УДК 004.492.3

MULTI-AGENT BASED TECHNIQUE OF BOTNET DETECTION IN COMPUTER SYSTEMS

PhD O. S. Savenko, PhD S. M. Lysenko, A. F. Kryschuk

Khmelnytsky National University
Ukraine, Khmelnytsky
kism@beta.tup.km.ua, sirogyk@ukr.net, rtandrey@rambler.ru

A new botnet technique based on multi-agent systems with the use of fuzzy logic is proposed. The analysis of the botnets' actions demonstrations in the situation of the intentionally computer system reconnection with the use of fuzzy logic is performed. Fuzzy expert system for making conclusion of botnet presence degree in computer systems is developed. It takes into account the demonstration degree of reconnected computer system, demonstration degree of probably infected computer systems and demonstration degree of other computer systems available in the corporate area network that probably were not infected.

Keywords: botnet, Trojan, worm-virus, antivirus detection, multi-agent system, agent, sensor, fuzzy logic.

The analysis of malware development shows dynamic growth of its quantity. The most numerous and danger malware during the last years are Trojans and worm-viruses that spread and penetrate into computer system (CS) for the purpose of information plunder, anonymous access to network, DDoS attacks, spamming etc. Such techniques as signature-based, code emulators, encryption, statistical analysis, heuristic analysis and behavioral blocking are used in modern antiviruses for botnet detection [1] show the decreasing of its efficiency for new malware detection. The efficiency of new malware detection in recent years is decreasing [2]. One of the main reasons of the low efficiency of detection is the spreading of a new malware class botnet.

Bot-nets are the most serious cyber-threats today. They are the main base for such danger acts as distributed denial of service attacks, malware distribution, phishing, theft of confidential corporate data, organization of anonymous proxy servers etc. The peculiarity of botnet is the using of specialized commands and controlled channels of interaction that provides the updating of functional bots' parts of and actions features. The term Botnet denotes a network of compromised end hosts (bots) under the remote command of a botmaster. After botnet construction they are controlled autonomously and automatically. Sometimes they perform some illicit monetary activities [1—4].

That is why the actual problem of computer systems safety is a development of a new more perfect technique for new botnet detection. One of possible way to increase the detection efficiency is a developing of multi-agent system for new botnet detection in computer systems.

In order to increase the efficiency of botnet detection the multi-agent system that allows us to make antivirus diagnosis via agents' communication within corporate area network was offered [6]. It uses the set of agents. Each agent implements antivirus diagnosis via a set of sensors $A = \langle S_1, S_2, S_3, S_4, S_5, S_6 \rangle$, where S_1 — agent sensor of signature-based analysis; S_2 — checksum sensor; S_3 — sensor of heuristics analysis; S_4 — behavioral analysis; S_5 — sensor of comparative analysis through application programming interface API and driver disk subsystem via IOS; S_6 — sensor — "virtual bait". Also agent includes a set of effectors that effect the computer system with purpose of blocking suspicious programs and then notify the other agents in the network about the infection in order to launch the suspicious programs detection with similar behavior. Agent has the CPU which processes the input data and determines the level of risk of specified object in the computer system based on some knowledge. In situation when agent cannot communicate with other agent it is as autonomous unit and is able to detect different malware relying on knowledge of the latest updates and corrections in the trusted software base.

The main disadvantage of this system is the decrease of the efficiency of antivirus detection by the recent period. Efficiency of detection is 67% (January, 2013) versus 70% (February, 2012). Other problem is the comparatively high level of the false detection which is about 7—10% (January, 2013) versus 3—7% (February, 2012).

To overcome mentioned problems the new techniques and methods are to be developed for the high efficiency botnet detection based on proposed multi-agent antivirus system.

The first step of the botnet detection is to construct a schematic map of connections which is formed by corresponding records in each antiviral agent of multi-agent systems for some corporate area network. All agents based on this information can perform communicative exchange data to each other. Botnet detection process can be presented as a scheme shown in Fig 1.

