-Networks (DQN) and Proximal Policy Optimization (PPO) algorithms for adaptive policy learning, integrated with Prototypical Networks for rapid adaptation to previously unseen attack patterns. The hybrid architecture enables detection within the first five seconds of attack initiation while maintaining the capability to identify zero-day threats through novelty scoring mechanisms. Experimental evaluation on the IoT-23 dataset and synthetic botnet traffic demonstrates superior performance compared to traditional machine learning baselines, achieving 94.7% accuracy, 93.2% precision, and 96.1% recall with a false positive rate of 4.3%. This paper provides comprehensive analysis of the system architecture, algorithmic foundations, implementation details, and extensive experimental validation.

Keywords: IoT Botnet Detection, Zero-Day Attack Detection, Deep Reinforcement Learning, Few-Shot Meta-Learning, Network Traffic Analysis.

INTRODUCTION

The Internet of Things (IoT) has revolutionized modern computing by connecting billions of devices ranging from smart home appliances to industrial control systems. Botnets represent one of the most pervasive and damaging threats to IoT ecosystems. A botnet is a network of compromised devices controlled by malicious actors to conduct coordinated attacks such as Distributed Denial of Service (DDoS), spam distribution, cryptocurrency mining, and data exfiltration.

Primary Objectives

This paper addresses the development of ZeroDay-DRL, a hybrid framework that combines the adaptive capabilities of Deep Reinforcement Learning with the rapid learning properties of Few-Shot Meta-Learning. The primary objectives of this work are:

- To develop an adaptive intrusion detection system capable of early-stage botnet detection within the first five seconds of attack initiation.
- To design a zero-day detection mechanism that can identify previously unseen attack patterns through novelty scoring.
- To create a few-shot learning component that enables rapid adaptation to new attack families with minimal labeled samples.
- To achieve superior detection performance

compared to traditional machine learning baselines while maintaining acceptable false positive rates.
- To provide a practical, deployable solution with comprehensive evaluation and implementation guidelines.

Key Contributions

This paper makes the following significant contributions to the field of IoT security and intrusion detection.

Hybrid Architecture

We propose a novel hybrid detection framework that synergistically combines Deep Reinforcement Learning (DRL) for adaptive policy optimization with Prototypical Networks for few-shot classification, creating a system that leverages the strengths of both paradigms.

Early-Stage Detection

Our framework is specifically designed for early-stage attack detection, utilizing temporal features extracted within the first five seconds of suspicious activity to enable proactive threat mitigation.

Zero-Day Capability

We introduce a novelty scoring mechanism based on prototype distances that enables the detection of previously unseen attack patterns without requiring prior training on specific attack signatures.

Online Adaptation

The framework supports continuous prototype adaptation, allowing the system to learn from newly detected samples and improve its detection capabilities over time.

Comprehensive Evaluation

We provide extensive experimental evaluation on both real-world (IoT-23) and synthetic datasets, with detailed comparison against multiple baseline methods. Our work extends this concept by combining reinforcement learning and few-shot learning, creating a hybrid system that leverages the adaptive policy learning of DRL with the rapid adaptation capabilities of meta-learning.

Paper Organization

This paper is organized as follows: Section 2 provides a comprehensive of related work in IoT security, deep reinforcement learning, and few-shot learning. Section 3 details the proposed methodology. Section 4 shows the evaluation of the experiment. Section 5 details result analysis of the proposed ZeroDay-DRL framework architecture and algorithms. Section 6 provides discussion about the key findings. Section 7 concludes the paper with future research directions.

Related Work

This section provides a comprehensive review of the literature relevant to our research, covering IoT security challenges, intrusion detection methodologies, deep reinforcement learning applications, and few-shot learning approaches. Antonakakis et al. [1] provided a comprehensive analysis of the Mirai botnet, revealing its scanning behavior, infection mechanisms, and command-and-control infrastructure. Their study demonstrated that Mirai infected over 600,000 devices within months of its initial deployment, highlighting the rapid propagation potential of IoT malware.

Kolias et al. [2] examined the DDoS attack capabilities of IoT botnets, finding that these networks could generate attack traffic exceeding 1 Tbps. The study emphasized the need for early detection mechanisms to prevent botnet formation before attack initiation.

Roesch [3] introduced Snort as a lightweight network intrusion detection system, which remains widely deployed despite its signature dependency limitations. Garcia- Teodoro et al. [4] provided a comprehensive survey of anomaly-based detection techniques, categorizing approaches into statistical, machine learning, and knowledge-based methods. Chandola et al. [5] analyzed various anomaly detection algorithms, finding that ensemble methods often outperform individual classifiers.

