Journal of Ubiquitous Systems & Pervasive Networks
Volume 15, No. 2 (2021) pp. 01-09

AI'S Contribution to Ubiquitous Systems and Pervasive Networks Security – Reinforcement Learning vs Recurrent Networks

Christophe Feltus *

Luxembourg Institute of Science and Technology (LIST), Esch-sur-Alzette, Luxembourg, L-4362

Abstract

Reinforcement learning and recurrent networks are two emerging machine-learning paradigms. The first learns the best actions an agent needs to perform to maximize its rewards in a particular environment and the second has the specificity to use an internal state to remember previous analysis results and consider them for the current one. Research into RL and recurrent network has been proven to have made a real contribution to the protection of ubiquitous systems and pervasive networks against intrusions and malwares. In this paper, a systematic review of this research was performed in regard to various attacks and an analysis of the trends and future fields of interest for the RL and recurrent network-based research in network security was complete.

Keywords: *Artificial Intelligence, Reinforcement Learning, Recurrent Networks, RNN, GRU, LSTM, RL, Security, Network Security, Malware Detection, Literature Review, Machine learning, Intrusion Detection, State of the Art.*

1. Introduction

The contribution of artificial intelligence to network and to ubiquitous system security is paramount, given that it has the potential to increase the security level of the defended system [8] up to the state-of-the-art level generally reached by the attackers. The field of artificial intelligence and machine learning (ML) is generally classified into three paradigms: supervised, unsupervised and reinforcement learning (RL). This paper will focus on the contribution of reinforcement learning and of recurrent networks algorithms (supervised, semi-supervised and unsupervised) to ubiquitous systems security, principally on the field of intrusions and malwares detection.

Research into RL has been proven to have made a real contribution to the protection of ubiquitous systems against intrusions and malwares. The principle of RL is that software agents learn to react on their own to an environment that they do not yet know [28]. In order to learn how to react, agents make decisions and take actions with the objective of accumulating rewards while avoiding errors. The volume of scientific contributions based on Recurrent networks is no less important. Recurrent networks consist in a family of network used to analyze an input based on the output of the previous analysis. this paper focuses on three types of recurrent networks, to know: RNN, LSTM and GRU.

So far as we know, depicting the contribution of RL and recurrent network to various attacks (phishing, domain generating algorithm, injection, sybil, jamming, adversarial, eavesdropping, spoofing, (D)DOS, botnet and ransomware) and understanding which algorithm is relevant for which attack has not been achieved yet.

Elaborated from the strategic literature review method [22], the paper will successively answer two research questions:

- *What is the reinforcement learning and recurrent networks' contribution and prospects for the field of malware and intrusion detection ?* and,
- *What are the most important contributions of the RL and of the recurrent networks to the most known security attacks ?*

1.1 Research material

We began this review of the literature, according to [22] by a systematic investigation of the IEEExplorer database, which includes 5,1 10^6 records and the ACM database, which includes 2,8 10^6 records. For both databases, the following search streams were defined:

At the level of IEEExplore database, we refined the number of records by applying the keywords "security", which gives 122,374 records, "network-security" which gives 109,199 records, and finally, "reinforcement learning" which gives 9,928 records. By searching both the "network security" and "reinforcement learning" keywords together at the abstract level,

* Corresponding author. Tel.: + 352 27 58 88
E-mail: Christophe.Feltus@list.lu

DOI: 10.5383/JUSPN.15.02.001

the result was refined to 105 papers for all dates. Then, the keywords “RNN”, ”LSTM” or ”GRU” were entered, giving 881 records. By entering both “security” together with “RNN”, ”LSTM” or ”GRU” at the abstract level, the result was refined to 163 papers for all dates. Given the recent developments of the RL and recurrent networks, we applied a final filter to our research, limiting the papers to the 2010-2021 period, which give us 253 abstracts to read. After reading them, it appeared that there were 212 remaining papers from IEEE worth being considered for the systematic review.

Afterward, we applied the same approach to the ACM Guide to Computing Literature. By looking for the keyword “security” at the abstract level, we obtained 3,841 records, and by looking for the keywords “network security”, we obtained 151,789 records. Concerning the keyword “reinforcement learning”, it gave 9,591 results. When applying both filters together (that is, “network security” and “reinforcement learning”) at the abstract level, it gave 9 papers. When applying both filters together (that is, “IT security” or “cyber-security” and “RNN”, “LSTM”, ”GRU”) at the abstract level, it gave 1,441 papers. Then when applying the filter only to the 2010-2020 period, it reduced the volume of relevant papers to 307. After reading the abstracts, only 121 papers appeared worth considering for the analysis and will be presented in the next sections.

In parallel, this selection of papers was applied to Springer, Science Direct, Google Scholar and Web of Science but only a limited amount of new papers were discovered. In addition to the selection criteria defined above, in order to limit the number of papers selected, the following exclusion criteria were retained: physical security and personal security are outside the scope of the studyIn the end, this increased the final volume of papers to 351. The evolution of the number of papers published the last decade is illustrated on Fig.1. This figure shows the exponential development of the number of contributions in the last two years.

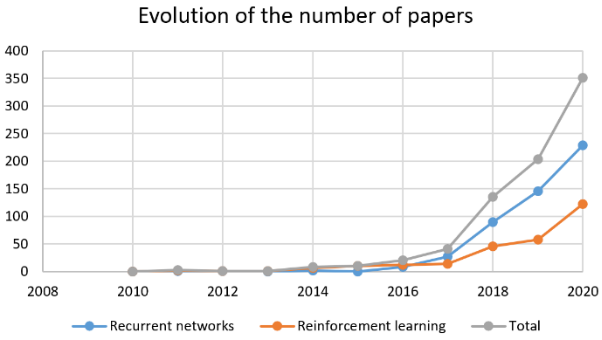

Fig. 1: **Evolution of the number of papers by years**

The paper is structured as follows: Section 2 reminds us of the theoretic background necessary to understand reinforcement learning and recurrent networks Section 3 reviews the literature related to the RL and recurrent networks’ contribution to IS and ubiquitous system security, and Section 4 concludes the paper.