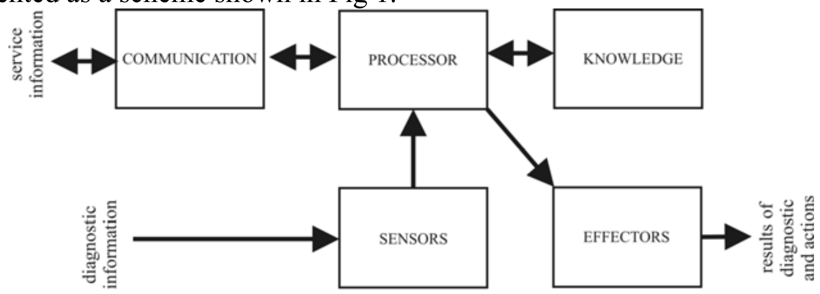

Fig. 1. The scheme of antiviral agent of multi-agent system operation

In order to overcome the problem of reducing the reliability of new botnet detection, a new method for determining the degree of presence of botnet is proposed. The offered method is based on analyzing the bots actions demonstration in situations of intentional change of connection. This approach is performed in the case of insufficient (low) values of suspicious software, but this suspicion is present in a definite amount of computer systems of the corporate area network.

During computer system functioning the antivirus detection via sensors available in each agent is performed. The antivirus diagnosis results are analyzed in order to define which of the sensors have triggered and what suspicion degree it has produced. If triggering sensors are signature S_1 or checksum S_2 analyzers, the results R_{S1} and R_{S2} are interpreted as a 100% malware detection. In this situation, the blocking of software implementation and its subsequent removal are performed.

For situations when the sensors of heuristic S_3 and behavioral S_4 analyzers have triggered, the suspicion degrees R_{S3} and R_{S4} are analyzed, and in the case of overcoming of the defined certain threshold n, $n \le \max(R_{S3}, R_{S4}) \le 100$, the blocking of software implementation and its subsequent removal are performed. If the specified threshold hasn't been overcome, the results R_{S3} , R_{S4} are analyzed whether they belong to the range $m \le \max(R_{S3}, R_{S4}) < n$, in order to make the final decision about malware presence in CS. If the value is $\max(R_{S3}, R_{S4}) > m$ then the new antivirus results from sensors are expected. In all cases the antiviral agents information of infection or suspicion software behavior in CS must be sent out to other agents.

The important point of this approach is to research the situation where the results of antivirus diagnosis belong to the range $m \le \max(R_{S3}, R_{S4}) < n$. In this case, the antiviral agent of CS asks other agents in the corporate area network about the similarity of suspicion behavior of some software that is similar to the botnet. If the interrogated agent receives information from one or more agents about the similar of software suspicious behavior. Then the probably infected computer systems are marked and map reconstruction is implemented (Figure 2). From the set of "marked» computer systems some CS is must be chosen for the changing of network connection type (reconnection) - specific network settings prevent the network functioning of the bot in the computer system (DNS change, non-standard port connection to network, etc).

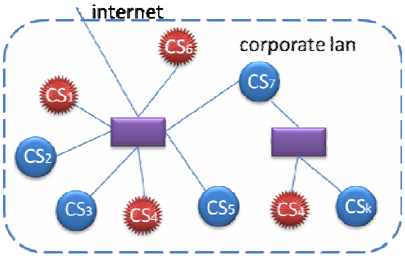

Fig. 2. Marked computer systems in the corporate area network

The means of choosing the one computer system from the “marked” is the expert system. It contains a set of rules that are present in the knowledge of each antiviral agent. This CS must meet the defined criteria.

After the reconnection of the chosen CS, the analysis of botnet demonstrations on reconnected computer system, on “marked” computer systems and other computer systems of the corporate area network and the definition of the degree of a new botnet presence in the network must be determined.