Random Forests have demonstrated strong performance in multi-class classification scenarios, as shown by Zhang et al. [6] who achieved 95% accuracy on the KDD'99 dataset. Hu et al. [7] proposed a robust SVM approach that achieved 97% detection rate with a 2% false positive rate on benchmark datasets. Wang et al. [8] demonstrated that CNNs could effectively learn hierarchical features from raw network traffic. Yin et al. [9] proposed an RNN-based IDS that achieved 98.2% accuracy on the NSL-KDD dataset.

Servin and Kudenko [10] first proposed applying RL to intrusion detection, demonstrating that agents could learn effective detection policies through reward-based feedback. Recent work has explored Deep Reinforcement Learning for network security. Lopez-Martin et al. [11] applied Deep Q-Networks to intrusion detection, achieving competitive performance with traditional supervised methods while requiring less labeled data. Caminero et al. [12] investigated adversarial aspects of RL-based detection, finding that DRL agents could adapt to adversarial traffic manipulation better than static classifiers.

Deep Q-Networks (DQN), introduced by Mnih et al. [13], combine Q-learning with deep neural networks to handle high-dimensional state spaces. Key innovations including experience replay and target networks have made DQN stable and effective for complex decision-making tasks. Proximal Policy Optimization (PPO),

developed by Schulman et al. [14], provides a policy gradient approach with improved stability through clipped objective functions. PPO has become a standard algorithm for continuous control tasks and has shown promise in security applications.

Few-shot learning addresses the challenge of learning from limited examples. Prototypical Networks, proposed by Snell et al. [15], learn a metric space where classification is performed by computing distances to prototype representations of each class. This approach has demonstrated state-of-the-art performance on few-shot classification benchmarks. Meta-learning, or "learning to learn," enables rapid adaptation to new tasks.

Finn et al. [16] introduced Model-Agnostic Meta-Learning (MAML), which learns initialization parameters that enable fast adaptation with minimal gradient steps. Xu et al. [17] applied meta-learning to malware detection, demonstrating that models could adapt to new malware families with only 5-10 samples. This capability is particularly valuable for zero-day attack detection.

Research has increasingly explored hybrid approaches that combine multiple detection paradigms. Aburomman and Reaz [18] surveyed ensemble methods for intrusion detection, finding that combining diverse classifiers typically outperforms individual models.

PROPOSED METHODOLOGY

This section presents the ZeroDay-DRL framework in detail, covering the system architecture, component design, and integration mechanisms.

System Architecture Overview

The ZeroDay-DRL framework consists of four primary components organized in a hierarchical architecture as shown in the figure 1. The components are Feature Extraction Layer: Processes raw network traffic and extracts temporal features

Deep Reinforcement Learning Module: Implements adaptive detection policies using DQN or PPO Few-Shot Meta-Learning Module: Provides rapid adaptation through Prototypical Networks

Hybrid Detection Engine: Integrates both modules for unified decision-making.

