CADHo:
*C*ollection and *A*nalysis of *D*ata from *Ho*neypots

E. Alata[1], M. Dacier[2], Y. Deswarte[1], M. Kaâniche[1], K. Kortchinsky[3],
V. Nicomette[1], V.H. Pham[2], F. Pouget[2]

[1]LAAS-CNRS
7 Avenue du Colonel Roche, 31077 Toulouse Cedex 4, France
{ealata, deswarte, kaaniche, nicomett}@laas.fr
http://www.laas.fr
[2]Eurécom
2229 Route des Crêtes, BP 193, 06904 Sophia Antipolis Cedex, France
{dacier, pham, pouget}@eurecom.fr
http://www.eurecom.fr
[3]CERT-RENATER
c/o ENSAM, 151 Boulevard de l'Hôpital, 75013, Paris, France
http://www.renater.fr/Securite/CERT_Renater.htm
kostya.kortchinsky@renater.fr

Abstract. The CADHO project is an on going research action funded by the French ACI "Security & Informatics" [1]. It aims at building an environment to better understand threats on the Internet and also at providing models to analyze the observed phenomena.

Introduction

Since the very first distributed denial of service attacks launched in February 2000, an apparently increasing number of major security problems have been reported. Surprisingly, the number of observed attacks does not seem to be influenced by the ever increasing deployment of efficient security tools, such as personal desktop firewalls. Is this impression of raise of number of attacks backed up by some undisputable data? If yes, what are the attack processes that lead to such phenomena?

As of today, we are unfortunately unable to answer these questions because of the lack of precise and unbiased data to assess the seriousness of the situation. A few qualitative indicators exist, but they are not rich enough to enable us to carry out any kind of serious analysis of the malicious behaviors at stake, and to model attack processes and their impact on the target systems security. CADHo addresses those issues by means of the following actions:

1. We are deploying and sharing with the scientific community a distributed platform that gathers data suitable to analyze the attack processes targeting machines connected to the Internet.

2. We have validated the usefulness of this platform by carrying out various analyzes based on the collected data to characterize the observed attacks and model their impact on security.
3. We aim at going beyond the study of the most frequent and automated attacks. We want to investigate and model the behavior of malicious attackers once they have managed to get access to a new host.

The honeypot platform we have built has been deployed in a couple of dozen places around the world. We collect all data in a centralized database. It enables us to have a better understanding of the attacks and of the attackers' behaviors. The data obtained thanks to this distributed setup constitutes a foundational element available to the various communities working on preventing, detecting or tolerating intrusions. Our objective is to provide solid rationales to those who need to validate the fault assumptions they make when designing, for instance, intrusion tolerant systems.

1. Honeypots: Initial Setup and Deployment

One of the goals of the CADHo project is to share with the scientific community an open distributed platform to collect data from a large number of honeypots. This platform is deemed to evolve over the years, well beyond the end of the CADHo project. New partners are allowed to get access to the data set if and only if they agree to set up a honeypot on their premises, thus enriching the overall setup by their presence. Names of the partners are protected by a Non Disclosure Agreement that each participating entity must sign. We have developed all the required software to automate the various regular maintenance tasks (new installation, reconfiguration, log collection, backups, etc.) to ensure the long term existence of this set up.

A honeypot is a machine connected to a network but that no one is supposed to use. In theory, no connection to or from that machine should be observed. If not, it must be, at best an accidental error or, more likely, an attempt to attack the machine. Recently, several approaches have been proposed to build environments where several honeypots are deployed. The generic term *honeynet* is used to represent them. The most visible honeynet project is the one carried out by the so called Honeynet research Alliance [2, 3]. The Alliance is made of national entities. Some CADHO members are active members of the French one, the French Honeynet Project [4]. So far, most of the attention has been paid to implementation issues. Institut Eurécom has been working for more than a year on the definition of a low-interaction honeypot dedicated to the tasks explained here above. A first environment has been deployed, based on the VMWare [5] technology. The results obtained have been published in international conferences in the course of 2004 [6,7]. An up to date list of publications on this topic can be found in [8]. Based on this acquired expertise, we are now convinced that, for these specific experiments, the freely available software called *honeyd* [9] can be used instead of VMWare. This conclusion is backed up by data obtained so far. Indeed, it is known that the major drawback of *honeyd* is that an environment using that software can be remotely identified by a skilled attacker. This is not the case for VMWare. Fortunately, data collected so far indicate that the risk of seeing attackers fingerprinting the environment under attack is negligible. This justifies the choice of a *honeyd* based solution.