2. Theoretic background

Reinforcement learning and recurrent neural networks are two machine learning techniques fundamentally different. This section summarizes the structure of the deep neural network frameworks analyzed and the advantages of each of them.

2.1. Reinforcement learning

Reinforcement learning involves agents, states (S), and actions per state (A). Agents evolves from state to state when they perform actions. In order to learn how to react, agents make decisions and take action at time t, A^t – (Fig. 2) with the objective of accumulating rewards (R^t) while avoiding errors. As RL algorithms mostly use dynamic programming techniques, this reward-based environment is typically represented in the of Markov decision processes. These processes reflect a straightforward description of the problem in order to learn to reach a desired goal. In practice, agents continually select actions while the form environment in which they behave responds and presents new situations (Fig. 2)

In contrast to classical dynamic programming methods, RL algorithms have no knowledge of the exact Markov decision processes. Q-Learning [50] is an RL algorithm, whose purpose is to learn the policy that informs agents of the action they have to achieve in determined situations. This policy is optimized and gives all the successive steps necessary to achieve a goal while maximizing the gain of the rewards. Agents that learn the environment must continuously choose between exploiting the knowledge learned and exploring new potential actions to perform. Hence, an important parameter to be considered while defining RL algorithms is the ε-greedy, which represents the proportion of exploration vs. exploitation actions (e.g., [51]).

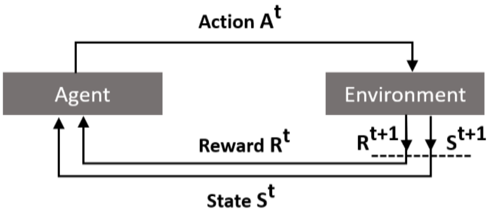

Fig. 2: **Structure of a reinforcement learning algorithm**

2.2. Recurrent network

Recurrent networks consist in a family of network used to analyze an input based on the output of the previous analysis. this paper focuses on three types of recurrent networks, to know: RNN, LSTM and GRU.

2.2.1. RNN

In a traditional neural network, the pieces of data injected into the network are independent from each other (Fig. 3).

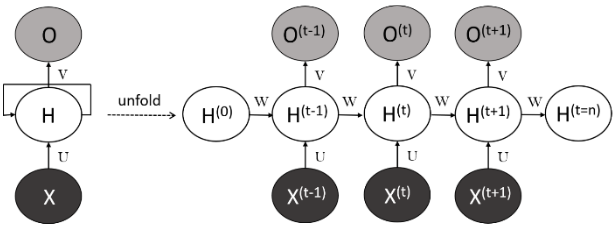

Fig. 3: **Structure of a reinforcement learning algorithm**

Sometimes, it is necessary to analyze connected data inputs, with their outputs being dependent on the previous ones (e.g. video frame sequences). In those cases, the analysis output of the first piece of data (e.g. a video frame) needs to be considered for the analysis of the subsequent ones. As a consequence, the neural network (NN) has to retain information that it passes from one iteration to another.

Recurrent neural networks are types of NN that can use an internal state, and the latter act as a memory. As explained in Fig. 3., RNN performs an identical function on each input (X^t) and the subsequent output of each piece of data depends on the previous computation (H^t). After computation, this output (O^t) is sent back to the RNN up to the last iteration.

2.2.2. LSTM

The results of the RNN are very good when they concern short-term dependencies between inputs. However, in more complex situations, where specific and important information (like the context) needs to be retained for a long time, RNN tends to fail. This problem is due to the fact that, as for all neural networks, to learn the weights of each neuron and to adapt them at each iteration, a gradient descent is used to minimize the sum squared error between the output values and the target values. This correction is than propagated to all layers of the network using a back-propagation algorithm. For long sequences, the derivative value (calculated by the gradient descent) is multiplied many times (as many times as there are inputs) and consequently tends to be insignificant in the end. This problem is known as the *vanishing gradient*.

Long-short Time Memory is a sophisticated type of RNN used to combat this problem (Fig. 4). It is composed of memory blocks called cells and of two states (cell state (C^t) and hidden state (h^t)) that represent the memories and that, as for the traditional RNN, are translated from one cell to another. The functioning of the cell (to retain relevant information) is achieved by means of gates whose roles are to add or remove information to the memories. LSTM includes three gates:

- The **input (i^t) gate**'s role is to add information to the cell state. Therefore, it regulates what information needs to be entered into the cell state, then it creates a vector that contains all this information, and finally, it adds this information to the cell state using an addition operation.
- The **forget (f^t) gate**'s role is to remove information from the cell state. Therefore, the information which is no longer necessary for comprehension purposes is withdrawn through the multiplication of a filter.
- The **output (o^t) gate**'s role is to select useful information from the cell state and transfer it as an output. Therefore, the output gate creates a vector with the relevant information, then creates a filter to regulate the information that needs to be transferred. Finally, it multiplies the value of this filter to the vector created with the relevant information and sends it (i) as an output of the cell, and (ii) to the hidden state of the following cell.

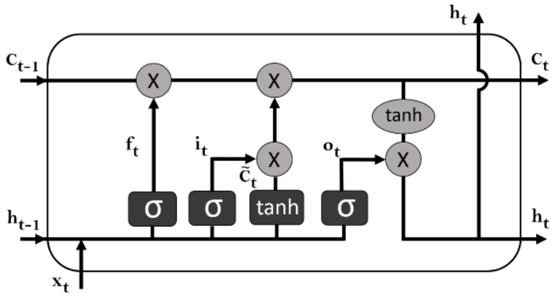

Fig. 4: Structure of the Long Short Time Memory.