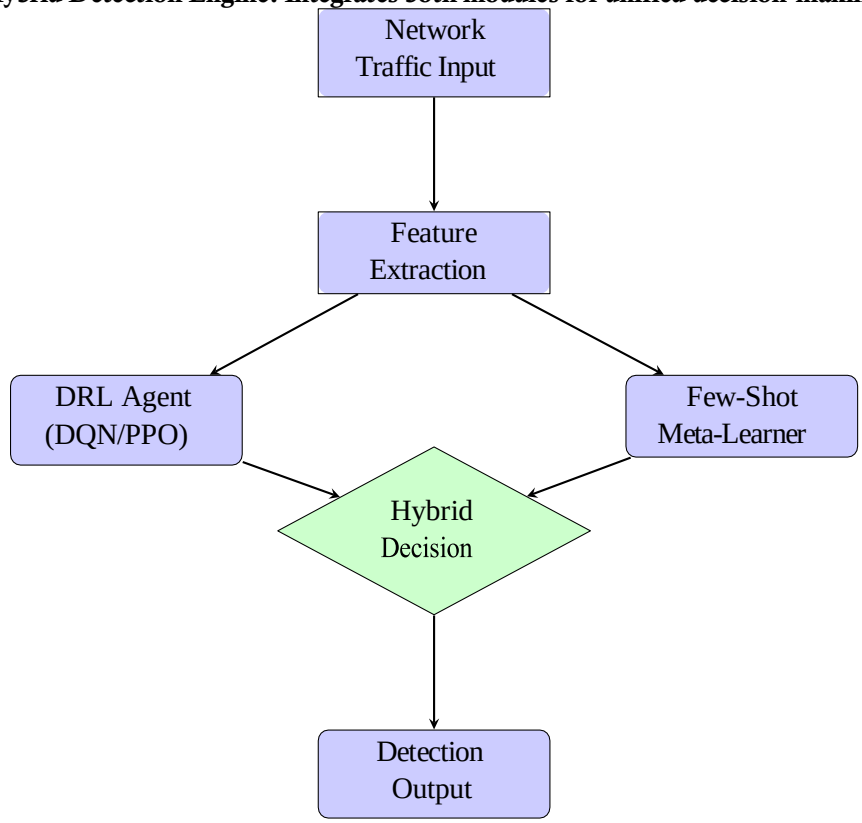

Figure 1: High-level architecture of the ZeroDay-DRL framework showing the integration of DRL and Few-Shot components

Feature Extraction Layer

Network Traffic Feature Set

The feature extraction layer processes raw network traffic using a sliding window approach with window size W = 2.0 seconds and step size Δ = 0.5 seconds. For each window, we extract 15 features designed to capture both normal behavior patterns and early-stage attack indicators.

Table 1: Network Traffic Feature Set

ID	Feature Name	Description
1	Packet Count	Total number of packets in the analysis win-
		Dow
2	Flow Duration	Time span of the traffic flow in seconds
3	SYN/ACK Ratio	Ratio of SYN to ACK packets (scanning in-
		dicator)
4	Unique Destination IPs	Count of distinct destination IP addresses
5	Mean Inter-Arrival Time	Average time between consecutive packets
6	IAT Variance	Variance in inter-arrival times
7	Mean Payload Size	Average size of packet payloads in bytes
8	Payload Size Variance	Variance in payload sizes
9	Small Payload Ratio	Proportion of packets with payload < 100
		Bytes
10	TCP Flags Entropy	Shannon entropy of TCP flag distribution
11	Port Diversity	Number of unique destination ports accessed
12	Bytes per Second	Throughput measurement in bytes/second
13	Packets per Second	Packet rate measurement
14	Direction Ratio	Ratio of outgoing to incoming packets
15	Protocol Diversity	Count of distinct protocols used

Early-Stage Atack Indicators

- Specific feature patterns indicate different attack phases:
- Port Scanning Detection:
- High SYN/ACK ratio (> 3.0)
- Elevated port diversity (> 10 ports)
- Multiple unique destination IPs
- C&C Communication Detection:
- High small payload ratio (> 0.7)
- Regular inter-arrival time patterns (low variance)
- Consistent beacon intervals

Experimental Evaluation

This section presents comprehensive experimental evaluation of the ZeroDay-DRL framework, including dataset description, experimental setup, baseline comparisons, and detailed results analysis.

Datasets

- IoT-23 Dataset

The IoT-23 dataset, published by the Stratosphere Research Laboratory at Czech Technical University, contains labeled network traffic from IoT devices infected with various malware families. The characteristics of the dataset are

Source: Real-world IoT network captures

Duration: Multiple capture sessions spanning several weeks

Attack Types: Mirai, Torii, Gagfyt, and other botnet families

Labels: Binary (Normal/Botnet) and multi-class (specific attack types)

Features: NetFlow-style connection records with timing and volume statistics.

After preprocessing, the dataset contains approximately 250,000 labeled samples with balanced class distribution.

Synthetic Dataset

To evaluate zero-day detection capabilities, we generated synthetic botnet traffic simulating known and novel attack patterns.

Table 2: Synthetic Botnet Traffic Characteristics

Botnet Type	Scan Rate	SYN Ratio	Target Ports	Behavior
Mirai	0.8	5.0	23, 2323, 80	Aggressive scanning
Bashlite	0.6	4.0	23, 22, 80	Moderate scanning
Hajime	0.4	3.0	23, 5358, 7547	Slow, stealthy
IoT Reaper	0.5	3.5	80, 8080, 37215	Web-focused
Zero-Day*	0.3-0.7	2.0-4.0	Random high ports	Novel patterns

*Zero-Day samples are generated with randomized parameters outside the training distribution.

