

# Multi-stage Attack Detection: Emerging Challenges for Wireless Networks

Moemedi Lefoane, Ibrahim Ghafir, Sohag Kabir, and Irfan-Ullah Awan

Department of Computer Science, University of Bradford, Bradford, UK

Email: m.lefoane@bradford.ac.uk, i.ghafir@bradford.ac.uk, s.kabir2@bradford.ac.uk, i.u.awan@bradford.ac.uk

**Abstract**—Multi-stage attacks (MSAs) are among the most serious threats in cyberspace today. Criminals target big organisations and government critical infrastructures mainly for financial gain. These attacks are becoming more advanced and stealthier, and thus have capabilities to evade Intrusion Detection Systems (IDSs). As a result, the attack strategies used in the attack render IDSs ineffective, particularly because of new security challenges introduced by some of the key emerging technologies such as 5G wireless networks, cloud computing infrastructure and Internet of Things (IoT). Advanced persistent threats (APTs) and botnet attacks are examples of MSAs, these are serious threats on the Internet. This work analyses recent MSAs, outlines and reveals open issues, challenges and opportunities with existing detection methods.

**Index Terms**—5G, Internet of Things, Multi-Stage attack, Botnet Attack, Advanced Persistent Threats, Machine Learning

## I. INTRODUCTION

Advancements in cloud computing technologies, processors and emerging 5G & 6G wireless networks have accelerated digital transformation and enabled the fourth Industrial Revolution (Industry 4.0) to take off, with benefits including increased productivity and automation with Artificial Intelligence (AI). These advancements are key enabling technologies for the IoT. Indeed, IoT is everywhere. Increasing deployment of IoT technologies can be observed across different organisational settings, from smart cities to Industrial Internet of Things (IIoT) like smart grids, transportation and healthcare.

These advancements have introduced inherent security challenges and new vulnerabilities, and thus expanded the attack surface. Cyber-criminals exploit vulnerabilities in these technologies to gain unauthorised access to these systems for financial gain and criminal activities. As a result, cybersecurity has become even more important than ever. MSAs are serious threats to some of the state-of-the-art technologies deployed today, and they are typically targeted at large organisations and critical national infrastructures. Examples of MSAs include botnet attacks and APTs. These attacks are launched in sequential stages, and often the stages are not malicious when considered individually [1]. The attack strategies have become more advanced and sophisticated. One of the key differences between MSAs and traditional once-off attacks is that for MSAs, an attack is launched in multiple and sequential steps, the steps are stealthier and therefore extremely challenging to detect [1].

This article provides a high-level overview of two MSAs: botnet attacks and APTs. Afterwards, it investigates IoT botnet detection and APTs detection techniques proposed over the last 5 years period to establish the state-of-the-art and suitability of proposed techniques for emerging research area that includes emerging wireless networks: 5G and 6G, and the explainability and trustworthiness of ML models. Challenges and trends relating to the reviewed solutions are discussed in the context of the emerging wireless networks (5G and 6G) and the IoT.

The rest of the paper is organised as follows: Section II gives an overview of MSAs, Section III presents existing MSA detection methods, Section IV outlines proposed solution for IoT Botnet Detection, Section V discusses the open challenges and finally Section VI concludes the paper.

## II. MULTI-STAGE ATTACKS

MSAs are executed in a number of stages and are among serious cyber security threats. The challenge is that they are implemented in stages that typically are not malicious when they are executed independently, but lethal when considered together [1]. The stages are shown in Fig. 1.

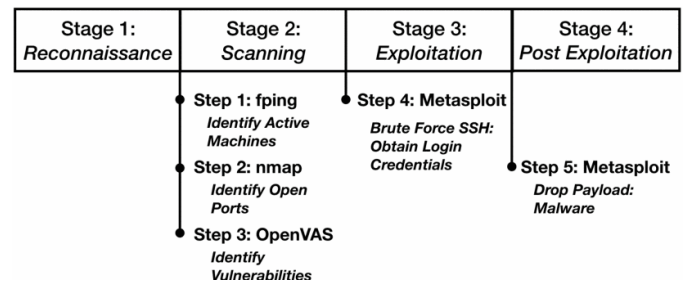


Fig. 1. Multi-stage attack (MSA) stages [1].

