International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:2, No:11, 2008

Specification of a Model of Honeypot Attack based on Raised Data

Souleymane Oumtanaga, Prosper Kimou, and Kouadio Gaza Kevin

***Abstract*—**The security of their network remains the priorities of almost all companies. Existing security systems have shown their limit; thus a new type of security systems was born: honeypots. Honeypots are defined as programs or intended servers which have to attract pirates to study theirs behaviours. It is in this context that the leurre.com project of gathering about twenty platforms was born. This article aims to specify a model of honeypots attack. Our model describes, on a given platform, the evolution of attacks according to theirs hours.

Afterward, we show the most attacked services by the studies of attacks on the various ports.

It is advisable to note that this article was elaborated within the framework of the research projects on honeyspots within the LABTIC (Laboratory of Information Technologies and Communication).

***Keywords*—**Honeypot, networks, attack, leurrecom, computer network

I. Introduction

NETWORK security is all methods and capacities implemented to guarantee integrity, confidentiality and totality of the information. The evolution of the world wide interconnections increases the venerability of information systems and the risk of misuse and sabotage. Indeed, data can be destroyed, amputated, falsified, seized, plagiarized or modified in several manners by evolved in system attacks.

Thus it seems necessary and essential to reassure information systems to guarantee data integrity.

In this optics several technologies were born: authentication, data encoding, firewalls and IDS (NESTS / HIDS).

All these technologies were passed-by by hackers, and crackers in the aim of showing their talent of pirate or for another devil reasons.

The limits of these technologies are essentially due to their functioning principle and their implementation. It is sensible to define a new approach of defence which has to be different from the existing technologies; in our paper we introduce honeypots. Honeypot is security technology which is based on attracting aggressors in a controlled zone of the network (bait) to take them away from real production servers. Any activity in the honeypots is considered as suspect and is thus stored in a file; what allows the study of tools and methods used by hackers.

Authors are with LARIT-Inphb 08 BP 475 Abidjan 08 Côte d'Ivoire (e-mails: Oumtana@nic.ci, kkimou02@yahoo.fr, kouadiokev@yahoo.fr).

Our work lies within Leurée.com framework environment. We use the data base of Cadho project [1]. This project aims at developing and giving to the scientific community a distributed platform of data collection. This platform will help in understanding and collecting information on hackers and crackers attacks through internet.

We present in this article a Model of honeypots attacks based on data collected from Leurré.com. In second part we outline the difference between research honeypots and production honeypots. The third part is dedicated to the modelling of attacks. In the fourth part we analyze the results.

II. Various Types of Honeypots

Honeypot is a server or a voluntarily vulnerable program, intended to attract pirates [2] in order to be attacked by them. It is designed to attract attackers, for example by hosting weak or interesting services like very old vulnerable wu-ftpd or IIS web server. This device has to persuade attackers that it is a real system, and has to limit their possibilities (no bounce on an external network).While being attacked; honeypot is a very interesting sensor for an intrusion detection purpose. In fact, as a honeypot [15] is not a real system, every single probe can become a security alert. For example, by deploying a honeypot on a company network [3], it shouldn't have any dialog with the real network system; so if anybody from this company tries to exchange with this very specific host, it may reveal a weird activity.

We can point out two types of honeypots: production and research honeypots. The purpose of a production honeypot is to help mitigate risk of attacks in an organization. While they help in a lesser measure in the prevention, they can greatly [16] contribute to the detection or the reaction. Research honeypots are designed to collect information on the "blackhat" community. These honeypots do not add direct value to a specific organization; instead, they are used to gather intelligence on general threats that organizations may face, and to help them improve their defence against those threats [17]. The next section gives more details about research and production honeypots.

A. Honeypots of Research

The purpose of research honeypots is to scan information in order to study the progress of a hacker, to learn their new attack techniques and encircle them better. So, they study hackers by using a great number of configurations. Hackers think they are working on some production computer. But, in reality they are working on honeypot which observes all theirs activities. In this case honeypot role isn't to improve the protection but to observe. It is even preferable to reduce the

access security level in order to make the network reachable. However it requires a permanent follow-up to avoid surprises and also recommend a good control of the network and the systems. It is necessary in this case systems have real time monitoring processes to be able to follow Aggressors and outline theirs behaviours at precise moments. It is however necessary to be discreet not to be tracked down by the aggressor [5].

B. Honeypots of Production

The reason of deployment of such system is the protection of the data and thus the network in which these are stored. A honeypot can protect an organization in three ways:

- The prevention of attacks: let a hacker play on the honeypot instead of the systems of production.
- The detection of roguish activities: In this case it is question of using intervention detection systems.