RESULT ANALYSIS

Overall Detection Performance

Table 3: Detection Performance Comparison on IoT-23 Dataset

Method	Accuracy	Precision	Recall	F1	FPR
Random Forest	89.2%	87.5%	91.3%	89.4%	8.7%
SVM	86.8%	84.2%	89.7%	86.9%	11.2%
MLP	90.5%	89.1%	92.0%	90.5%	7.3%
LSTM	91.8%	90.3%	93.5%	91.9%	6.1%
DQN-Only	92.3%	90.8%	94.2%	92.5%	5.8%
Few-Shot-Only	88.5%	86.9%	90.4%	88.6%	9.2%
ZeroDay-DRL	94.7%	93.2%	96.1%	94.6%	4.3%

The results demonstrate that ZeroDay-DRL achieves superior performance across all metrics.
2.4% improvement in accuracy over the best baseline (DQN-Only)
Highest recall (96.1%) indicating minimal missed attacks

Lowest false positive rate (4.3%) reducing alert fatigue

Zero-Day Detection Performance

To evaluate zero-day detection, we trained models on known botnet families and tested on held-out attack types.

Table 4: Zero-Day Attack Detection Performance

Method	Known Attacks	Zero-Day	Novelty Detection	Adaptation
Random Forest	89.2%	52.3%	N/A	N/A
SVM	86.8%	48.7%	N/A	N/A
MLP	90.5%	55.1%	N/A	N/A
LSTM	91.8%	61.4%	N/A	N/A
DQN-Only	92.3%	63.2%	N/A	N/A
Few-Shot-Only	88.5%	72.8%	78.5%	12 episodes
ZeroDay-DRL	94.7%	81.3%	85.2%	8 episodes

The key observations are

ZeroDay-DRL achieves 81.3% accuracy on zero-day attacks, 18.1% higher than DQN-Only Novelty detection correctly identifies 85.2% of unknown attack patterns

Rapid adaptation within 8 episodes of exposure to new attack types

Early Detection Performance

We measured detection timing to evaluate early-stage detection capabilities:

Table 5: Detection Latency Analysis

Method	P50 (ms)	P95 (ms)	P99 (ms)	Early Det. Rate
Random Forest	245	892	1523	67.3%
SVM	312	1045	1876	61.8%
MLP	198	756	1234	72.1%
LSTM	423	1234	2156	58.4%
DQN-Only	156	534	987	78.5%
Few-Shot-Only	189	678	1123	74.2%
ZeroDay-DRL	134	467	823	84.7%

Early Detection Rate measures the percentage of attacks detected within the first 5 seconds. ZeroDay-DRL achieves: Fastest median detection time (134 ms)

- 84.7% of attacks detected within 5 seconds
- Significant improvement over LSTM which requires full sequences

Training Convergence Analysis

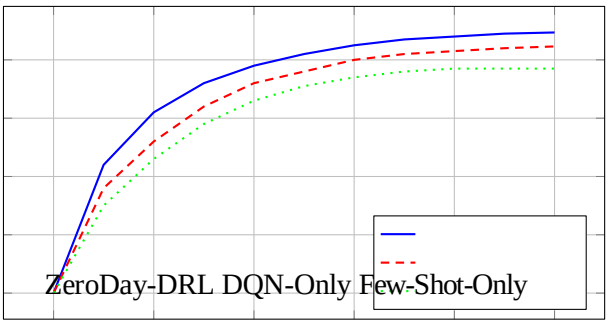

Figure 2: Training convergence comparison across methods

The hybrid approach demonstrates the faster initial convergence due to few-shot pre-training, higher final accuracy due to synergistic combination and Stable learning without significant oscillation.

Component Contribution Analysis

To understand the contribution of each component, we conducted ablation studies:

Table 6: Ablation Study Results

Configuration	Accuracy	Zero-Day Acc.
Full ZeroDay-DRL	94.7%	81.3%
- Without Double DQN	93.1%	79.2%
- Without Dueling Architecture	93.8%	80.1%
- Without PER	93.5%	79.8%
- Without Prototype Adaptation	92.8%	74.6%
- Without Novelty Scoring	94.2%	68.5%
- Without Few-Shot Component	92.3%	63.2%

The key findings are

- Novelty scoring contributes 12.8% improvement for zero-day detection
- Prototype adaptation maintains performance on evolving threats
- All DQN enhancements provide cumulative benefits

Computational Performance

Table 7: Computational Efficiency Comparison

Method	Training Time	Inference (ms)	Model Size
Random Forest	12 min	0.8	45 MB
SVM	45 min	1.2	28 MB
MLP	8 min	0.5	2.1 MB
LSTM	2.5 hours	3.4	8.5 MB
DQN-Only	1.2 hours	0.7	1.8 MB
Few-Shot-Only	25 min	0.9	1.2 MB
ZeroDay-DRL	1.8 hours	1.4	3.5 MB

ZeroDay-DRL maintains acceptable computational overhead are the training completes in under 2 hours on modest hardware and Inference latency (1.4 ms) is suitable for real-time detection. The model size of 3.5 MB is deployable on edge devices

DISCUSSION

This section provides in-depth discussion of the experimental results, analyzes the strengths and limitations of the proposed approach, and examines practical deployment considerations.