APTs are example of MSAs. Fig. 2 shows APTs attack stages. The attack begins with reconnaissance, which includes enumeration and social engineering [2], Point of Entry, C&C Server communication through to ultimate objectives such as data ex-filtration.

Botnet attacks are another example of MSAs, botnet attack stages are illustrated in Fig. 3. Several IoT botnet detection approaches have been proposed in recent years, particularly in the area of Machine Learning (ML). Specifically, detection approaches have been proposed from supervised and unsupervised learning to deep learning approaches. By and large, the proposed approaches yield promising results. Largely, the



Fig. 2. APT stages of attack [3].

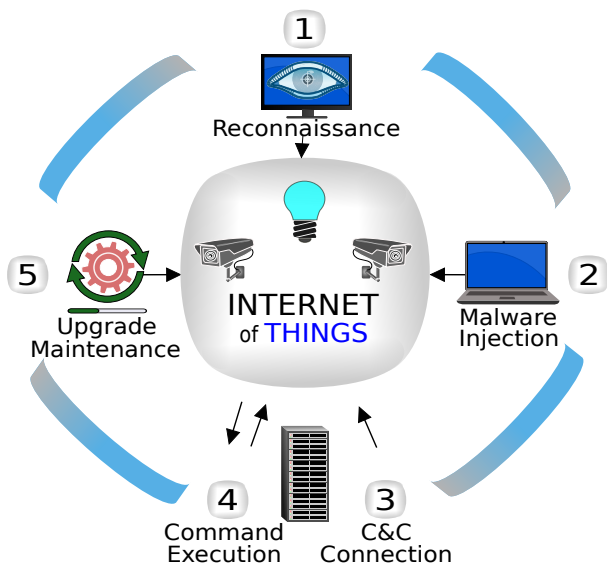


Fig. 3. Botnet stages of attack (Adapted from [4]).

success of ML approaches is attributed to feature engineering and related preprocessing tasks [5], [6]. This is the stage in the ML process that utilises techniques resulting from expert knowledge about a domain. The curse of dimensionality [7] problem is addressed by dimensionality reduction and feature selection approaches [8].

### III. RELATED WORK

This section presents related work to MSA detection and discusses current existing solutions applied to MSAs. In their work, Ghafir et al. [9] proposed APTs defence module. Their work considers that persistent communication is crucial for a successful APTs attack and communication protection by

Secure Socket Layer (SSL) encryption. Given encrypted communication, analysis of network traffic with the aim of picking attack trail is extremely challenging. To address these challenges their module aims to detect Malicious SSL Certificate within a target network, specifically the detection module is based on detection of blacklisted SSL Certificates within a target network. Their detection method leverages Intelligence feed for access to updated blacklisted SSL Certificates. Their proposed method is evaluated on an implemented Virtual Network.

Aparicio-Navarro *et al.* [1] proposed a real-time multi-stage attack detection approach; their proposed approach utilises contextual information and leverages expert judgment relating to network behaviour. Their approach does not require a training process and they reported 58% performance enhancement of multi-stage attack detection rate. Khosri and Ladani [10] described APTs as characterised by long-term and stealthy stepwise attack strategies that are slow, and thus alerts for these attacks are typically below threshold levels for IDSs and in some instances, they are not malicious at all. In their study, they modelled causal relationships between APTs stages. Shawly et al. [11] investigated MSAs and proposed a detection model based on Hidden Markov Model (HMM). Within the proposed model, HMM templates are deployed with known attacks. Their work is evaluated in a simulated environment.

APT's detection model based on a Game-Theoretic approach is explored in [12]. The proposed model track the flow of information changing dynamically to detect traces of APTs attacks. The proposed is a nonzero-sum game. The model proposed utilises graph theory, with their nodes represented by file and network points. The interaction between nodes and processes are captured by the edges of the graphs. For detection of stages of APTs that are inter-related, correlated equilibrium is computed based on polynomial-time algorithm. Their proposed model is evaluated on Attack Investigation (RAIN) data.