Honeyd is a free software and it runs on various flavors of Linux and Windows. It does not consume a lot of resources and, therefore, old PCs can be used without any trouble. These are very interesting features since we are interested in building a large environment where several honeypots would run. The fact that we can add honeypots for almost no cost makes this solution very attractive. It is indeed unlikely that we could identify interested partners to join this platform on a voluntary basis otherwise.

The distributed platform itself is made of a potentially large number of honeypots and of a large centralized database. All the honeypots are centrally managed to ensure that they have exactly the same configuration. This is very important if we want to keep the experiment under control. The data gathered by all honeypots are securely uploaded to a centralized database. This database contains in a highly structured and efficient way, the content, including the payload, of all packets sent to or from these honeypots. Furthermore, the data are enriched by means of several techniques to facilitate their analysis. All partners have the possibility to send queries to that database through a secure web interface.

2. High-Interaction Honeypots

The honeypots that we have already deployed in the context of this project belong to the family of so-called “low interaction honeypots”. This means that their design is such that attackers have not the possibility, at any point in time, to actually get access to the machine they are attacking. This property is enforced by the fact that there is no real machine but that, instead, targets are implemented by means of virtual machines without any real operating system or server to compromise. Thus, hackers can only scan ports and send requests to fake servers without ever succeeding in taking control over them.

In the CADHo project, we are also interested in running experiments with “high interaction” honeypots where attackers can really compromise the targets. Collecting data from such honeypots would enable us to study the behaviors of attackers once they have managed to get access to a target. Of course, we will not let them bounce from these machines to run attacks against third party machines. Instead, we will construct a simulated environment within which they could evolve. An important feature of the environment we are planning to build is that it will “select” the attackers that we will, or will not, let compromise our machines. Indeed, we are not interested in monitoring all kinds of attackers. Instead, we want to monitor only those that are representative of large classes of attackers so that the knowledge derived from their observation is symptomatic of a large amount of real attacks. Such high interaction honeypots will be deployed within a limited number of highly controlled environments.

The experiments and the data that will be collected based on the high-interaction honeypots will enable us to address two distinct objectives. First, we are interested in better understanding the attack scenarios, in particular those carried out by skilled intruders. This acquired knowledge will be useful to build concrete answers and develop tools to counter this form of attack which is known to be very costly but which has received little attention up to now. Second, we want to propose concrete and efficient techniques to assess the impact of such ongoing attacks on the security of the target system. Along this line, we propose to use observations from this setup to validate a theoretical model initially developed in our previous work on quantitative analysis of operational security in the 90's [10, 11]. The original method is a probabilistic one that differs from classical qualitative approaches (red book, ITSEC, common criteria, etc.). The core of the method lies in a so called privilege graph which highlights the various possibilities offered to an insider to increase his privileges thanks to identified vulnerabilities or features of the system he has access to. We have shown how to use that graph to derive probabilistic estimations of the ability of a system to resist attacks. The limitations of that approach reside in the absence of real world validation of the assumptions made about the behaviors of the insiders. Common sense dictated our design but a more rigorous approach requires running some experiments to validate our claims. Thanks to high-interaction honeypots, this is something that now becomes feasible and something that we aim to do within the CADHo project.

References

1. ACI Sécurité et Informatique, http://acisi.loria.fr.
2. Home Page of the Honeynet Project, http://www.honeynet.org/, last visited 03/2005
3. L. Spitzner, *Honeypots: Tracking Hackers*, Add.-Wesley, ISBN from-321-10895-7, 2002
4. Home page of the French Honeynet Project, http://honeynet.rstack.org.
5. VMWARE, Home page, http://www.vmware.com.
6. M. Dacier, F. Pouget, H. Debar, "Honeypots: Practical Means to Validate Malicious Fault Assumptions on the Internet", *Proc. 10th IEEE International Symposium Pacific Rim Dependable Computing (PRDC10)*, March 2004, pages 383-388.
7. M. Dacier, F. Pouget, H. Debar, "Attack Processes found on the Internet", *Proc. OTAN Symposium on Adaptive Defense in Unclassified Networks*, April 2004.
8. F. Pouget, Publications web page, http://www.eurecom.fr/~pouget/papers.htm.
9. Honeyd Home page, http://www.citi.umich.edu/u/provos/honeyd/
10. M. Dacier, Y. Deswarte, M. Kaâniche, "Models and tools for quantitative assessment of operational security", *Proc. 12th International Information Security Conference* (IFIP SEC'96), Samos (Greece), May 1996, pages 177-186
11. R. Ortalo, Y. Deswarte, M. Kaâniche, "Experimenting with Quantitative Evaluation Tools for Monitoring Operational Security", *IEEE Transactions on Software Engineering*, Vol.25, N°5, pages 633-650, September/October 1999.