2.2.3. GRU

The Gated Recurrent Unit (Fig. 5) is a variation of LSTM that also aims to solve the problem of the vanishing gradient. The GRU has eliminated the cell state and uses the hidden state (h^t) to transfer information. It also only has two gates, a reset gate and update gate:

- The **update (Z^t) Gate**'s role acts similarly to the forget and input gate of a LSTM. It decides what information to throw away and what new information to add.
- The **reset (r^t) Gate** is another gate used to decide how much past information to forget.

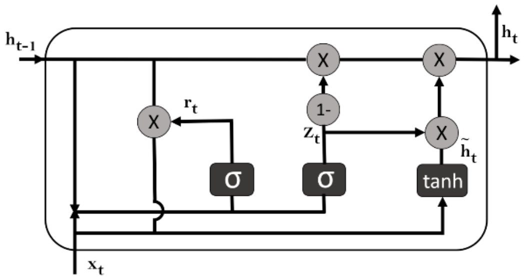

Fig. 5: Structure of the Gated Recurrent Unit.

3. Literature review and analysis

Given the increasing number of network vulnerabilities and attacks [54], developing sound malware detection appears essential for protecting information systems [27]. Hence, it is not surprising that this security topic of ubiquitous systems is the most widely addressed RL research and development field [10, 20, 21, 24, 32, 33, 37, 39]. In this regard, dedicated architectures have been a particular focus of attention of the various research works related to RL and recurrent networks.

In [24], Divyatmika et al. propose a novel approach to building a network-based IDS using a ML approach and suggest a two-tier architecture to detect intrusions at the network level. In the proposed architecture, RL allows anomaly detection considering network agents that learn from and make decisions based on the environment. Navarro-Lara et al. [20] emphasize the contribution of a human expert for threat detection and accordingly, propose the Morwilog framework to integrate alert correlation into security systems and inject human expert feedback into the system using RL.

According to [44], a lot of research is currently dedicated to wireless networks and wireless sensors as an integral part of cyber physical systems. Recently, Otoum et al. [21] exploited RL techniques on a hybrid IDS framework in wireless networks [56]. Considering a big data-driven [65] IDS approach, the authors compare and demonstrate the better performances of the RL-based IDS compared to the previously existing adaptive ML-based ones. In the field of autonomous vehicles [62], Xing et al. exploit a trust evaluation model to support a two-level IDS. Here, an attack warning is established based on (i) trust evaluation with the coverage of a roadside unit, and (ii) the information exchanged between RSUs through the cloud server. Then, an RL-based incentive mechanism reports warning by stimulating the vehicle [37]. In the same vein, [32] investigates and presents ways to use deep learning (DL) methods, including RL approaches, to improve methods for mobile crowd sensing and Zolotukhin et al. [39] stress the fact that traditional IDS approaches are unsuitable for IoT networks due to two elements: the limited computational capacity of devices and the diversity of technology. Therefore, an RL agent is proposed as a core component of an IoT defense system in order to evaluate the risk

of potential attacks [17] and mitigate them using the most optimal actions.
In [10], to maintain the high-level security of data in the Cloud, RL is incorporated to the Reinforcement Learning Automata for detecting and classifying attacks [3]. Effective rules are generated using learning automata from a vast training set to improve the learning process. Xiao et al., in [33], propose a MD scheme with Q-learning. This IDS is applied to mobile devices with the aim of deriving the optimal offloading rate without knowing (i) the trace generation and (ii) the radio bandwidth model of other mobile devices. Other malwares are addressed more specifically, such as: jamming attacks, adversarial attacks, eavesdropping, spoofing, (D)DoS, Botnet, ransomware, and some others. The impact of these attacks and their consideration by the RL-based IDS literature is shown in Fig. 6 and is discussed in the following sections.

Concerning the recurrent networks, numerous research has addressed the improvement of IDS with an LSTM, such as Li et al. [57], who propose an LSTM to establish a time correlation between situation data while improving the LSTM with a rectified linear unit, layer stacking and cross entropy function, or Wang et al. [58], who exploit LSTM in RNN units to generate an improved LSTM tree that has the ability of secondary detection to solve the problem of a high false negative in traditional RNN.

In the same vein, in He et al. [59], authors extracted various levels of features from the network connection (the opposite of traditional long feature vectors) in order to process information more efficiently separately. They also present a multi-modal-sequential IDS supported by multi-modal deep auto-encoder and LSTM technologies, which provide the advantage of automatically learning temporal information between connections amongst adjacent networks. In Le at al. [60], to limit the false positive and false alarm rate of traditional machine-learning approaches in IDS, the authors analyze the most suitable optimizer among six optimizers for LSTM-RNN and found that the *Nadam optimizer* outperformed previous solutions.

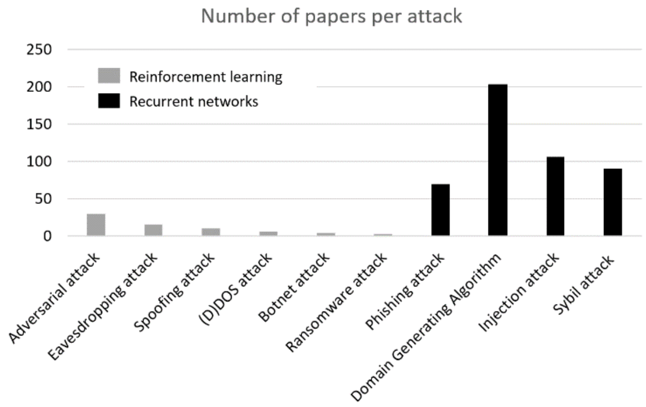

Fig. 6. Percentage of papers by type of attacks and by type of algorithm

To review the various contributions, each type of attack will be reviewed and the state of the art of the RL and recurrent networks' contributions will be presented in the following subsections.

2.1. Reinforcement learning contribution to attacks

Reinforcement learning algorithms focus mainly on jamming attack, adversarial attack, eavesdropping attack, spoofing attack, (D)DOS attack, botnet attack and ransomware attack.