Ghafir et al. [13] investigated botnet attacks and proposed a botnet attack detection approach. Their proposed approach focuses on detection of network traffic from the command and control (C&C) server.

Vinayakumar et al. [14] proposed a deep learning botnet detection framework, which operates at DNS services of the application layer. It works by distinguishing normal behaviour from botnet behaviour. The proposed approach works on two levels: 1) first DNS query similarity scores are determined and this is done based on a predefined threshold. 2) the second level focuses on the normal and abnormal classification of domain names generated by domain generation algorithms. The domain generation algorithm is developed based on deep learning architectures.

Anthi et al. [15] notes an exponential growth in the adoption of IoT technologies. Admittedly, there are massive benefits, mainly by enabling automation of services that simplify many aspects of our daily activities/tasks, part of the process that enables this automation involves regular collection and trans-

mission of sensitive data between the IoT devices and the cloud. Unfortunately, the benefits of IoT come with huge security flaws as security mechanisms do not keep up with the proliferation rate, thus creating a big gap in research to be filled to defend and secure this technology.

Ali et al. [16] reviewed literature on IoT botnet attacks systematically. Their work revealed trends and challenges relating to botnet detection methods, the trends shows that research in IoT botnet detection has been gaining momentum recently motivated by some of the serious threats in the internet (DDoS attacks particularly launched from IoT device). The adversaries are developing more sophisticated techniques that have the capabilities to evade detection methods. Given the challenges outlined, research towards countering these attack strategies is of paramount importance.

Sudheera et al. [17] investigated MSAs and proposed a distributed approach for detection of MSAs. Their work addresses botnet attack spacio-temporal characteristics challenges. The proposed solution undergoes three phases: first, the normal behaviour of connected devices are learned and normal profiles for these behaviour are generated and stored in cuckoo hash table for quick retrieval. These profiles are stored within the perimeter of an IoT network. In the second phase, monitoring of the network commences for detection of MSA traffic. An alert generation is triggered by any deviation from the expected behaviour of devices connected to the network. The generated alerts are sent to the security manager in the cloud where they are accumulated from different networks and analysed together to enhance detection and therefore improve detection rate. The proposed approach utilises Frequent Itemset Mining. They evaluated performance of the proposed approach on their generated dataset and a publicly available dataset.

Stephens et al. [18] reviewed IoT botnet detection literature over a period of five years leading to year 2021. Their work looked into the impact of IoT botnets. Specifically, they looked at advances, proposed detection and prevention approaches. The comparison of the different recent botnet attacks in the literature revealed that 95% of IoT formed botnets were successful due to exploitation of default password credentials, this shows that a significant number of attacks are not due to novel vulnerabilities exploited but rather basic ones. The results of such attacks lead to even more catastrophic consequences. Their work further identifies different methods of detection that are utilised namely, host-based detection and network-based detection methods. The focus of the detection methods can either be reactive or proactive (offensive vs defensive). In host-based methodology, an analysis to effectively detect breaches is performed on the device. The host-based analysis can be done in one of two ways; first, the static method can be employed where the analysis is performed on binaries without execution of the code. For the second methodology, the analysis is performed dynamically, where the sandbox is created and monitored, subsequently, the effects of executing the code are observed.

#### IV. PROPOSED SOLUTION FOR IoT BOTNET DETECTION

Fig. 4 shows a proposed topology for an IoT system with IDS for the detection of malicious intrusions including botnet attacks.

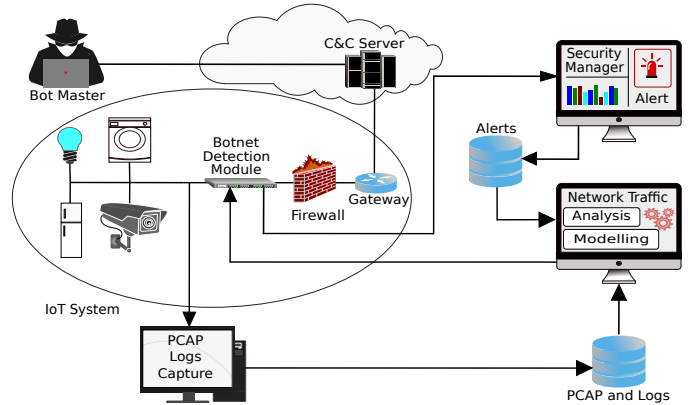


Fig. 4. A proposed IoT IDS topology.