The presence of botnet in the corporate area network is concluded by the fuzzy expert system that confirms or disproves this fact. The determining of the botnet presence degree in computer system in situation of changed connection is shown in the following algorithm:

```
for i=1 to k of CS_i do
   while CS_i is_on do
      if $R_{S1} = true \cap R_{S2} = true$ then block and delete malware;
        else if $R_{S3} = true \cap R_{S4} = true$ and $n \le \max(R_{S3}, R_{S4}) \le 100$ then block and delete malware;
               else if $R_{S3} = true \cap R_{S4} = true$ and $m \le \max(R_{S3}, R_{S4}) < n$ then communicate with
               other agents;
               choose suitable CS to reconnect;
               analyze the degree of botnet demonstration in corporate area network
                     else if $R < m$ wait for results $R_{S1}, R_{S2}, R_{S3}, R_{S4}$.
```

Choosing the computer system to change its type connection in corporate area network

Determination of the presence of botnet network is possible due to the fact that when we change the type of connection of some computer system, bots can demonstrate itself in some way (bots can try to communicate with other elements of botnet, update lists of active bots, reconfigure itself taking to account the new lists, etc.).

It is important to pay attention to the place of the computer system in the topology of the corporate area network. If the computer system is a unifying node with neighboring computer systems in corporate area network (e.g., CS_7 in Fig. 3), which can be a server or a firewall, we are cannot not change the type connection of this computer system.

Each agent of probably infected CS calculates the rate of its «suitability» and then communicates with other agents in order to choose CS as the most «suitable» one for the changing the type of network connection. In order to choose some CS we must analyze the features and properties of probably infected computer systems with botnet. For this purpose let take the concept of «suitability» of some computer system. Thus, we are interested in the computer system with the most relevant antivirus databases, with the highest uptime duration, with the lowest vulnerability degree of the operating system and the best result of virus diagnosis. Determination of computer system «suitability» is performed with the use of a fuzzy inference system which is present in the agent structure. Each agent of probably infected CS calculates the rate of its «suitability» and then communicates with other agents in order to choose CS as the most «suitable» one for the changing the type of network connection.

The analysis of botnet demonstrations and the conclusion about computer system infection

For the determination of the presence degree of botnet in CS we must analyze botnet's demonstrations when some CS was reconnected. For this purpose all demonstrations are divided into three categories and the degrees, each of them must be determined: demonstration degree of reconnected CS, demonstration degree of probably infected computer systems and demonstration degree of other computer systems belonging to the corporate area network that probably weren't infected. To determine the possibility of the botnet presence in CS, the estimation of the demonstration degree for each of the three categories is performed. Demonstrations' degrees of three categories are presented as the fuzzy linguistic variables "demonstration degree" with three terms ("low", "medium" and "high").

The task of determination of membership function for input variable «demonstration degree» of reconnected computer we will consider as the task of the ranking for each of functions of penetration ports with the set of indications of danger and a choice of the most possible with activation of some function.

The task of determination of membership function for input variables «demonstration degree» of «marked» computers and common (not infected) computer systems are considered as the calculating the botnet demonstration degree. We must take into account the botnet action danger, the number of computer systems and where the demonstrations took place.

Let accept ω_j^i, $0 \le \omega_j^i \le 1$ - one of the signs of the demonstration, $j = \overline{1,n}$, $i = \overline{1,\gamma}$, where γ – number of botnet demonstration, k – number of computer systems in corporate area network. The estimation of each CS can be performed with the use of formula $\omega^1 = \sum_{i=1}^{q} \alpha_i^1 \omega_i^1 / \gamma$, $\omega^1 = \sum_{i=1}^{q} \alpha_i^2 \omega_i^1 / \gamma$, … , $\omega^j = \sum_{i=1}^{q} \alpha_i^{\Sigma} \omega_i^j / \gamma$, where - α_i - coefficients of the danger of some demonstration, $\alpha_1 + \alpha_2 + ... + \alpha_q = 1$, $0 \le \omega^i \le 1$.