The answer to attacks is the routing of any suspect traffic towards honeypots. This kind of system does not really require specific follow-up if it has been configured with efficiency. Indeed, it will divert the pirates on zones wished by the network without opening any breach on the sensitive data. However, it will be necessary to foresee log files which keep the interventions tracks. These must be placed on a different server so that they cannot be modified by the pirate who usually erases the tracks of his passage.

We distinguish two types of interaction: honeypots with weak interaction and honeypots with strong interaction:

- Honeypots with weak interaction collect a maximum of information while offering a minimum of privileges to the pirates. They allow to limit the risks at most.
- Honeypots with strong interaction can be considered as highly-rated extreme of the subject because it is based on the principle of real services access on a machine of the network more or less reassured. The risks are naturally much more important than in the case of honeypots with weak interaction.

C. Some Existing Platforms

1. The Project CADHo

The CADHo project within the framework of computer security has set up the Leurre.com environment. It is a platform based on a collaboration between many partners who spread a set of honeypot all configured in an identical way [6]. The following figure gives the general architecture of the honeypot platform [7].

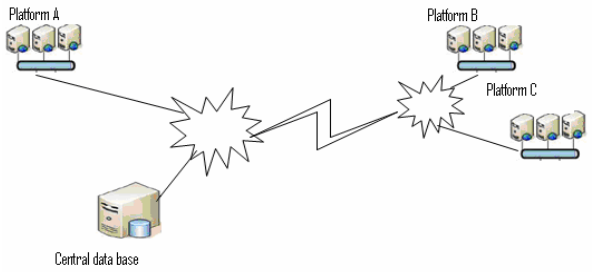

Fig. 1 Honeypots' distributed platforms

It is a network of about thirty platforms spread in about twenty countries all around the world. On every platform are installed honeypots to watch aggressors' behaviours. Data collected in every platform are daily sent to a central data base installed in Eurecom.

This base is then enriched by other information on the geographic localization, the operating system and the domain name. These data can be shared between people in charge of the networks who spread a platform to central data base.

2. Detailed Architecture

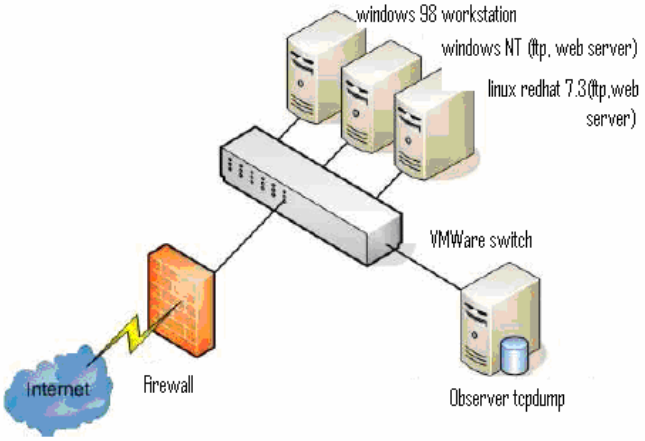

Fig. 2 Detailed architecture of a platform

In great detail a platform consists of three honeypots [8]. These virtual machines are feigned by using VMWare [9] or Honeyd [10]. Operational systems installed on these machines are Windows 98, Windows NT and Linux Readhat 7.3. In these three virtual machines, which moreover appear to the aggressor as true machines, we add an observer which arrests all traffics which arrive on the platform. Data arrested during a day are stored in a tcpdump file and repatriated within a server of data base administered by Eurecom Institute. The sent or successful received packages are enriched by additional information such as the geographic localization of the source, the jet lag, the type of operating system. A firewall is used to refuse connections of the virtual machines towards the outside and to accept those in opposite way. This measure allows avoiding the case where the honeypots are used to attack other networks.

III. Data Modelling

Our source is the data stemming from the central data base of Leurrée.com. The model is characterized by the evolution of all attacks stemming from all existing platforms per hour. We represent the number of attacks in a circular diagram according to the accorded ports.

A. Steps of the Modelling

It is advisable to note that the data that we are going to exploit are those stemming from all the platforms. In the model, we consider two factors which are, in one hand, attacks according to the hour of the day to highlight the periods of the day in which activities are strong. In the other hand we shall see the services aimed through the study of attacks on various connection ports.

