ELECTRONICS AND ELECTRICAL ENGINEERING
ISSN 1392 – 1215 *2009. No. 2(90)*
ELEKTRONIKA IR ELEKTROTECHNIKA

ELECTRONICS
T170
ELEKTRONIKA

Network Scan Detection Simulation

N. Paulauskas, E. Garšva
Department of Computer Engineering, Vilnius Gediminas Technical University,
Naugarduko str. 41, LT-03227, phone: +370 5 2744767; e-mail: nerijus.paulauskas@el.vgtu.lt
J. Skudutis
Department of Electronic Systems, Vilnius Gediminas Technical University,
Naugarduko str. 41, LT-03227, phone: +370 5 2744755; e-mail: julius.skudutis@el.vgtu.lt

Introduction

With the increase of communications, economy, industry and business dependence on the information technologies, the risk related to the pervasive intrusions in the electronic space is also increasing. Malicious intruders more frequently overcome protection systems installed in banks or companies intended to restrict the access to the computer network resources of the organization. Seeking to reduce the risk and possible consequences it is very important to identify the intrusions at the initial stage of their realization and to react to them properly [1, 2, 3].

One of the most widely spread network incident forms is the network scanning, which is used by the attacker for the configuration determination of the target network. For instance, the attacker can be interested in the active network hosts (servers, operating computers, etc.) and their services (web, e-mail, file sharing, etc.).

The first intrusion detection system, which realized a simple scan detection method when a particular flags value was directly compared with the threshold value, was *Network Security Monitor* (NSM). In the open source network Intrusion Detection System *Snort*, especially for the port scan detection, the *sfPortscan* preprocessor is used. Leckie and Kotagiri [4] applied a probabilistic model for the recognition of a normal user from the scanning sources. The model is based on the assumption that the active network hosts are accessed more frequently, while queries for not existing network hosts are rather rare. Jung et al. [5] proposed a *Sequential Hypothesis Testing* method for the scan detection. Its main idea is that the ratio of successful connections of good and scanning sources differs.

The shortcoming of the mentioned methods is that they require that the changing network infrastructure and the intended data stream should be evaluated. For the signature-based detection methods a constant update of data about new intrusions is needed. A general shortcoming of known intrusion detection systems is that before implementation of the system it is necessary to perform a lot of initial settings (e.g. configuration of various sets of rules), which most frequently have to be chosen based on subjective assessment criteria because they are unknown without additional investigations.

The aim of this work is to develop an intrusion detection model which allows operative investigation of various scanning types (horizontal, vertical, block), the influence of parameters of the simulated computer system on scan detection efficiency and proper configuration of the intrusion detection system.

Model description

The simulated computer system consists of the N-sized addressed computer subnetwork, which involves n computers and m open ports. The number of ports interesting to the malicious user is M. The number of external users who reach the computer system is L, out of whom l are authorized.

The simulated computer system is described by the set of server destination addresses D_n and open ports Q_m:

$$SYS = \{D_n, Q_m\} = \{(d_i, q_j), \quad i = 1,\ldots,n, \quad j = 1,\ldots,m\}. \quad (1)$$

The set of active server addresses is $D_n=\{d_1,\ldots,d_n\}$, the whole set of possible target addresses is $D_N=\{d_1,\ldots,d_N\}$, where $D_n \subset D_N$. It is likely that the number of used addresses is lower than the number of all possible addresses in the subnetwork: $|D_n|<|D_N|$.

Open ports in the computer system are determined by the set $Q_m=\{q_1,\ldots,q_m\}$, and ports which can be scanned by the attacker in the whole computer system are determined by the set $Q_M=\{q_1,\ldots,q_M\}$, where $Q_m \subset Q_M$.

The analyzed computer system can be correctly used by a certain number of users l, who are identified in the model by the source addresses $s_1,\ldots,s_l$, and the probability that the packet comes from the authorized user is P_g. The computer system is scanned by L-l scanners, who are identified by the source addresses $s_{l+1},\ldots,s_L$, and the scanning packet appearance probability is P_b. The set of the source addresses reaching the whole computer system

is: $S_L=S_g\cup S_b$, where $S_g=\{s_1,\ldots,s_l\}$ is the set of authorized users, $S_b=\{s_{l+1},\ldots,s_L\}$ is the set of malicious users. It is probable that the number of users correctly using the computer system will be significantly larger than the number of malicious users scanning the system (L-l), therefore the probability that the computer system will be used by its users will be higher than the probability of malicious scanning $P_g>P_b$, where $P_g+P_b=1$.