The majority of the work proposed utilises ML techniques. Research opportunities that tend to advance the state of the art are found in the pre-processing stages of the ML process: feature selection, feature extraction and feature Engineering. Fig. 5 illustrates the steps followed in the pre-processing stage that prepares network traffic data for building the model for botnet detection.

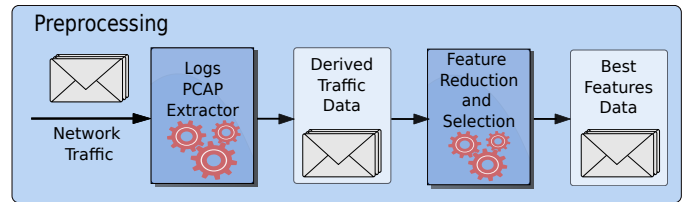


Fig. 5. Steps in pre-processing stage for IoT botnet detection model creation

Once the pre-processing stage is complete, the resulting best set of features data is passed to the next stage: training of ML models and evaluation of performance. Where labelled data is available, supervised learning models are trained and for unlabelled datasets, unsupervised learning approaches are used to build models. Fig. 6 illustrates the steps followed.

As an emerging topic within the research community, explainability of ML models utilised for building detection model is crucial for the success of botnet detection models. Explainability ensures that applied models are explainable, with sufficient and justified reasons behind actions taken by the model, this ensures safety and trustworthiness of the model, which is yet another important aspect that must be considered when developing detection models. Currently, the existing approaches for ML-based botnet detection do not consider this aspect, therefore, in the future, this aspect needs to be explored. As a potential avenue, one may consider methods like SafeML [19] to address the issues of explainability and trustworthiness.

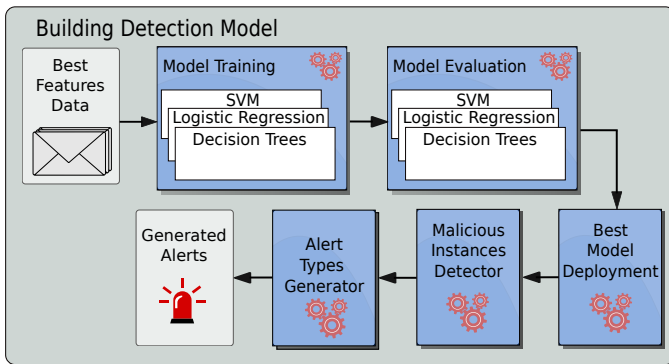


Fig. 6. Steps of creating a detection model

## V. CHALLENGES

While several IoT botnet detection techniques have been proposed in the literature with good performance, considering the latency requirements of emerging wireless networks, the proposed approaches are best suited for 3G and 4G networks and not for 5G & 6G. 5G is one of the key enabling technologies for IoT networks and latency requirements for 5G is 1ms [16]. This is relevant because, if the proposed approaches do not meet the latency requirements of 5G and 6G networks, they will be rendered ineffective for IoT botnet detection. A challenge or limitation associated with datasets used for analysis is that the majority of the datasets are imbalanced, which might affect the performance of the approaches, and potentially make it difficult to evaluate the performance and benchmark against state-of-the-art. Therefore extra caution needs to be taken when evaluating these models to determine how effective they are in predicting malicious traffic. Botnet attacks occur at a large scale spanning different IoT networks for launching DDoS attacks. The solutions for detection against botnet attacks focus more on individual networks [17], this means that the detection approaches may miss some trails of attacks that happen across different networks.