1. Attak according to the Hour of the Day

The first step of the modelling was the collection of data. We consider the number of attacks during the 24 hours of every day within the period of one year. A part of the results is shown in the following board [11]:

Hour	Number of sources	Rate
0	89469	3.99
1	79422	3.54
2	72593	3.23
3	66506	2.96
4	63130	2.81
5	60730	2.71

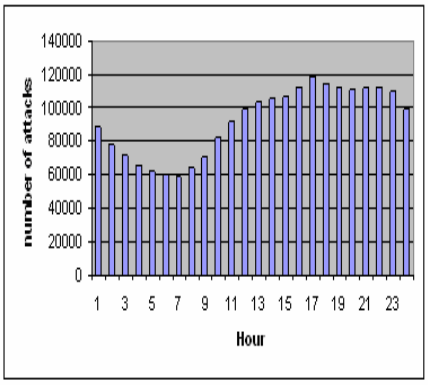

Fig. 3 Results and graph of attacks relative at the hour

2. Data of the Ports Attacks

After the study of attacks per hour; we have, afterward for the same period and according to ports, gotten back the number of attacks undergone by platforms. These data are recapitulated in the following picture: [11]

Services	Ports	Nombre d'attaque
http	80	56756
Telnet	23	1819
Microsoft-DS	445	226978
Netbios-SSN	139	112983
FTP	21	3383
Microsoft-SQL Server	1433	56906
SMTP	25	2222

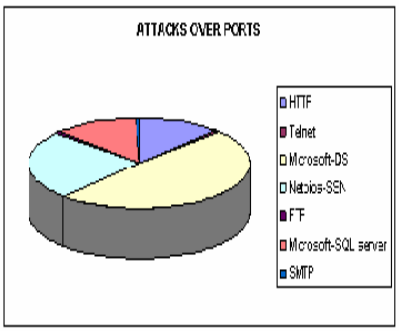

Fig. 4 Results of attacks over ports

B. Mathematical Equations

The search of a mathematical model brought us to make a representation of the data obtained according to time but by considering the period of time from 6 am to 5 am of the next day. For that, we realized an exchange of variable in the time that leads us to consider time from 6 to 29. These changes lead to the following representation:

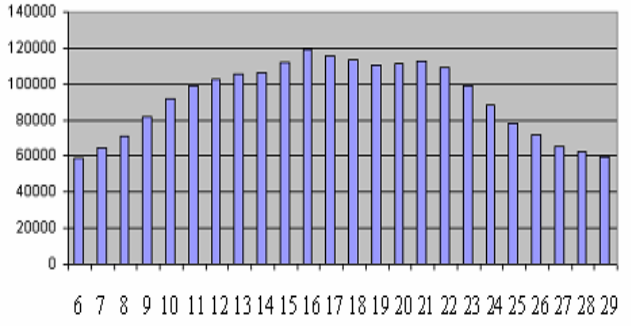

Fig. 5 Graph of the results with the change of variable

The analysis of this graph shows that the evolution can be approached to by the normal law probability density which is shaped as follows: [12]

$$F(X) = b\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) + a(X-u)(X-v)+c$$

Where

- a, b, c, u and v are constants depending on the platform:

Here a=-382.4688; b=39046400.447; c=60046; u=6 and v=29.

- σ represent the standard deviation, which allows characterizing the dispersal of the values in regard to the average

$$\sigma = \sqrt{VAR(X)} \quad ; \quad VAR = \sum_{1}^{24} p_i * (X_i - \bar{X})^2$$ Where

$\bar{X}$ represents the average;

Calculations give $\sigma = 5967.2587$

- μ: Represents the median which separates the number of sample in two parts. It is enough to organise values in an increasing order and to choose the middle as median: $\mu = \frac{(u+v)}{2}$; here μ=17.5.

Fig. 6 shows the graph of the estimated and real results:

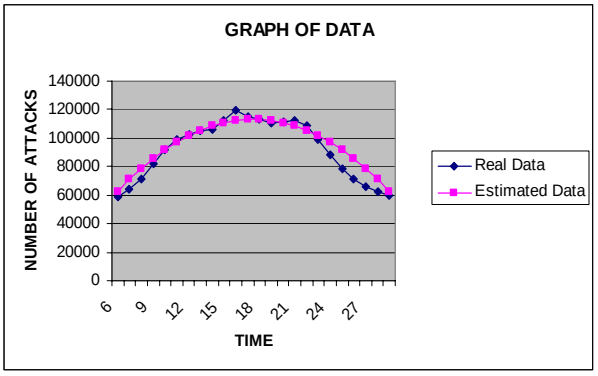

Fig. 6 Graph of the estimated and real data

IV. Results' Analysis

A. Analysis of the Attacks According to the Hour of the Day

This model aims at estimating the number of attacks on a given platform every hour of the day. To validate our model, we are going to apply it to the platform of Abidjan.
The results are shown in the following picture and graph.
Constants of the platform are:

a= -2.185255 ; b=1 ; c=135 ; u=1 ; v=24 ; μ=12.5 ;

Hour	Number of sources	Rate
0	187	2.26
1	192	1,91
2	216	1,84
3	245	2,55
4	375	2,62
5	382	2,94

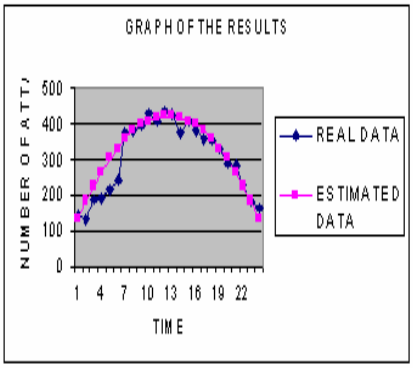

Fig. 7 Results of the application of the model has the platform of Abidjan

Obtained results show that the model can be applied to the Abidjan's platform. It is however necessary to specify that the evolution of the number of attacks varies strongly according to platforms. The results above remain to be confirmed in the time.

