Attacking and Defending with Intelligent Botnets

Moises Danziger and Marco Aurelio Amaral Henriques

Abstract—Machine learning (ML) has been seen as a great ally of security. All his potential to automate actions with some level of intelligence has called the attention of industry which is using it on security systems. However, attackers have also noted all ML potential. In a first moment, attackers have tried to fight ML-based security tools through the study and exploitation of weak points in ML techniques. It is named as adversarial machine learning. Besides this first application, someone could apply a ML-based tool directly against a security system. It is the case of intelligent bots - a different type of botnet made of relatively intelligent bots which can take decisions by their own during the attack. So, in this work we are making a reflection on the future of botnets within the context of ML and showing that this kind of botnet could break the current detection approaches. We also point out to the need for creating new approaches to combat bots with some kind of intelligence. Moreover, we propose a theoretical model of intelligent bots, their possible impacts and combat strategies.

Keywords—Botnets; Machine Learning; Autonomous systems, Intelligent Agents

I. INTRODUCTION

Botnets are one of the most feared virtual threats after the Internet advent. It brings several dangerous problems for users and systems. Nowadays most botnet attacks are successful due to their evasion abilities and, to improve defense techniques, researchers and the security industry presented a large number of detection methods as can be seen in [1]. Most of these works have presented some good results and some of them might be used in the real world. However, fighting botnets is a complex task and many approaches have been tried, but restricted to laboratory experiments [2].

At the same time, works as those of Bijalwan et al. [3] and Karim et al. [4] presented some challenges, trends and possible applications for botnets in the future. Currently we already can see botnets over IoT (Internet of Things) [5], smartphones (mobile botnets) [6] and cloud environments (botclouds) [7]. These approaches showed us some technological trends in botnet actions. Looking at such trends we can see that the botnet’s complexity has been growing as new technologies are appearing in the world. If we pay attention to the technology advances as, for example, cognitive computing [8] and deep learning [9], we can note that there are new possibilities in the botnet area. In this scenario, a new botnet characteristic can be the autonomy of bots. It means that bots could decide by themselves what must be done according to the context, targets and mission. It could be done through strong abilities to do reconnaissance on the environment and through an embedded intelligent process that helps the decision taking.

Moises Danziger and Marco Aurelio Amaral Henriques Faculty of Electrical Engineering and Computation, University of Campinas (UNICAMP), Campinas-SP, Brazil, E-mails: (danziger,marcoj@dca.leeunicamp.br

The first consequence of adding intelligence to botnets is the new type of operation that becomes possible. Now the bots can limit or break their communication with the botmaster or command centers all the time or, in the simplest scenario, they just communicate when the mission is accomplished. This can be disastrous for botnet detection because many detection approaches were created based on the command & control (C&C) communication process between the bots and the command centers or botmasters. If this process is eliminated or limited, the detection efficiency will probably decrease. But this new model is not so straightforward. It needs a deep change in present botnet concepts and the first actor to look at in this case is the botmaster. He will need new knowledge and skills.

In this work we present a discussion about the impact of attacks applying methods based on ML to botnets with bots that can take decisions on-the-fly. We also present a theoretical model based on multi-agent systems (MAS). Our objective is to anticipate the discussion about the impact of this new kind of botnet and the different modus operandi over the current detection systems.

We call attention to the fact that we are not discussing about the adversarial machine learning as showed by Huang et. al. [10]. Some discussions are showing the high impact of sophisticated adversaries that frequently attempt to break the training processes by, for example, crafting input data that has similar feature properties to normal data as seen in [11] and [12]. In these cases, the attackers are trying to evade ML-based classifiers systems by exploiting weak points. They are not applying ML to the attack, as we are discussing here.

Navigating through this paper we can see, in the second section, the current botnet research efforts and some results from the application of ML techniques in this area. Following, in section three we present a theoretical model of intelligent bots. Section four presents a discussion about the impacts of the model and possible ways to combat the new menace it imposes. Finally, in sections five and six we draw some conclusions and discuss possible future works.