2.1.1 Jamming attack

Jamming consists of creating interference within radio channels. In a jamming attack, malicious node block legitimate communication by causing intentional perturbations. This attack is a subset of denial of service attacks (Section 2.5) but, given that 25.35% of these attacks are considered in the RL-based IDS research, the related literature is reviewed independently in this section, including the solution proposed by [32] that aims to improve mobile crowd sensing security methods, including anti-jamming transmissions. In [35], Xiao et al. investigate attack models for IoT systems and review ML-based IoT security solutions based on RL.

Later, Abuzainab et al. [1] proposed an interference-aware routing protocol to ensure robust communication against jamming. This protocol has the purpose of allowing nodes to avoid communication holes created by jamming attacks. The authors use RL to elaborate a distributed cooperation framework to assess network conditions and make real-time decisions on whether to defend the network against a jamming attack.
In [23] and [19], an RL-based control framework is developed to prevent unauthorized unmanned aerial vehicles (UAV) from entering a target area. The challenge addressed by the authors is to accelerate the learning speed to achieve the optimal UAV control policy. This UAV control scheme enables a target estate to choose the optimal control policy to expel nearby UAVs (e.g., jamming the global positioning system signals). Wang et al. study defense strategies against DRL-based jamming attackers and put forth three diversified defense approaches: (i) proportional-integral-derivative control, (ii) usage of an imitation attacker, and (iii) development of orthogonal policies.

2.1.2. Adversarial attack

ML classifiers are vulnerable to inputs (named adversarial examples) maliciously constructed by adversarial attacks. These attacks consist of a generic subset of attacks funded on adversarial examples (e.g., a strategically-timed attack and an enchanting attack [14]) and represents 26.17% of all attacks considered by the RL-based IDS (Fig. 6) According to [30], adversarial attacks are also effective when targeting neural network policies in RL and have exposed a significant security vulnerability in ML-models [12].

Similarly, Inkawhich et al. present a new class of threat models where the adversary does not have the ability to interact with the target agent's environment, in contrast to existing methods against RL agents that assume that the adversary either has access to the target agent's learned parameters or to the environment [13]. In parallel to this, some researchers have highlighted that intruders are able to bypass the IDS model by constructing samples vulnerable to almost imperceptible perturbations of the inputs. To solve this, Wu et al. [31] built an RL framework able to generate adversarial traffic flows to deceive the detection.

2.1.3. Eavesdropping attack

An eavesdropping attack consists of a theft of information transmitted over a network and concerns 14.08% of the cases. It is also referred to as a sniffing attack or a snooping attack and may concern all connected devices such as a computer, a laptop or a cell phone. In [34], Xiao et al. develop a physical-layer anti-eavesdropping solution to diminish the capability of unapproved eavesdroppers to infer information in the context of visible light communication. Therefore, the authors exploit an RL-based control scheme to discover the theoretically optimal solution of the secrecy rate and, at the same time, define the most efficient beamforming policy against attackers.

In the field of internet of things (IoT), [35] discuss the challenge of using ML-based techniques, including RL, to protect user privacy (e.g., against eavesdropping attacks [7]). In this regard, the cooperation framework previously explained in [1] also aims to make decisions on eavesdropping attacks using a dedicated deep RL approach. In another area, to protect wireless networks, Xie and Xiao [36] apply prospect theory (theory based on the observation that people react differently to potential losses and potential gains - wikipedia) to formulate the interaction between a smart attacker and a mobile user. The first makes subjective decisions on the attack model and the second on the security mechanism layer to be applied. This allows the Nash equilibria of the static smart attack game to be derived and a defense strategy based on Q-learning to be proposed.

2.1.4. Spoofing attack

A spoofing attack consists of an attacker pretending to be someone else or something else in an attempt to gain the confidence of the defender. By spoofing a system, the attacker attempts to get access to defenders' systems, to steal data, or to spread malwares. This attack was the subject of consideration in 8.45% of the cases. The autopilot system of an autonomous or unmanned aerial vehicle is particularly sensitive to a spoofing attack given, for instance, the physical consequences that being hacked could imply.

In [2], Arthur identifies that drones need to identify their intruders and ensure their safe return-to-home and accordingly, he develops an RL-based adaptive IDS including a self-healing method enforced with a deep-Q network for dynamic route learning. Likewise, in [5], Dai et al. stress the fact that in vehicular ad hoc networks (VANETs), malicious on-board units (OBUs) may potentially try to gain illegal access to other OBUs. To face this situation, Dai et al. propose (i) an indirect reciprocity security framework to evaluate the OBU level of dangerousness to the VANET and (ii) an RL-based action selection strategy, which allow OBUs in the VANET to select a reliable relay OBU or determine whether or not to follow the request of another source OBU.

Bezzo [1] demonstrated the vulnerability of autonomous cyber-physical systems to attacks like sensor spoofing and used RL techniques to determine which sensors are compromised. Therefore, he proposes a reachability-based approach and a Bayesian Inverse RL technique [6] to leverage the history of sensor data and predict the attack [11].

2.1.5. (Distribute) Denial of Service attack ((D)DOS)

(D)DoS attacks concern 5.63% of the cases encountered. It is a typical attack in which the attacker tries to make the defender system services unavailable in order, mainly, to steal system information.
In [38], Zhang et al. analyze the resilience of cyber-physical systems to DoS and define, first, an RL method able to obtain the defense and attack policies at the cyber layer, and second, a dynamical programming method to obtain the physical layer control strategy and judge whether a system is capable of protecting the underlying control system.

Malialis et al. [16] propose Multiagent Router Throttling. This approach aims to defend the system against DDoS attacks and consists of a set of RL agents installed on multiple routers. The goal of these RL agents is to learn to rate-limit or throttle traffic towards a victim server. The particularity of this approach stays in the online learning process and in the incorporation of task decomposition, team rewards and a form of reward shaping.