Inherent limitations in IoT make it difficult to implement state-of-the-art cryptographic security mechanisms. Heterogeneity is another major challenge. Detection methods tend to be mostly deployed at local network while the attacks span multiple networks. ML has been widely applied across different fields ranging from computer vision, cyber-physical systems, threat detection e.t.c [20]–[22], these include statistical and anomaly detection approaches [23], [24]. While ML approaches have proven to be effective in terms of performance, often the solutions are black-box in nature [25], as such not explainable to humans, this affect trustworthiness of these ML solutions. Indeed this has sparked interest in explainability and trustworthiness, aimed at optimizing and therefore maximising performance of ML approaches [25]. Existing ML detection solutions do not consider explainability and trustworthiness of the proposed systems [8], [16]. Furthermore, proposed systems do not test solutions for suitability to emerging wireless networks which will form significant components in some of the critical infrastructure deployments,

and thus may be rendered inefficient. For APTs, the lack of standard datasets is still a challenge.

## VI. CONCLUSION

While MSAs are becoming more advanced and stealthier, continuous efforts have been made to develop intelligent methodologies to combat these attacks. Although the existing solutions have their proven strength in detecting MSAs, they have their own limitations. This article provides a high-level overview of a couple of examples of MSAs together with some of the proposed solutions found in the literature. The open issues and challenges revealed in this article will help to direct future research efforts to address them, thus improving the effectiveness of the approaches and further improving their application potential in a wide range of networks.

## REFERENCES

- [1] F. J. Aparicio-Navarro, K. G. Kyriakopoulos, I. Ghafir, S. Lambotharan, and J. A. Chambers, "Multi-stage attack detection using contextual information," in *MILCOM 2018 - 2018 IEEE Military Communications Conference (MILCOM)*, 2018, pp. 1–9.
- [2] I. Ghafir, K. G. Kyriakopoulos, S. Lambotharan, F. J. Aparicio-Navarro, B. Assadhan, H. Binsalleh, and D. M. Diab, "Hidden markov models and alert correlations for the prediction of advanced persistent threats," *IEEE Access*, vol. 7, pp. 99 508–99 520, 2019.
- [3] Trend-Micro, "The custom defense against targeted attacks," <http://www.trendmicro.fr/media/wp/custom-defense-against-targeted-attacks-whitepaper-en.pdf>, accessed: 2019-03-25.
- [4] F. Hussain, S. G. Abbas, I. M. Pires, S. Tanveer, U. U. Fayyaz, N. M. Garcia, G. A. Shah, and F. Shahzad, "A two-fold machine learning approach to prevent and detect iot botnet attacks," *IEEE Access*, vol. 9, pp. 163 412–163 430, 2021.
- [5] M. Panda, A. A. A. Mousa, and A. E. Hassanien, "Developing an efficient feature engineering and machine learning model for detecting iot-botnet cyber attacks," *IEEE Access*, vol. 9, pp. 91 038–91 052, 2021.
- [6] M. Lefoane, I. Ghafir, S. Kabir, and I.-U. Awan, "Unsupervised learning for feature selection: A proposed solution for botnet detection in 5g networks," *IEEE Transactions on Industrial Informatics*, pp. 1–9, 2022.
- [7] E.-w. Bai, C. Cheng, W. Zhao, and H.-F. Chen, "Variable selection of high-dimensional non-parametric nonlinear systems: A way to avoid the curse of dimensionality," in *2017 IEEE 56th Annual Conference on Decision and Control (CDC)*, 2017, pp. 6469–6474.
- [8] M. Lefoane, I. Ghafir, S. Kabir, and I.-U. Awan, "Machine learning for botnet detection: An optimized feature selection approach," in *The 5th International Conference on Future Networks & Distributed Systems*, ser. ICFNDS 2021. New York, NY, USA: Association for Computing Machinery, 2021, p. 195–200.
- [9] I. Ghafir, V. Prenosil, M. Hammoudeh, L. Han, and U. Raza, "Malicious ssl certificate detection: A step towards advanced persistent threat defence," in *Proceedings of the international conference on future networks and distributed systems*, 2017.
- [10] M. Khosravi and B. T. Ladani, "Alerts correlation and causal analysis for apt based cyber attack detection," *IEEE Access*, vol. 8, pp. 162 642–162 656, 2020.
- [11] T. Shawly, A. Elghariani, J. Kobes, and A. Ghafoor, "Architectures for detecting interleaved multi-stage network attacks using hidden markov models," *IEEE Transactions on Dependable and Secure Computing*, vol. 18, no. 5, pp. 2316–2330, 2021.
- [12] S. Moothedath, D. Sahabandu, J. Allen, A. Clark, L. Bushnell, W. Lee, and R. Poovendran, "A game-theoretic approach for dynamic information flow tracking to detect multistage advanced persistent threats," *IEEE Transactions on Automatic Control*, vol. 65, no. 12, pp. 5248–5263, 2020.
- [13] I. Ghafir, V. Prenosil, M. Hammoudeh, T. Baker, S. Jabbar, S. Khalid, and S. Jaf, "Botdet: A system for real time botnet command and control traffic detection," *IEEE Access*, vol. 6, pp. 38 947–38 958, 2018.