II. STATE OF ART

Chen and Ji [13] presented an approach of a self-learning worm using importance scanning. It is well known that to spread worms, the developers employ distinct scanning methods as, for example, topological, random, localized, hitlist, routable and the importance scanning. The authors contend that in the Internet it may not be easy for attackers to collect information on vulnerable hosts. So, they pointed to future worms that can become more intelligent and potentially learn a certain knowledge about the vulnerable hosts while propagating. According to the article, the key capability of the worm
is to learn an underlying vulnerable host group distribution. So, the worm can collect information and estimate the group distribution. The results showed that the self-learning worm can spread far faster than a random-scanning, permutation-scanning, and a Class A routing worm.

Castiglione et al. [14] applied a swarm-based approach to improve botnet-based C&C infrastructures. They presented a methodology based on stigmergic communication models. A stigmergic system is a natural method in MAS that is based on collective agent’s behavior according to a distributed intelligence paradigm [15]. Working as a set of agents that act together with their operating environment, they are described at every instant by a set of state variables. Stigmergic (from stigmergy) is a form of self-organization that can produce complex and seemingly intelligent structures without prior planning, control or even direct communication between the agents [16]. This is a characteristic of social agents in an ant colony that was applied by Castiglione et al. in the building of a P2P-based C&C infrastructure. With it, the authors facilitate the interconnection among bots spread over the network. The results showed a C&C architecture more robust and scalable, a better collaboration and coordination environment for botmaster, improved fault tolerance and dynamic adaptation to varying network conditions. Their work gives an idea about the power of ML if applied to develop more efficient botnets. Yet, because of the stigmergy, the agents do not need to possess any previous memory, intelligence or coordination. Stigmergic cooperation significantly improves the infrastructure’s ability to dynamically adapt to changes in network conditions, greatly enhancing fault tolerance, robustness, scalability and survivability.

Gaudesi et al. [17] developed a new obfuscation mechanism based on evolutionary algorithms. Attackers use obfuscation to avoid reverse engineering in their malware/botnet code. They embedded an evolutionary core in the malware to generate a different, optimized hiding strategy for every single infection. It means that the malware program is able to evolve its own packer, creating a brand new encoding routine in each infection. The ML technique applied here is the genetic algorithm which is an approach to search and define the best values for the variables. Note that this kind of malware hiding technique can be a challenge for the security community.

III. A THEORETICAL MODEL OF INTELLIGENT BOTNET

Imagine a group of intelligent agents in a MAS [18] that were created with a mission. They have autonomy to decide how they will perform their actions to reach the target and accomplish their mission. This means that they do not need the C&Cs or botmasters monitoring their activities. The botmaster has a target and prepares a group of agents to spread on the internet, and their mission is to infect and steal sensitive information from a specific target. Bots do not have contact with the C&C after deployment.

After the agents are spread on the network/hosts, a learning process is started and they will use it to take decisions according to the challenges they face in their way. It means that agents/bots have the ability to learn from the events that they experience in the environment and can decide the better way to accomplish their mission. The whole decision process is done without the botmaster knowledge. For example, imagine an agent (bot) in front of a firewall. It can create a plan to bypass the firewall using a strategy based on team game from Game Theory [19]. In this case, the agent could create new agents to do specific tasks and create a set of agents to be sacrificed for the sake of a goal. Now we describe the main components of this model: The Agents, The Operation Model and The Learning Process.

A. The Agents

There are four kinds of agents in our proposal.

1) Super Agent (SA): It is the main agent that can be seen as a local botmaster. SA is responsible for the complex part of the bot actions. He must create and maintain the environment map and the surface attack. To avoid databases, he can use simple structures to store the data used to help in the decision process. He can also create other agents to help him to perform his mission. In addition, he can also create fake agents to distract a defending system or some specific agent to attack an enemy botnet. The communication that does not have a coordinator agent is his best choice to avoid him being discovered. For the decision process, he needs to have more ability to work on-the-fly. Note that the agent’s ability depends on the ability of the botnet creator in defining the best techniques to be embedded in the agent. For example, an agent can employ the bio-inspired optimization techniques, as for example, that implemented by [14], to discover the best way to do a coordinated attack with a set of agents. More details about the learning process of SA is given in the subsection about learning process.