2.1.6. Botnet attack

A botnet is a set of devices connected to the internet, compromised by an attacker, which act as a force multiplier to break into the defense system. Generally, botnets are performed in the context of distributed denial of service attacks, but their computing power may also be exploited (i) to send large volumes of spam, (ii) to steal large amounts of credentials, or (iii) to spy on persons and organizations. Botnet attacks are consequently addressed by the literature together with D(DoS) attacks. They represent 3.82% of the cases analyzed in this review. In [29], Venkatesan et al. observe the persistence of modern botnets when they operate in a stealthy manner over a long period of time. To reduce the lifetime of stealthy botnets and identify the maximum number of bots, the authors propose an RL-based solution to dynamically and optimally deploy a limited number of defensive mechanisms within the target network, including honeypots and network-based detectors.

2.1.7. Ransomware attack.

A ransomware attack consists of the attacker encrypting important business information stored on the victim's system, and to demand the payment of a ransom in exchange (i) for the data being decrypted and (ii) for the victim regaining access right [55] to the system. Hence, ransomware is often motivated by the gain of money usually transferred from the victim to the attacker by bitcoin. This type of attack only targets 2.41% of the cases. Existing ransomware detection approaches usually exploit machine learning, which needs large amounts of data to train the model, like the Domain Generational Algorithm (DGA), a method to quickly generate domains using a mathematical algorithm.

DGA has been considered by Cheng et al. [4] as a relevant technology for detecting ransomwares. However, given the difficulty of getting enough data to train specific models in a short period of time, Cheng et al. have developed a new DGA generation model based on RL and the Long Short Time Memory (LSTM) models. First, LSTM aims to provide the advantage of being able to generate a lot of new data learnt from a short set of real DGA samples and second, RL aims to guide the LSTM generation model to be enhanced by evaluating its newly generated domain name. This development aims to create a specific DGA trained with little data without leading to the over-fitting of the detection model. [15] in the field of communication and networking, or [26] which focuses on systems in general and proposes a method which consists of translating the system components and behavior into a multi-objective Markov process.

2.2. Recurrent network's contribution to attacks

Recurrent algorithms, by their nature, focus mainly on type attacks: phishing attack, domain generating algorithm, injection attack, and sybil attack while

2.2.1. Phishing attack

This attack consists of dissimulating oneself as a trustworthy person in an attempt to collect sensitive information from an end user. This attack represents 14.95% of the cases tackled with recurrent networks and consists of building four RNN models that only use URL's lexical features for identifying phishing attacks, like in [41]. In parallel, this research also proposes a set of visualization techniques to interpret the way RNN behaves internally to detect the phishing attempt.

2.2.2. Domain Generating Algorithm

A Domain Generating Algorithm (DGA) consists of a piece of code that provides malware with on-the-fly generated domain names. DGAs can be blocked using blacklists, but their coverage is widely deficient and inconsistent most of the time. In the light of artificial intelligence and machine learning, a DGA is examined as a classification issue [40]. DGAs represents 43,12% of the cases tackled with recurrent networks.

In Spaulding and Mohaisen [42], the DNS filtering system and system for network extraction (*FENS*) was developed and used CNN and LSTM for real-time classification together with blacklists. This system allows the prediction of malicious domain names that have never before been reported.

In [43], it is suggested that LSTM DGA are potentially able to learn of much new data from a small number of real samples and it considers the association of reinforcement learning and LSTM with the specific objective of detecting ransomware attack threats (consisting of the attacker encrypting important business information stored on the victim's system, and demanding the payment of a ransom in exchange for the data being decrypted). In a benchmarking model, [40] demonstrate that the classification of DGA has the best precision (with more than 96%) of a CNN-LSTM model in comparison to a simple CNN or LSTM model, and [63] go a step further and incorporate *attention mechanism* [45] into the LSTM model in order to tackle the problem of long domain expression. The contribution of the attention mechanism is to focus on more important substrings.

2.2.3. Injection attack

The injection attack consists of an attacker supplying untrusted data to a software program and an interpreter processing this data as part of a command or query, which in the end alters the software execution. It represents 22.52% of the cases tackled with recurrent networks. Two examples of RLG contributions that protect the system against injection attacks are observed in the literature.

First, Wang et al. [46], who propose an LSTM-RNN approach to predict temporal sequences in the field of Industrial Control Systems and second, Wang et al. [47], who propose a time-series state estimation method based on a deep RNN-LSTM network called *PGDL* (physics-guided deep learning). *PGDL* learns the temporal correlations among states by taking real-time measurements for inputs to the neural network, by outputting the new estimated states, and then by reconstructing measurements considering power system physics.

2.2.4. Sybil attack

In a Sybil attack, an attacker creates a large number of false identities and uses them to gain an overwhelming influence on the system and, *the facto*, on its reputation. This attack represents 19.41% of the cases tackled with recurrent networks. Gao et al. [48] propose a three-step neural network built upon, first, a CNN for extracting lower features from the multi-dimensional input, second a bidirectional self-normalizing LSTM network (bi-SN-LSTM) for extracting higher features from the feature map sequence generated, and third a classical dense layer and softmax classifier.

In Huang et al. [49], the authors tackle the protection of IoT against Sybil-like attacks and propose an automatic modulation classification method based on a densely connected LSTM network. The method first extracts features from cyclic correntropy vectors using the signals it receives and then, uses the extracted CCV feature as an input to the LSTM and the dense network.

3. Conclusions

In this paper, we have systematically reviewed reinforcement learning and recurrent networks-based security literature in order to analyze their current and future contributions to ubiquitous network security. This analysis has more particularly focused on the malware and intrusion detection and, on the contribution of each type of algorithms to various attacks. After applying filters to the most relevant databases, 351 papers have appeared relevant for scrutiny and the most important contributions have been presented in the paper. The most important findings are that RL and recurrent networks-based contributions to network security have been increasing exponentially for the last two years. We observe that the two types of algorithms analyzed contribute differently to the state of the art. First, recurrent networks contribute more (79.64%) to malware detection than reinforcement learning (20.36%). Second, the contribution is very specific to each of the algorithms. Recurrent algorithms, by their nature, focus mainly on type attacks: phishing attack, domain generating algorithm, injection attack, and sybil attack while reinforcement learning algorithms focus on jamming attack, adversarial attack, eavesdropping attack, spoofing attack, (D)DOS attack, botnet attack and ransomware attack. From the state of the art, we observe that most contributions agree that future research will mainly consist of improving the true positive rate in IDS, as well as in the precision of biometric recognition.