Thus, if we choose some threshold value for each computer system with the estimation ω^j, for example $\tau \in (0;1]$, then we can select some group g of ‘suspicious’ computer systems if $\omega^j > \tau$. Then we calculate d_i - number of nonzero demonstrations of d_i^j in each computer system and average value ω_i with nonzero demonstrations ω_i^j. If number of nonzero demonstrations $d_i \neq 0$ then number of nonzero demonstrations is calculated with the use of formula $\omega_i = \sum_{j=1}^{n} \omega_j^i / d_i$, $d = \sum_{i=1}^{q} d_i \leq \gamma \cdot k$. We have to normalize the number $\omega_i, i = \overline{1,\gamma}$, so that $\omega_1 + \omega_2 + ... + \omega_\gamma = 1$. Then general demonstration degree of botnet presence in “marked” computer systems is $P_d(d_1, d_2, ..., d_\gamma) = \frac{d!}{d_1! d_2! .. d_\gamma!} \cdot \omega_1^{d_1} \cdot \omega_2^{d_2} \cdot ... \cdot \omega_\gamma^{d_\gamma}$. Let k', $k' \leq k$ – number of “marked” as infected computer systems. Then the arithmetic middling $\overline{\omega}$ of its correspondent ω^j must be calculated. After that the number P_d is determined and is interpreted as degree of botnet demonstration in «marked» SCs.

The resulting conclusion of botnet presence degree in computer systems is performed by fuzzy inference system based on Mamdani algorithm. It operates on determined demonstration degrees for three categories of computer systems (reconnected, «marked», and other computer systems of the network).

The new botnet technique based on multi-agent system with the use of fuzzy logic is proposed.

The detection is performed in the situations of a priori uncertainty of the botnet presence in the corporate area network with taking into account the botnet demonstrations in the several computer systems available in the network. With the use of fuzzy logic, the analysis of the botnets actions demonstrations in the situation of the intentionally computer system reconnection is performed

Fuzzy expert system for making conclusion about botnet presence degree in computer systems is developed. Fuzzy expert system takes into account the demonstration degree of reconnected computer system, demonstration degree of probably infected computer systems and demonstration degree of other computer systems available in the corporate area network that probably weren’t infected. The consistency of agents in order to improve the efficiency of botnet detection is the direction of the further research.

References

1. Buxbaum P. Battling botnets // Military Information Technology (MIT).— 2008.— Vol. 12.

2. Zhaosheng Z., Guohan L., Yan C. et al. Botnet research survey // Proceed. Of the 32nd Annual IEEE International Conference on Computer Software and Applications.— 2008.— P. 967—972.

3. Livadas C., Walsh R., Lapsley D., Strayer W. T. Using machine learning techniques to identify botnet raffic // Proceedings of the 2nd IEEE LCN Workshop on Network Security (WoNS'2006).— P. 967—974.

4. Lee W., Wang C., Dagon D. A taxonomy of botnet structures. Botnet Detection Countering the Largest Security Threat.— Springer Science, LLC, 2008.— P. 143—164.

5. Stern H. A survey of modern spam tools // Proceedings of the 5th Conference on Email and Anti-Spam (CEAS).— Mountain View, CA.— 2008.

6. Savenko O., Lysenko S., Kryschuk A. Multi-agent based approach of botnet detection in computer systems // Proceedings of the 19th 19th International Conference, CN 2012.— Poland, Szczyrk.— 2012.— P. 171—180.

7. Cristian Florian. The most vulnerable operating systems and applications in 2011. http://www.gfi.com (2012)

О. С. Савенко, С. М. Лысенко, А. Ф. Крыщук

Технологии обнаружения ботнетов в компьютерных системах на основе мультиагентной системы.

Разработан новый метод выявления ботнет сетей на основе мультиагентной системы с использованием нечеткой логики. Выполняется анализ проявления активности ботнет сетей в ситуации изменения типа подключения компьютерной системы с использованием нечеткой логики. Разработана нечеткая экспертная система для принятия окончательного вывода о степени присутствия ботнет сети в компьютерных системах. Она учитывает степень активности в переподключенных компьютерных системах, которые могли быть инфицированы, и других компьютерных системах корпоративной сети, которые, вероятно, не были инфицированы.

Ключевые слова: *ботнет сети, троянские программы, обнаружение вирусов, мультиагентая система, агент, датчик, нечеткая логика.*