B. Analysis according to the Ports

The figure below shows the percentages of attacks over ports:

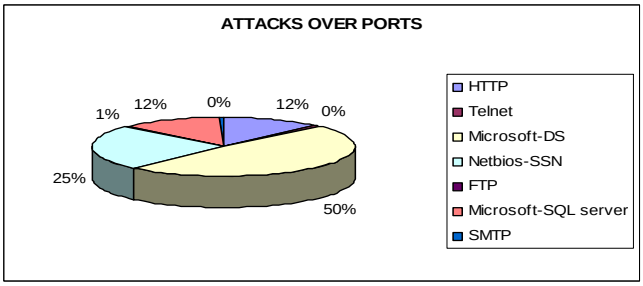

Fig. 8 Analysis of the attacks on ports

The results relative to the aimed services show that half of the attacks is steered towards the services Microsoft-DS which are used for the file sharing on machines turning on Windows 2000, XP, or on 2003 [13]. It is used for example to connect the shared files.

NetBIOS-SSN is the second target of attacks. They are still directed to the shared resources but on Windows 9x, ME and NT. Then come at the same proportions http and Microsoft-SQL server. These models could constitute tools intended for the network administrators. They would allow having a precise idea of the most aimed services by attacks, and estimate the number of attacks at any hour of the day.

V. Conclusion

The leurre.com project represents, today, a world effort in the field of abnormal traffic collection on Internet and also for the study of their property and causes [14]. From these studies, more clear models describing the activities of hacker will be elaborate. We note a growing interest towards the development of quantitative evaluation methods and representative measure. It is within the framework of this project that the team base in Abidjan specified a model of attacks according to the hour of the day. We can notice that the obtained model presents a good adequacy with the observed data. Other studies also allowed accentuating the most aimed services by analyzing attacks on ports. The results remain to be confirmed in the duration. It is important to justify the results to validate their aptness. It is however necessary to clarify that this model can show itself unusable for certain platforms because the behaviour varies strongly according to platforms. So, later tests shall certainly allow to bring improvements but also to describe other more generic models. These models will allow projected evaluations, and can constitute a help for the improvement of the security systems.

References

[1] E. Alata, M Dacier ''Leurré.com : retour d'expérience sur plusieurs mois d'utilisation d'un pot de miel distribué mondialement''.
[2] Home Page du projet Honeynet, http ://www.honeynet.org/, dernière visite 19 /09/2005.
[3] L. Spitzner, Honeypots: Tracking Hackers, Add.-Wesley, ISBN from-321-10895-7, 2002.
[4] French Honeynet Project, http: //honeynet.rstack.org.

[5] F. Pouget, T. Holz, _A Pointillist Approach for Comparing Honeypots_, Proc. Conference on Detection of Intrusions and Malware & Vulnerability Assessment (DIMVA 2005), Vienne 9.
[6] ACI Sécurité et Informatique, http ://acisi.loria.fr.
[7] Fabien Pouget. Leurré.com, the Eurecom Honeypot Project introduction. http://www.eurecom.fr/~pouget/leurrecom.htm.
[8] NGUYEN Programme d'alerte base sur des pots de mile, septembre 2005.
[9] VMware Corporation Home Page, http: //www.vmware.com
[10] Honeyd Home page, http: //www.citi.umich.edu/u/provos/honeyd
[11] https://riviera.eurecom.fr/.
[12] fr.wikipedia.org/wiki/Loi_normale; last access sept 2006.
[13] Honeypot-Based Forensics F Pouget and M Dacier, Proceedings of the Asia Pacific Information Technology Security Conference, (Auscert), 2004.
[14] What's port 445 used for in windows 2000/XP; www.petri.co.il
[15] Global Intrusion Detection: Prelude Hybrid IDS Mathieu Blanc1, Laurent Oudot1, and Vincent Glaume, rapport de recherche 2002.
[16] The Value of Honeypots, Part Two: Honeypot Solutions and Legal Issues by Lance Spitzner with extensive help from Marty Roesch last updated October 23, 2001.
[17] The value of Honeypots, Part one: Definitions and Values of Honeypots *Lance Spitzner* 2001-10-10.