This paper is an extension of the paper presented in EUSPN 2020 [53]. In [52], the analysis of the contribution of reinforcement learning to the cybersecurity has been extended to various security domains (e.g. attacker-defender game, policy elaboration [61], biometric authentication...) associated to various technology (e.g. IoT, autonomous vehicles, critical infrastructures).

References

[1] Abuzainab N, Erpek T, Davaslioglu K, Sagduyu, YE, Shi Y, Mackey SJ, Patel M., Panettieri F, Qureshi MA, Isler, V, et al., 2019. Qos and jamming-aware wireless networking using deep reinforcement learning. arXiv preprint arXiv:1910.05766 . https://doi.org/10.1109/MILCOM47813.2019.9020985

[2] Arthur MP. Detecting signal spoofing and jamming attacks in uav networks using a lightweight ids, in: 2019 International Conference on Computer, Information and Telecommunication Systems (CITS), 2019. IEEE. pp. 1–5. https://doi.org/10.1109/CITS.2019.8862148

[3] Band I, Engelsman W, Feltus C, Paredes SG, and Diligens D. Modeling enterprise risk management and security with the archimate ®.Language, 2015. The Open Group.

[4] Cheng H., Fang Y, Chen L, and Cai J. Detecting domain generation algorithms based on reinforcement learning, in: 2019 CyberC. IEEE. pp. 261–264. https://doi.org/10.1109/CyberC.2019.00051

[5] Dai C, Xiao X, Ding Y, Xiao L, Tang Y, and Zhou S. Learning based security for vanet with blockchain, in: 2018 IEEE International Conference on Communication Systems (ICCS), 2018. IEEE. pp. 210–215. https://doi.org/10.1109/ICCS.2018.8689228

[6] Elnaggar M and Bezzo N. An irl approach for cyber-physical attack intention prediction and recovery ,in:2018 Annual American Control Conference (ACC), 2018. IEEE. pp. 222–227. https://doi.org/10.23919/ACC.2018.8430922

[7] Feltus C, Grandry E, Kupper T, and Colin JN. Model-driven approach for privacy management in business ecosystem., in: MODELSWARD, 2017, pp. 392–400. https://doi.org/10.5220/0006142203920400

[8] Feltus C, Khadraoui D, De Remont B, and Rifaut A. Business governance based policy regulation for security incident response. 2007. Crisis 7.

[9] Feltus C, Petit M, Dubois E. Strengthening employee's responsibility to enhance governance of it: Cobit raci chart case study, in: Proceedings of the first ACM workshop on Information security governance, 2009, pp. 23–32. https://doi.org/10.1145/1655168.1655174

[10] Ghosh P, Bardhan M, Chowdhury NR, Phadikar S, et al. Ids using reinforcement learning automata for preserving security in cloud environment. International Journal of Information System Modeling and Design (IJISMD), 2017, 8, 21–37. https://doi.org/10.4018/IJISMD.2017100102

[11] Grandry E, Feltus C, and Dubois E. Conceptual integration of enterprise architecture management and security risk management, in: 2013 17th IEEE International Enterprise Distributed Object Computing Conference Workshops, IEEE. pp. 114–123. https://doi.org/10.1109/EDOCW.2013.19

[12] Huang S, Papernot N, Goodfellow I, Duan Y, and Abbeel P, Adversarial attacks on neural network policies. arXiv preprint arXiv, 2017. :1702.02284 .

[13] Inkawhich M, Chen Y, and Li H. Snooping attacks on deep reinforcement learning, in: 19th International Conference on Autonomous Agents and MultiAgent Systems (AAMAS '20), 2020. p. 557–565.

[14] Lin YC, Hong ZW, Liao YH, Shih M., Liu MY, and Sun M. Tactics of adversarial attack on deep reinforcement learning agents. arXiv 2017 preprint arXiv:1703.06748. https://doi.org/10.24963/ijcai.2017/525

[15] Luong NC, Hoang DT, Gong S, Niyato D, Wang P, Liang YC, and Kim DI. Applications of deep reinforcement learning in communications and networking: A survey. 2019. IEEE Communications Surveys & Tutorials 21, 3133–3174. https://doi.org/10.1109/COMST.2019.2916583

[16] Malialis K, Devlin S, and Kudenko D. Distributed reinforcement learning for adaptive and robust network intrusion response. 2015. Connection Science 27, 234–252. https://doi.org/10.1080/09540091.2015.1031082

[17] Mayer N, Grandry E, Feltus C, and Goettelmann E. Towards the entri framework: security risk management enhanced by the use of enterprise architectures, in: International Conference on Advanced Information Systems Engineering, Springer. 2015. pp. 459–469. https://doi.org/10.1007/978-3-319-19243-7_42

[18] Melo FS. Convergence of q-learning: A simple proof. Institute Of Systems and Robotics, 2001. Tech. Rep , 1–4.