The computer system model consists of the packet generator and the scanning detection system. In the packet generator output the packet sequence is obtained, part of which consists of scanning packets. The scanning detection system is connected so that the whole network packet flow can be monitored. It determines the number of unique scanners and the number of scanning cases.

The generated network packets of the computer system users and its scanners are described by the source address s_i, the destination address d_j and the destination port q_k. The set of all possible packets in the model is:

$$B=\{(s_i,d_j,q_k),\ 1\le i\le L,\ 1\le j\le N,\ 1\le k\le M\ \}. \qquad (2)$$

Packets are generated at the time Δt_{gen}, which is distributed according to the random stochastic law (e.g., Poisson law). During the whole simulation time Δt a certain number of packets is generated. A separate packet $b_z=(s_{z,i},d_{z,j},q_{z,k})$, where z is the sequence number of network packets in the model, which is common both to the correct use of the network and scanning packets. In the network packets the source addresses, destination addresses and destination ports can recur.

The generated packet can be of correct or malicious use of the network. The corresponding probabilities of appearance of such packets are as follows:

$$P\{b_z\in B_g\}=P_g;\quad P\{b_z\in B_b\}=P_b, \qquad (2)$$

where B_g is the set of good packets ($B_g\subset B$), B_b is the set of malicious packets ($B_b\subset B$), $B=B_g\cup B_b$.

In the first generated packet b_1 of the model the source address $s_{1,i}$ is generated from the known computer system authorized user source addresses set S_g or the scanning malicious source addresses set S_b based on the following requirements for event probabilities:

$$P\{s_{1,i}\in S_g\}=P_g;\ P\{s_{1,i}\in S_b\}=P_b, \qquad (3)$$

i.e., the probability that in the first packet the source address will be of the authorized user is P_g, and the probability that the first packet will be sent by the scanner is P_b, where $S_L=S_g\cup S_b$ ir $P_g+P_b=1$.

For the second and all other packets, the probabilities of good and malicious packet event appearance remain the same, but there appear additional conditions of malicious packet source address generation: the probability P_s that another generated packet will be sent by the same scanner (source address remains the same) is higher than the probability that the packet will be sent by another scanner (1-P_s), i.e. the probability that in another scanning packet the source address will change is lower.

Thus, the probability of event appearance that the second packet will be good does not change

$$P\{s_{2,i}\in S_g\}=P_g, \qquad (4)$$

while event probabilities that the second packet will be bad and it will be sent by the same scanner are equal, respectively:

$$P\{s_{2,i}\in S_b\}=P_b \text{ ir } P\{s_{2,i}=s_{1,i}\mid s_{1,i}\in S_b\ \}=P_s. \qquad (5)$$

Based on these conditions in all other packets:

$$P\{s_{z,i}\in S_g\}=P_g, \qquad (6)$$

$$P\{s_{z,i}\in S_b\}=P_b \text{ ir } P\{s_{z,i}=s_{z-1,i}\mid s_{z-1,i}\in S_b\ \}=P_s. \qquad (7)$$

Network packets of authorized users are addressed to servers and their open ports. If the packet is of malicious scanning (with probability P_b), it can be either of horizontal b_h or vertical b_v scanning. The respective probabilities of appearance of such events are:

$$P\{b_v\in b_z\}=P_v;\ P\{b_h\in b_z\}=P_h, \qquad (8)$$

where P_v is the probability of appearance of vertical scanning packets, P_h is the probability of appearance of horizontal scanning packets, $P_v+P_h=1$.