2) Recon Agent (RA): This agent has two modules: communication and scanner. The first is similar to SA. The second means that this agent can do a lot of things as detecting services, operating system, open ports, defense systems and so on. It can also collect network information, location and other monitoring data. It is very important that this agent is able to do its activities secretly to avoid data exfiltration techniques. One of its most important tasks is updating the environment map.

3) Defense Agent (DA): During the actions, the botnet needs to act in a stealthy way to avoid being detected. It means that the botnet needs to follow some strategy to defend its perimeter. So, the DA is created to prepare the environment against attacks from other botnets and protects the perimeter to avoid defense systems. It can act alone or in group and can also help to aggregate information about the attack surface. It has two modules: communication and defender. The second module is used to detect any attack against the intelligent bots done, for example, by other botnets and/or worms. It is also responsible for the detection of any attempt of defending tools to capture bot core codes.

4) Attack Agent (AA): It is responsible for exploiting vulnerabilities, and opening the way to the botnet movement. Sometimes, depending on the situation, the attack agents could be created simply to fool the enemy.
B. The Operation Model

Figure 1 shows a macro vision of the proposal. This figure represents only one intelligent bot. Note that the attacker can spread the malware and infect a lot of devices and each one will be a new intelligent bot. In this first version the bots do not have any external communication. But, each bot is controled by the SA agent and it has an id that is used when it meets another intelligent bot during its lifetime (generaly in another already infected device). Once identified as a friend they can exchange knowledge. This is done using messages in a specific languages used by intelligent agents.

C. The Learning Process

Learning is essential for an intelligent agent and ML techniques using reinforcement learning (RL) seems to be suitable to the proposed model [21]. Differently from supervised and unsupervised learning, RL has an online learning process. It means that it learns and acts simultaneously and it seems to be a good choice in an intelligent bot because he can experience new scenarios and get new kinds of data that need retraining. In this case, we need to decide the basic attributes of the RL adopted to give bots ability to take decisions. Behind RL there is the well-known Markov Decision Process (MDP), which is a tool for modeling sequential decision-making problems where a decision maker interacts with a system in a sequential fashion [22]. We need to define the MDP attributes: (i) Non-empty set of states, (ii) non-empty set of actions, (iii) the transition probability kernel and (iv) a reward function. For example, the states can be the position in the network infected by the agent. The set of actions can move the bot from the current device to another in the neighborhood, the reward can be based on the number of captured devices in a period of time or the time spent to infect a device. Anyway, the learning module is the most important for intelligent bots.

IV. IMPACTS OF INTELLIGENT BOTNETS

We can follow the work of Stinson and Mitchell [23] and look at some botnet characteristics upon which automated detection methods rely to think about some impacts on detection systems. We can see that our proposed botnet model can impact detection methods that rely on traffic statistics from network communication between the C&C and the bots. The agents (bots) could calculate the statistics and act to...
avoid anomalies. Also, methods that rely on flow-charts can be impacted because there is no C&C in our proposal and the communication among agents can be done using a local-based network structure or through the ports 80 or 443 using HTTP. Methods that rely on syntax can also be impacted when cryptography is used in the agent’s communication. Our proposal could also impact the methods that rely on observing network communication. For example, it can be difficult to do information flow tracking in communication flow because there is no RP (rendezvous point).

Certainly, the communication can be the weak point of our model because a system could discover the agents. To mitigate it we can apply evasion techniques to avoid the defence system. For example, we can apply DGA to establish rendezvous points (it can be considered heavy for a simple agent but not for an expanded botnet with several intelligent bots working in a collaborative mode). So, for example, the communication model among agents can be done with P2P-based network techniques.