[19] Min M, Xiao L, Xu D, Huang L, and Peng M. Learning-based defense against malicious unmanned aerial vehicles, in: IEEE 87th Vehicular Technology Conference (VTC Spring), 2018. IEEE. pp. 1–5. https://doi.org/10.1109/VTCSpring.2018.8417685

[20] Navarro-Lara J, Deruyver A, and Parrend P. Morwilog: an aco-based system for outlining multi-step attacks, in: 2016 IEEE Symposium Series on Computational Intelligence (SSCI), IEEE. pp. 1–8. https://doi.org/10.1109/SSCI.2016.7849902

[21] Otoum S, Kantarci B, and Mouftah H. Empowering reinforcement learning on big sensed data for intrusion detection ,in: ICC2019-2019 IEEE International Conference on Communications (ICC), IEEE. pp. 1–7. https://doi.org/10.1109/ICC.2019.8761575

[22] Petersen K, Vakkalanka S, and Kuzniarz L. Guidelines for conducting systematic mapping studies in software engineering: An update. Information and Software Technology 2015 64, 1–18. https://doi.org/10.1016/j.infsof.2015.03.007

[23] Sheng G, Min M, Xiao L, and Liu S. Reinforcement learning-based control for unmanned aerial vehicles. Journal of Communications and Information Networks 2018. 3, 39–48. https://doi.org/10.1007/s41650-018-0029-y

[24] Sreekesh M., et al. A two-tier network based intrusion detection system architecture using machine learning approach, in: 2016 International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT), IEEE. pp. 42–47.

[25] Sundararajan K, and Woodard DL. Deep learning for biometrics: A survey. 2018. ACM Computing Surveys (CSUR) 51, 1–34. https://doi.org/10.1145/3190618

[26] Tozer B, Mazzuchi T, and Sarkani S. Optimizing attack surface and configuration diversity using multi-objective reinforcement learning, in: 2015 ieee 14th international conference on machine learning and applications (icmla),

IEEE. pp. 144–149. https://doi.org/10.1109/ICMLA.2015.144

[27] Tsochev G, Trifonov R, Yoshinov R, Manolov S, and Pavlova G. Improving the efficiency of idps by using hybrid methods from artificial intelligence, in: 2019 International Conference on Information Technologies (InfoTech), IEEE. pp. 1–4. https://doi.org/10.1109/InfoTech.2019.8860895

[28] Van Otterlo M, and Wiering M. Reinforcement learning and markov decision processes, in: Reinforcement Learning. Springer, 2012. pp. 3–42. https://doi.org/10.1007/978-3-642-27645-3_1

[29] Vekatesan S, Albanese M, Shah A, Ganesan R, and Jaodia S. Detecting stealthy botnets in a resource-constrained environment using reinforcement learning, in: Proceedings of the 2017 Workshop on Moving Target Defense, 2017. pp. 75–85. https://doi.org/10.1145/3140549.3140552

[30] Wang F, Zhong C, Gursoy MC, and Velipasalar S. Defense strategies against adversarial jamming attacks via deep reinforcement learning, in: 2020 54th Annual Conference on Information Sciences and Systems (CISS), IEEE. pp. 1–6. https://doi.org/10.1109/CISS48834.2020.1570629719

[31] Wu D, Fang B, Wang J, Liu Q, and Cui X. Evading machine learning botnet detection models via deep reinforcement learning, in: ICC 2019-2019 IEEE International Conference on Communications (ICC), IEEE. pp. 1–6. https://doi.org/10.1109/ICC.2019.8761337

[32] Xiao L, Jiang D, Xu D, Su W, An N, Wang D. Secure mobile crowdsensing based on deep learning. China Communications 2018. 15, 1–11. https://doi.org/10.1109/CC.2018.8485464

[33] Xiao L, Li Y, Huang X, Du X, 2017. Cloud-based malware detection game for mobile devices with offloading. IEEE Transactions on Mobile Computing 16, 2742–2750. https://doi.org/10.1109/TMC.2017.2687918

[34] Xiao L, Sheng G, Liu S, Dai H, Peng M, Song J. Deep reinforcement learning-enabled secure visible light communication against eavesdropping. IEEE Transactions on Communications2019. 67, 6994–7005. https://doi.org/10.1109/TCOMM.2019.2930247

[35] Xiao L, Wan X, Lu X, Zhang Y, Wu D. Iot security techniques based on machine learning: How do iot devices use ai to enhance security? IEEE Signal Processing Magazine 2018. 35, 41–49. https://doi.org/10.1109/MSP.2018.2825478

[36] Xie C, Xiao L. User-centric view of smart attacks in wireless networks, in: 2016 IEEE International Conference on Ubiquitous Wireless Broadband (ICUWB), IEEE. pp. 1–6. https://doi.org/10.1109/ICUWB.2016.7790439

[37] Xing R, Su Z, Zhang N, Peng Y, Pu H, Luo J. Trust-evaluation-based intrusion detection and reinforcement learning in autonomous driving. IEEE Network 2019. 33, 54–60. https://doi.org/10.1109/MNET.001.1800535

[38] Zhang P, Yuan Y, Wang Z, Sun C. A hierarchical game approach to the coupled resilient control of cps against denial-of-service attack, in: 2019 IEEE 15th International Conference on Control and Automation (ICCA), IEEE. pp. 15–20. https://doi.org/10.1109/ICCA.2019.8899933

[39] Zolotukhin M, Hämäläinen T. On artificial intelligent malware tolerant networking for iot, in: 2018 IEEE Conference on Network Function Virtualization and Software Defined Networks (NFV-SDN), IEEE. pp. 1–6. https://doi.org/10.1109/NFV-SDN.2018.8725767

[40] Pham TTT, Hoang VN, Ha TN, Exploring efficiency of character-level convolution neuron network and long short term memory on malicious url detection, in Proceedings of the 2018 VII ICNCC 2018. New York, NY, USA: Association for Computing Machinery, 2018, p. 82–86. https://doi.org/10.1145/3301326.3301336

[41] Feng T, Yue C, Visualizing and interpreting rnn models in url-based phishing detection, in Proceedings of the 25th ACM SACMAT '20. New York, NY, USA: Association for Computing Machinery, 2020, p. 13–24. https://doiorg.proxy.bnl.lu/10.1145/3381991.3395602 https://doi.org/10.1145/3381991.3395602

[42] Spaulding J, Mohaisen A, Defending internet of things against malicious domain names using d-fens, in 2018 IEEE/ACM Symposium on Edge Computing (SEC), Oct 2018, pp. 387–392. https://doi.org/10.1109/SEC.2018.00051

[43] Cheng H, Fang Y, Chen L, Cai J, Detecting domain generation algorithms based on reinforcement learning, in 2019 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC), Oct 2019, pp. 261–264. https://doi.org/10.1109/CyberC.2019.00051

[44] Shakshuki EM, Malik H, Sheltami T. WSN in cyber physical systems: Enhanced energy management routing approach using software agents. Future Generation Computer Systems. 2014 Feb 1;31:93-104. https://doi.org/10.1016/j.future.2013.03.001

[45] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, and Polosukhin I, Attention is all you need, Advances in neural information processing systems, vol. 30, pp. 5998–6008, 2017.