In case of horizontal scanning, the destination address d_j is from the set of all the subnetwork D_N, and the probability of access to the same port P_p is much higher than that of access to another one (1-P_p). In case of vertical scanning the port q_k will be from the set of all ports interesting to the scanner Q_M, and the probability of access to another computer of the system (1-P_d) will be much lower than to the same computer P_d. Taking into account this assumption we can state that vertical and horizontal scanning packets are generated according to these conditions:

$$b_v=\left(s_i\in S_b, P\left\{d_{z,j}=d_{z-1,j}\mid d_j\in D_N\right\}=P_d, q_k\in Q_M\right);$$

$$b_h=\left(s_i\in S_b, d_j\in D_N, P\left\{q_{z,k}=q_{z-1,k}\mid q_k\in Q_M\right\}=P_q\right). \qquad (9)$$

Data about source addresses from which it was accessed to the computer system destination addresses and ports are collected and stored for the time interval Δt_{rst}. If during this time interval the number of queries exceeds the threshold value, the scan is detected. The scanning detection system detects the number of horizontal scans C_{hscan}, the number of vertical scans C_{vscan}, and the number of malicious scanners $C_{scanners}$. The scan detection system checks accumulated data during time intervals Δt_{IDS}, where $\Delta t_{IDS}<\Delta t_{rst}$. When $\Delta t_{IDS}\to 0$, the scan detection approaches to the real time. Time Δt_{IDS} was introduced in order to reduce the simulation time, because having introduced this time during the time interval Δt_{rst} accumulated data are checked more rarely by saving the computer resources used for simulation. If the number of accesses during the determined time Δt_{rst} exceeds the threshold value W_h, a horizontal scan is detected. The number of horizontal scans C_{hscan} is:

$$C_{hscan}=\sum_{i=1}^{L}\left(\sum_{d_j\in D_N}\left(s_i,d_j\right)\ge W_{h,}\right). \qquad (10)$$

If the number of accesses during the determined time Δt_{rst} exceeds the threshold value W_v, the vertical scan is detected. The number of vertical scans C_{vscan} is:

$$C_{vscan} = \sum_{i=1}^{L} \left(\sum_{q_j \in Q_M} (s_i, q_j) \geq W_v, \right). \quad (11)$$

Here W_v is the threshold value of vertical scans.

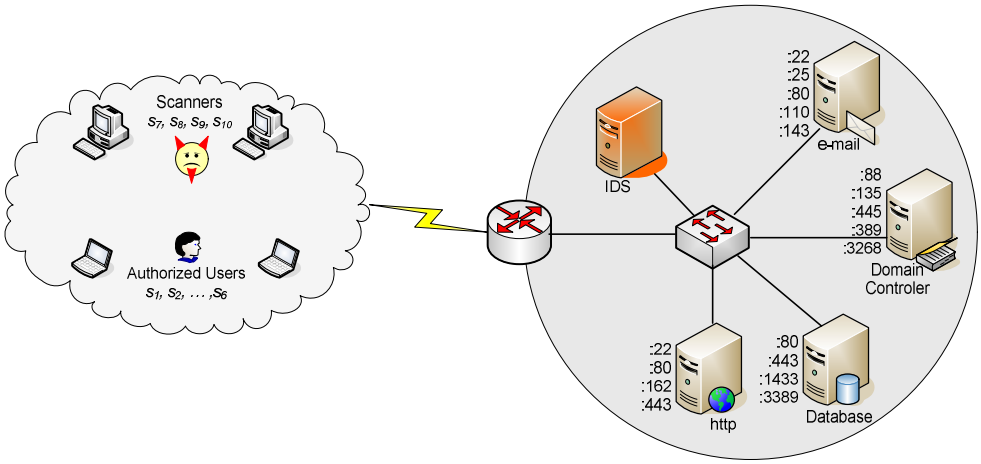

Fig. 1. Simulated computer system

The number of scanners $C_{scanners}$ is determined by adding up all source addresses, from which the scan was done. Let us denote them as s_i', then:

$$C_{scanners} = \sum_{i=1}^{L} s_i' \, . \quad (12)$$

Case study

The model of the scan detection system was developed using stochastic activity networks (SANs). A model of the horizontal scan of a small business company computer network is shown in Fig. 1.

The model consists of a packet generator and a scan detection system (detector). The scan detection system is connected so that it could see the whole network packet flow and could determine the number of scan cases and malicious users scanning the network.

The initial parameters of the simulated system are as follows: N=14, n=4, L=10, l=4, m=6, M=18, P_g = 0.99, P_b = 0.01, P_s=0.8, P_p=0.8, Δt_{gen}=1000. As the scanner has no initial data about the scanned network, the probabilities of access to each network host are equal.