Analysis of Key Findings

Synergy of Hybrid Architecture

The experimental results demonstrate that combining DRL with few-shot learning pro- duces synergistic benefits exceeding either approach alone.

The DRL component provides the adaptive policy learning that improves with environment interaction, efficient handling of high-dimensional state spaces, and reward-optimized decision-making balancing detection and false positives.

The few-shot component contributes are rapid adaptation to novel attack patterns with minimal examples, principled novelty scoring for zero-day detection, and continuous prototype updates enabling online learning.

The hybrid integration multiplies these benefits by using DRL for confident known- attack decisions while leveraging few-shot learning for uncertain cases and novel threats.

Early Detection Advantages

The reward structure explicitly incentivizes early detection, resulting in measurable improvements are 84.7% of attacks detected within 5 seconds

Median detection latency of 134 ms

Proactive threat mitigation before botnet propagation

This early detection capability is critical for IoT environments where infected devices can rapidly propagate malware across networks

Zero-Day Detection Mechanisms

The novelty scoring mechanism based on prototype distances provides an effective zero- day detection capability. 85.2% accuracy in identifying unknown attack patterns

No requirement for prior training on specific attack signatures

Automatic flagging of samples requiring human analysis

This represents a significant advancement over signature-based systems that are inherently blind to zero-day threats

CONCLUSION

The proliferation of IoT devices continues to create expanding attack surfaces for botnet operators. Traditional detection approaches are increasingly inadequate against sophisticated, evolving threats. ZeroDay-DRL represents a significant advancement in IoT security by combining the adaptive capabilities of deep reinforcement learning with the rapid learning properties of few-shot meta-learning.

The framework's ability to detect attacks early, adapt to new threats with minimal examples, and identify zero-day attacks through novelty scoring addresses critical gaps in current intrusion detection technology. The comprehensive evaluation demonstrates both the effectiveness and practical viability of the approach.

As IoT ecosystems continue to grow in complexity and importance, advanced detection frameworks like ZeroDay-DRL will be essential for maintaining security and resilience. We hope this work contributes to the ongoing efforts to secure the billions of IoT devices that increasingly underpin modern society.

