The WOMBAT Attack Attribution Method: Some Results

Marc Dacier[1], Van-Hau Pham[2], and Olivier Thonnard[3]

[1] Symantec Research, Sophia Antipolis, France
marc_dacier@symantec.com
[2] Institut Eurecom, 2229 Route des Crètes,
Sophia Antipolis, France
van-hau.pham@eurecom.fr
[3] Royal Military Academy, Polytechnic Faculty
Brussels, Belgium
olivier.thonnard@rma.ac.be

Abstract. In this paper, we present a new *attack attribution* method that has been developed within the WOMBAT[1] project. We illustrate the method with some real-world results obtained when applying it to almost two years of attack traces collected by low interaction honeypots. This analytical method aims at identifying large scale attack phenomena composed of IP sources that are linked to the same root cause. All malicious sources involved in a same phenomenon constitute what we call a *Misbehaving Cloud* (MC). The paper offers an overview of the various steps the method goes through to identify these clouds, providing pointers to external references for more detailed information. Four instances of misbehaving clouds are then described in some more depth to demonstrate the meaningfulness of the concept.

1 Introduction

There is no real consensus on the definition of "attack attribution" in the cyber domain. Most previous work related to that field tend to use the term "attribution" as a synonym for *traceback*, which consists in "determining the identity or location of an attacker or an attacker's intermediary" [25]. In the context of a cyber-attack, the obtained identity can refer to a person's name, an account, an alias, or similar information associated with a person or an organisation. The location may include physical (geographic) location, or any virtual address such as an IP address or Ethernet address. The rationale for developing such attribution techniques is mainly due to the untrusted nature of the IP protocol, in which the source IP address is not authenticated and can thus be easily spoofed. An extensive survey of attack attribution techniques used in the context of IP traceback can be found in [25].

[1] Worldwide Observatory of Malicious Behaviors and Threats - http://www.wombat-project.eu

A. Prakash and I. Sen Gupta (Eds.): ICISS 2009, LNCS 5905, pp. 19–37, 2009.

In this paper, we refer to "attack attribution" as something quite different from what is described here above. We are primarily concerned with larger scale attacks. In this context, we aim at developing an analytical method to help security analysts in determining their root causes and in deriving their *modus operandi*. These phenomena can be observed through many different means (e.g., honeypots, IDS's, sandboxes, web crawlers, malware collecting systems, etc). In most cases, we believe that attack phenomena manifest themselves through so-called "attack events", which can be observed with distributed sensors that are deployed in the Internet. Typical examples of attack phenomena that we want to identify vary from worm or malware families that propagate through code injection attacks [9], to established botnets controlled by the same people and targeting machines in the IP space. All malicious sources involved in the same root phenomenon constitute what we call a *Misbehaving Cloud* (MC).

The structure of the paper is as follows: Section 2 describes the experimental environment used to validate the method presented. Section 3 offers a high level overview of the attack attribution method defined within the WOMBAT project and Section 4 gives some more information on the multi criteria fusion approach used in the method. Section 5 discusses a couple of illustrative examples obtained by applying the method on honeynet traces, and Section 6 concludes the paper.

2 Description of the Experimental Environment

This paper offers an empirical analysis of some attacks collected during two years by a set of low interaction honeypots deployed all over the world by the Leurré.com Project [10]. We refer the interested reader to [8,19] for an in-depth presentation of the data collection infrastructure. From an analytical viewpoint, our attack attribution method builds upon previous results, namely [18,4,16,24,17]. For the sake of clarity, we start by introducing some important terms that have been defined in these previous publications.

2.1 Terminology

1. **Platform:** A physical machine running three virtual honeypots, which emulate three distinct machines thanks to *honeyd* [20]. A platform is connected directly to the Internet and collects tcpdump traces that are gathered on a daily basis in a centralized database [10].
2. **Source:** An IP address that has sent at least one packet to, at least, one platform. An IP address remains associated to a given Source as long as no more than 25 hours[2] elapse between two packets sent by that IP. After such a delay, the IP will be associated to a new source identifier if we observe it again.

[2] By grouping packets by originating sources instead of by IPs, we minimize the risk of mixing together the activities of two distinct physical machines (as a side effect of the dynamic address allocation implemented by ISP's).