In order to obtain more reliable simulation results, the model was calibrated based on the following assumptions:

- if all generated packets are the result of correct use of the network (P_g = 1), then the number of detected vertical and horizontal scans should not be higher than 2 (C_{hscan}, $C_{vscan} \leq 2$);
- if all generated packets are malicious packets of the horizontal scan (P_b = 1), then the number of detected vertical and horizontal scans should be higher than or equal to the number of possible scanners (C_{hscan}, $C_{vscan} \geq L\text{-}l$).

Based on these assumptions the intervals of possible values of parameters used for the scan detection were determined: the threshold value W_h=[5, 6], the scan detection interval Δt_{IDS}=[0.2, 2.5], the detector reset interval Δt_{rst}=[1, 5]. The scan detection interval Δt_{IDS} and the reset time Δt_{rst} do not have influence on the scan detection when all generated network packets are of authorized users (P_g=1) and the number of horizontal and vertical scans is equal 0.

Dependences of the number of horizontal scans C_{hscan} and the number of malicious scanners $C_{scanners}$ on the threshold value W_h are presented in Fig. 2. When the threshold value is equal to 2 (it means that all IP addresses from which two or more network hosts were

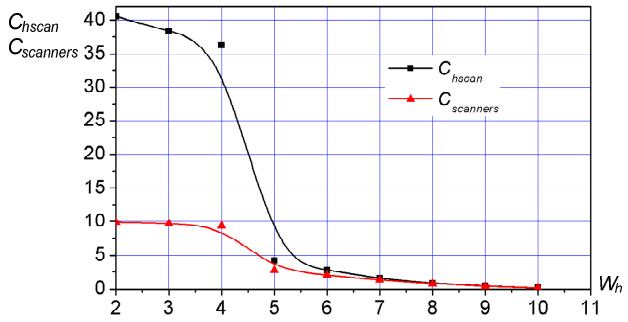

Fig. 2. Dependences of the number of horizontal scans C_{hscan} and the number of malicious scanners $C_{scanners}$ on the threshold value W_h

accessed are detected), the number of scans and scanners is the highest. This is due to the fact that at this threshold value, good network users are mistaken for scanners and false positive detection occurs. A further increase of the threshold value lowers the number of scans and scanners and both curves reach the breaking point. The breaking point appears at the threshold value at which good network users are not recognized as scanners. Since the model foresees that good network users can connect only to the services providing computers, which in this case are 4, at the threshold value equal to 5 they are not recognized as scanners. By adjusting the detector and seeking to reduce the number of false positives, the threshold value higher than the number of active computers in the network should

be chosen or the access to active ports of these computers should not be fixed when detecting scans. It can be seen from the diagram that a further increase of the threshold value reduces the number of detected scans and scanners. The reason is that when the threshold value is too high the attempts of scanning are not detected. This is called false negative.

The dependence of horizontal scans C_{hscan} and scanners $C_{scanners}$ on Δt_{rst}, when P_g=0,99, is presented in Fig. 3.

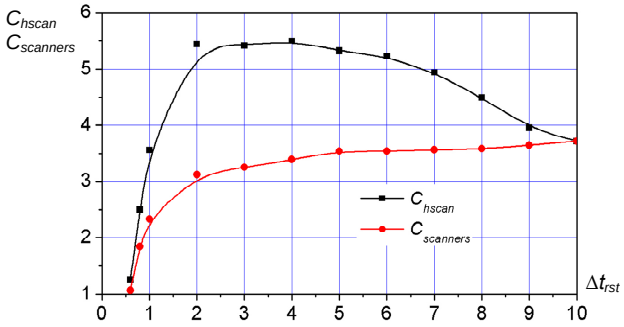

Fig. 3. Dependence of the number of scans C_{hscan} and scanners $C_{scanners}$ on Δt_{rst}, when P_g=0,99

By making the time interval longer the number of scans increases faster because the same scanner can scan more than once. By further increasing the time interval, the number of scans decreases. The number of scanners and scans becomes equal when the duration of the reset interval becomes equal to the simulation duration. The reason is that the detector system having detected the scanner fixes its IP address and further scans from the same IP address are not registered until the detector counters are reset.