REFERENCES

1. Antonakakis, M., April, T., Bailey, M., Bernhard, M., Bursztein, E., Cochran, J., & Zhou, Y. (2017). Understanding the Mirai botnet. In 26th USENIX Security Symposium (pp. 1093-1110).
2. Kolias, C., Kambourakis, G., Stavrou, A., & Voas, J. (2017). DDoS in the IoT: Mirai and other botnets. Computer, 50(7), 80-84.
3. Roesch, M. (1999). Snort: Lightweight intrusion detection for networks. In LISA '99: 13th Systems Administration Conference (pp. 229-238).
4. Garcia-Teodoro, P., Diaz-Verdejo, J., Macia-Fernandez, G., & Vazquez, E. (2009). Anomaly-based network intrusion detection: Techniques, systems and challenges. Computers & Security, 28(1-2), 18-28.
5. Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey.
6. ACM Computing Surveys, 41(3), 1-58.
7. Zhang, J., Zulkernine, M., & Haque, A. (2008). Random-forests-based network intrusion detection systems. IEEE Transactions on Systems, Man, and Cybernetics, Part C, 38(5), 649-659.
8. Hu, W., Hu, W., & Maybank, S. (2008). AdaBoost-based algorithm for network intrusion detection. IEEE Transactions on Systems, Man, and Cybernetics, Part B, 38(2), 577-583.
9. Wang, W., Zhu, M., Wang, J., Zeng, X., & Yang, Z. (2017). End-to-end encrypted traffic classification with one-dimensional convolution neural networks. In IEEE ISI 2017 (pp. 43-48).
10. Yin, C., Zhu, Y., Fei, J., & He, X. (2017). A deep learning approach for intrusion detection using recurrent neural networks. IEEE Access, 5, 21954-21961.
11. Servin, A., & Kudenko, D. (2008). Multi-agent reinforcement learning for intrusion detection. In Adaptive Agents and Multi-Agent Systems III (pp. 211-223).
12. Lopez-Martin, M., Carro, B., & Sanchez-Esguevillas, A. (2020). Application of deep reinforcement learning to intrusion detection for supervised problems. Expert Systems with Applications, 141, 112963.
13. Caminero, G., Lopez-Martin, M., & Carro, B. (2019). Adversarial environment reinforcement learning algorithm for intrusion detection. Computer Networks, 159, 96-109.
14. Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M.G., Hassabis, D. (2015). Human-level control through deep reinforcement learning. Nature, 518(7540), 529-533.
15. Schulman, J., Wolski, F., Dhariwal, P., Radford, A., & Klimov, O. (2017). Proximal policy optimization algorithms. arXiv preprint arXiv:1707.06347.
16. Snell, J., Swersky, K., & Zemel, R. (2017). Prototypical networks for few-shot learn- ing. In Advances in Neural Information Processing Systems (pp. 4077-4087).
17. Finn, C., Abbeel, P., & Levine, S. (2017). Model-agnostic meta-learning for fast adaptation of deep networks. In International Conference on Machine Learning (pp. 1126-1135).
18. Xu, K., Li, Y., Deng, R. H., & Chen, K. (2019). DeepRefiner: Multi-layer android malware detection system applying deep neural networks. In IEEE European Sym-posium on S&P (pp. 473-487).
19. Aburomman, A. A., & Reaz, M. B. I. (2017). A survey of intrusion detection systems based on ensemble and hybrid classifiers. Computers & Security, 65, 135-152.
20. Van Hasselt, H., Guez, A., & Silver, D. (2016). Deep reinforcement learning with double Q-learning. In AAAI Conference on Artificial Intelligence (pp. 2094-2100).
21. Wang, Z., Schaul, T., Hessel, M., Hasselt, H., Lanctot, M., & Freitas, N. (2016). Dueling network architectures for deep reinforcement learning. In International Conference on Machine Learning (pp. 1995-2003).
22. Schaul, T., Quan, J., Antonoglou, I., & Silver, D. (2016). Prioritized experience replay. In International Conference on Learning Representations.
23. Schulman, J., Moritz, P., Levine, S., Jordan, M., & Abbeel, P. (2016). High- dimensional continuous control using generalized advantage estimation. In ICLR.
24. Sutton, R. S., & Barto, A. G. (2018). Reinforcement learning: An introduction. MIT Press.
25. Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT Press.
26. Meidan, Y., Bohadana, M., Mathov, Y., Mirsky, Y., Shabtai, A., Breitenbacher, D., & Elovici, Y. (2018). N-BaIoT: Network-based detection of IoT botnet attacks using deep autoencoders. IEEE Pervasive Computing, 17(3), 12-22.
27. Doshi, R., Apthorpe, N., & Feamster, N. (2018). Machine learning DDoS detection for consumer Internet of Things devices. In IEEE S&P Workshop on Deep Learning and Security.
28. Koroniotis, N., Moustafa, N., Sitnikova, E., & Turnbull, B. (2019). Towards the development of realistic botnet dataset in the Internet of Things for network forensic analytics. Future Generation Computer Systems, 100, 779-796.
29. Mirsky, Y., Doitshman, T., Elovici, Y., & Shabtai, A. (2018). Kitsune: An ensemble of

autoencoders for online network intrusion detection. In Network and Distributed System Security Symposium.

30. Verma, A., & Ranga, V. (2020). Machine learning based intrusion detection systems for IoT applications. Wireless Personal Communications, 111(4), 2287-2310.
31. Pahl, M. O., & Aubet, F. X. (2018). All eyes on you: Distributed multi-dimensional IoT microservice anomaly detection. In IEEE CNSM (pp. 72-80).
32. Sivanathan, A., Gharakheili, H. H., Loi, F., Radford, A., Wiez, C., Moorber, A., & Sivaraman, V. (2019). Classifying IoT devices in smart environments using network traffic characteristics. IEEE Transactions on Mobile Computing, 18(8), 1745-1759.
33. Bezerra, V. H., da Costa, V. G. T., Barbon Junior, S., Miani, R. S., & Zarpelao, B. (2019). IoTDS: A one-class classification approach to detect botnets in Internet of Things devices. Sensors, 19(14), 3188