The concept of a self-learning worm from Chen and Ji [13] is very interesting for our approach. In fact, the idea of an intelligent worm is similar to an intelligent bot. However, the main difference between our approach and that is the ability of an agent to work in a collaborative way. In addition, the concepts of mobile bots and the super agent as a local botmaster are other important differences.

The intelligence behind ML techniques has a number of attractive features for military systems and they can call the attention of many governments. We also point out that this kind of botnet can have military focus because it could be an army of virtual soldiers. It can also be used for cyber espionage where the agents would be responsible for discovering sensitive information about, for example, countries, industry and arms.

Going to the bottom, we could extrapolate the concept of intelligent bots and think about botnets with autonomy. This is the next step of our research and brings more challenges. Yet, we can also think in hybrid botnets where there are intelligent bots and common bots. In this case the intelligent bots could be implemented to defend those common bots from being captured or stopped. Certainly, all of these conjectures are very interesting and we are planning simulations to show the autonomy concepts applied to botnets in several scenarios.

V. COMBATING INTELLIGENT BOTS

Probably, the most efficient way to combat an autonomous botnet is an approach that also uses some kind of intelligent agents. There are some MAS-based detection approaches as those presented by Pomorova et al. [24] and Savenko et al. [25]. In this way, we could create a MAS-based approach to show a possible combat method against intelligent botnets. Our work is different from that proposed by Kotenko et. al. in [26] because they do not consider intelligent bots nor detection models for them.

Another approach could be that from Salloum and Wolthusen [27] where the authors proposed a semi-intelligent link layer vulnerability discovery to operate in networks. They developed an agent-based (vulnerability) detection mechanism using semi-intelligent propagation strategies as the self-replication from worms. The idea is to reconstruct the topology information found through the link layer discovery protocol to detect neighboring nodes and propagate gradually until total coverage of an enterprise network is reached. No ML technique was implemented in that work and a possible proposal is adding it to improve the ability of those agents and force them to work without any kind of human interaction. This idea could be implemented to help agents to defend network perimeters.
It is possible to apply the nematodes idea from Dave Aitel [28]. Nematodes, good, beneficial or benevolent are names for a kind of worm created to combat other worms. Although the concept was used before by rival groups (attackers), it is a very controversial topic as showed by Salloum [29]. In [30] Salloum showed the Seawave, the first compute worm that utilize the second layer of the OSI model (Data Link Layer) as its main propagation way. This worm is a controlled interactive, self-replicating, self-propagating, and self-contained vulnerability mitigation mechanism. The author defined the taxonomy of viruses, worms and botnets. For each one, he point to a defensive model, so, we can have defensive viruses, worms and botnets. Maurushat [31] also did a benevolent worm design and a long discussion about the ethical and legal analysis of this approach to combat worms.

We noted that no learning process was added in those works with nematodes, so, we could apply intelligence on the nematode worms and give them more autonomy to fight against the intelligent bots. In this way, the intelligent bots would have a enemy in the same level of abilities.

VI. CONCLUSIONS
The botnet model discussed here is completely different from previous botnets because the C&C communication channel is eliminated or limited. In fact, because the intelligent bots have abilities to decide their next steps, the botmaster could not have online control over the botnet operation. As a result, the current botnet detection approaches that are based only on network analysis should be reviewed. We presented a theoretical model based on MAS that is used as a basic model to typify botnets with intelligent bots. According to our analysis, we can point to a dangerous evolution on botnets which can cause high impact on security systems. On the other hand, we call attention to the fact that for combating this new kind of botnet new tools will be necessary and intelligent techniques, as ML, can be a valuable asset on their development. Finally, we strongly recommend security researchers to review their security systems and applications and apply a threat modeling to evaluate the impact of attacks by tools and methods using ML. In the future, ML can be the best ally to fight more dangerous treats as intelligent botnets and we have to be prepared to better understand and use it.

REFERENCES