- [14] R. Vinayakumar, M. Alazab, S. Srinivasan, Q.-V. Pham, S. K. Padanayil, and K. Simran, "A visualized botnet detection system based deep learning for the internet of things networks of smart cities," *IEEE Transactions on Industry Applications*, vol. 56, no. 4, pp. 4436–4456, 2020.
- [15] E. Anthi, L. Williams, M. Słowińska, G. Theodorakopoulos, and P. Burnap, "A supervised intrusion detection system for smart home iot devices," *IEEE Internet of Things Journal*, vol. 6, no. 5, pp. 9042–9053, 2019.
- [16] I. Ali, A. I. A. Ahmed, A. Almogren, M. A. Raza, S. A. Shah, A. Khan, and A. Gani, "Systematic literature review on iot-based botnet attack," *IEEE Access*, vol. 8, pp. 212 220–212 232, 2020.
- [17] K. L. K. Sudheera, D. M. Divakaran, R. P. Singh, and M. Gurusamy, "Adept: Detection and identification of correlated attack stages in iot networks," *IEEE Internet of Things Journal*, vol. 8, no. 8, pp. 6591–6607, 2021.
- [18] B. Stephens, A. Shaghghi, R. Doss, and S. S. Kanhere, "Detecting internet of things bots: A comparative study," *IEEE Access*, vol. 9, pp. 160 391–160 401, 2021.
- [19] K. Aslansefat, S. Kabir, A. Abdullatif, V. Vasudevan, and Y. Papadopoulos, "Toward improving confidence in autonomous vehicle software: A study on traffic sign recognition systems," *Computer*, vol. 54, no. 8, pp. 66–76, 2021.
- [20] Q. Liu, P. Li, W. Zhao, W. Cai, S. Yu, and V. C. M. Leung, "A survey on security threats and defensive techniques of machine learning: A data driven view," *IEEE Access*, vol. 6, pp. 12 103–12 117, 2018.
- [21] J. White, T. Kameneva, and C. McCarthy, "Vision processing for assistive vision: A deep reinforcement learning approach," *IEEE Transactions on Human-Machine Systems*, vol. 52, no. 1, pp. 123–133, 2022.
- [22] B. A. Salau, A. Rawal, and D. B. Rawat, "Recent advances in artificial intelligence for wireless internet of things and cyber-physical systems: A comprehensive survey," *IEEE Internet of Things Journal*, vol. 9, no. 15, pp. 12 916–12 930, 2022.
- [23] D. M. Diab, B. AsSadhan, H. Binsalleeh, S. Lambbotharan, K. G. Kyriakopoulos, and I. Ghafir, "Anomaly detection using dynamic time warping," in *2019 IEEE International Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC)*, 2019, pp. 193–198.
- [24] I. Ghafir, K. G. Kyriakopoulos, F. J. Aparicio-Navarro, S. Lambbotharan, B. Assadhan, and H. Binsalleeh, "A basic probability assignment methodology for unsupervised wireless intrusion detection," *IEEE Access*, vol. 6, pp. 40 008–40 023, 2018.
- [25] C. S. Wickramasinghe, K. Amarasinghe, D. L. Marino, C. Rieger, and M. Manic, "Explainable unsupervised machine learning for cyber-physical systems," *IEEE Access*, vol. 9, pp. 131 824–131 843, 2021.