Conclusions

The model of the scan detection system consisting of the packet generator and the detector allowing operative evaluation of the influence of parameters of the simulated computer system on the efficiency of horizontal and vertical scan detection has been developed.

The threshold value W and the detector reset interval Δt_{rst} are the main parameters having influence on the efficiency of the scan detection.

Scanning is detected most effectively when the threshold value W is higher than the number of hosts in the network, which are accessed most frequently (higher than the number of active servers).

References

1. **Garšva E., Skudutis J.** Secure Computer System Design // Electronics and Electrical Engineering. – 2004. – No. 6 (55). – P. 43–48.
2. **Paulauskas N., Skudutis J.** Investigation of the Intrusion Detection System “Snort” Performance // Electronics and Electrical Engineering. – 2008. – No. 7 (87). – P. 15–18.
3. **Garšva E.** Computer System Survivability Modelling by Using Stochastic Activity Network // Proc. SAFECOMP’06. – Springer-Verlag. – 2006. – P. 71–78.
4. **Leckie C., Kotagiri R.** A Probabilistic Approach to Detecting Network Scans // Proc. of the Eighth IEEE NOMS2002. – 2002. – P. 359–372.
5. **Jung J., Paxson V., Berger A., Balakrishnan H.** Fast Portscan Detection Using Sequential Hypothesis Testing // Proc. of the IEEE Symposium on Security and Privacy, Berkeley, CA, United States. – 2004. – P. 211–225.

Received 2008 12 22

N. Paulauskas, E. Garšva, J. Skudutis. Network Scan Detection Simulation // Electronics and Electrical Engineering. – Kaunas: Technologija, 2009. – No. 2(90). – P. 43–46.

The network scan detection provides information about the attempts of attackers to determine servers and to prevent potential trials to compromise them. In the work the model for simulation of the network scanning and its detection using the stochastic activity network and their simulation tool Mobius is proposed. The model allows operative evaluation of the influence of simulated computer system parameters on the scan detection efficiency and proper configuration of the intrusion detection system. The results of the horizontal scan simulation are presented. The IDS parameters having the highest influence on the intrusion detection performance for a specific case have been determined. Ill. 3, ref. 5 (in English; summaries in English, Russian and Lithuanian).

Н. Паулаускас, Е. Гаршва, Ю. Скудутис. Моделирование системы обнаружения вторжений // Электроника и электротехника. – Каунас: Технология, 2009. – № 2(90). – С. 43–46.

Установление факта сканирования портов позволяет пресечь попытки атакующих определить потенциально уязвимые точки воздействия и скомпромитировать систему. Предложена модель обнаружения атак с использованием стохастических сетей и пакета Mobius. Модель позволяет оперативно определить влияние параметров информационной системы на эффективность обнаружения атак и правильно сконфигурировать систему их обнаружения. На примере детектора горизонтальных атак определены наибольшее влияние оказывающие факторы. Ил. 3, библ. 5 (на английском языке; рефераты на английском, русском и литовском яз.).

N. Paulauskas, E. Garšva, J. Skudutis. Tinklo žvalgos aptikimo sistemos modeliavimas // Elektronika ir elektrotechnika. – Kaunas: Technologija, 2009. – Nr. 2(90). – P. 43–46.

Žvalgos aptikimas suteikia informacijos apie atakuojančiųjų bandymus nustatyti paslaugas teikiančias tarnybines stotis ir užkirsti kelią potencialiems bandymams jas sukompromituoti. Straipsnyje pasiūlytas modelis tinklo žvalgai ir jos atpažinimui modeliuoti, panaudojant stochastinius veiklos tinklus ir jų modeliavimo įrankį *Mobius*. Modelis leidžia operatyviai įvertinti modeliuojamos kompiuterių sistemos parametrų įtaką žvalgos aptikimo efektyvumui ir teisingai sukonfigūruoti atakų atpažinimo sistemą. Pateikti horizontaliosios žvalgos modeliavimo rezultatai. Nustatyti IDS parametrai, turintys daugiausia įtakos atakų aptikimo efektyvumą konkrečiu atveju. Il. 3, bibl.5 (anglų kalba; santraukos anglų, rusų ir lietuvių k.).

DOI: 10.5755/j02.eie.10500