[46] Wang W, Xie Y, Ren L, Zhu X, Chang R, and Yin Q, Detection of data injection attack in industrial control system using long short term memory recurrent neural network, in 2018 13th IEEE Conference on Industrial Electronics and Applications (ICIEA), May 2018, pp. 2710– 2715. https://doi.org/10.1109/ICIEA.2018.8398169

[47] Wang L and Zhou Q, Physics-guided deep learning for time-series state estimation against false data injection attacks, in 2019 North American Power Symposium (NAPS), Oct 2019, pp. 1–6. https://doi.org/10.1109/NAPS46351.2019.9000305

[48] Gao T, Yang J, Peng W, Jiang L, Sun Y, and Li F, A content-based method for sybil detection in online social networks via deep learning, IEEE Access, vol. 8, pp. 38753–38766, 2020. https://doi.org/10.1109/ACCESS.2020.2975877

[49] Huang S, Lin C, Xu W, Gao Y, Feng Z, and Zhu F, Identification of active attacks in internet of things: Joint model-and data-driven automatic modulation classification approach, IEEE Internet of Things Journal, pp. 1–1, 2020.

[50] Melo FS. Convergence of Q-learning: A simple proof. Institute Of Systems and Robotics, Tech. Rep (2001), 1–4

[51] Li G, Zhou H., Feng B, Li G, and Yu S. (2018, December). Automatic selection of security service function chaining using reinforcement learning. In 2018 IEEE Globecom Workshops (GC Wkshps) (pp. 1-6). IEEE https://doi.org/10.1109/GLOCOMW.2018.8644122

[52] Feltus C. Reinforcement Learning's Contribution to the Cyber Security of Distributed Systems: Systematization of Knowledge. International Journal of Distributed Artificial Intelligence (IJDAI). 2020 Jul 1;12(2):35-55. https://doi.org/10.4018/IJDAI.2020070103

[53] Feltus C. Current and Future RL's Contribution to Emerging Network Security. Procedia Computer Science. 2020 Jan 1;177:516-21. https://doi.org/10.1016/j.procs.2020.10.071

[54] Nabi F, Yong J, Tao X. Classification of Logical Vulnerability Based on Group Attack Method. Journal of Ubiquitous Systems & Pervasive Networks.;14(1):19-26. https://doi.org/10.5383/JUSPN.14.01.004

[55] Feltus C, Dubois E, Proper E, Band I, Petit M. Enhancing the ArchiMate® standard with a responsibility modeling language for access rights management. Proceedings of the Fifth International Conference on Security of Information and Networks 2012 Oct 25 (pp. 12-19). https://doi.org/10.1145/2388576.2388577

[56] Pikulík T, Starchon P. Insight into PDP challenges of data transfer in wireless mobile devices. J. Ubiquitous Syst. Pervasive Networks. 2020;12(1):17-24. https://doi.org/10.5383/JUSPN.12.01.003

[57] Li S and Zhao D, A lstm-based method for comprehension and evaluation of network security situation, in 2019 18th IEEE International Conference TrustCom/13th IEEE International Conference BigDataSE, 2019, pp. 723– 728. https://doi.org/10.1109/TrustCom/BigDataSE.2019.00103

[58] Wang S, Xia C, and Wang T, A novel intrusion detector based on deep learning hybrid methods, in 2019 IEEE 5th Intl Conference on Big Data Security on Cloud (Big Data Security), IEEE Intl Conference HPSC and IEEE Intl Conference on IDS, May 2019, pp. 300– 305. https://doi.org/10.1109/BigDataSecurity-HPSC-IDS.2019.00062

[59] He H, Sun X, He H, Zhao G, He L, and Ren J, A novel multimodal sequential approach based on multi-view features for network intrusion detection, IEEE Access, vol. 7, pp. 183207–183221, 2019. https://doi.org/10.1109/ACCESS.2019.2959131

[60] Le T, Kim J, and Kim H, An effective intrusion detection classifier using long short-term memory with gradient descent optimization, in 2017 International Conference on Platform Technology and Service (PlatCon), Feb 2017, pp. 1–6. https://doi.org/10.1109/PlatCon.2017.7883684

[61] Feltus C. Preliminary Literature Review of Policy Engineering Methods; Toward Responsibility Concept. In2008 3rd International Conference on Information and Communication Technologies: From Theory to Applications 2008 Apr 7 (pp. 1-6). IEEE. https://doi.org/10.1109/ICTTA.2008.4529912

[62] Shakshuki EM, Kang N, Sheltami TR. EAACK—a secure intrusion-detection system for MANETs. IEEE Transactions on industrial electronics. 2012 Apr 26;60(3):1089-98. https://doi.org/10.1109/TIE.2012.2196010

[63] Chen Y, Zhang S, Liu J, and Li B, Towards a deep learning approach for detecting malicious domains, in 2018 IEEE International Conference on Smart Cloud (SmartCloud), Sep. 2018, pp. 190–195. https://doi.org/10.1109/SmartCloud.2018.00039

[65] Talha M, Elmarzouqi N, Abou El Kalam A. Quality and Security in Big Data: Challenges as opportunities to build a powerful wrap-up solution. J. Ubiquitous Syst. Pervasive Networks. 2020;12(1):9-15. https://doi.org/10.5383/JUSPN.12.01.002