Chapter 8.4
Challenges in Sharing Computer and Network Logs

Adam Slagell
University of Illinois at Urbana-Champaign, USA

Kiran Lakkaraju
University of Illinois at Urbana-Champaign, USA

ABSTRACT

It is desirable for many reasons to share information, particularly computer and network logs. Researchers need it for experiments, incident responders need it for collaborative security, and educators need this data for real world examples. However, the sensitive nature of this information often prevents its sharing. Anonymization techniques have been developed in recent years that help reduce risk and navigate the trade-offs between privacy, security and the need to openly share information. This chapter looks at the progress made in this area of research over the past several years, identifies the major problems left to solve and sets a roadmap for future research.

DOI: 10.4018/978-1-60566-414-9.ch004

INTRODUCTION

On March 20, 2004, the security incident response team at the National Center for Supercomputing Applications (NCSA) at the University of Illinois received an automated alert indicating that a particular NCSA machine was making an atypical number of outbound connections to external hosts. Often, when something like this happened in the past, it was because a machine had been infected with a worm or become part of a botnet. Naturally, the team investigated the anomaly, and they found that unauthorized ports were open. By scanning the machine and reviewing their network flows, they found that the host was running a backdoor SSH client granting remote access to an unauthorized

user. Worse yet, a subsequent scan of the network revealed that other machines had the same strange port open and were also compromised. Little did they realize that this was only the very smallest tip of the iceberg.

Rather quickly, it was discovered that the attacker, who later started identifying himself as "Stakkato," spread his attacks across much more than the NCSA network. He exploited a number of specific vulnerabilities across many of the TeraGrid sites. The TeraGrid was at the time the world's largest, most comprehensive distributed computing infrastructure for open scientific research, with high-performance computing resources spread across 11 institutions. While the attacks were expanding to encompass more and more institutions, they were also escalating in frequency. Because the attacker installed Trojaned SSH daemons on many infected machines, he was able to compromise accounts faster than they could be closed or have their passwords changed. This problem was exacerbated by the fact that many of the TeraGrid resources shared authentication credentials, as a typical user could run jobs on any of the TeraGrid supercomputers. Some of the sites were at times just trying to keep their heads above water to stay on top of this problem; eventually, all users were forced to change their passwords at these sites.

As the scope of the problem grew, even beyond TeraGrid, the FBI was brought in on the matter. A few key institutions became the points of contact between the FBI and the many other institutions involved with the case (which was named Major Case 216 by the FBI). Before the investigation was finally complete, the attacks had spanned 19 months and thousands of sites, including high-security military sites and federal research laboratories, university sites, private sector sites, and machines owned by individuals, both in the U.S. and in Europe. It was finally tracked back to a teenager in Sweden after whose apprehension the attacks suddenly stopped (Nixon, 2006).

Lessons Learned

We learned a great deal as one of the victim sites in this experience. First, not only can attacks be very large and sustained, but such attacks can be perpetrated by a single individual. In fact, if your organization is the target of a focused digital intrusion—not just worms or script-kiddies collecting bots—it is likely that your organization is just one of many involved in the same attack. Understanding the specific attack that we experienced required a very broad picture of the incident and the cooperation and collaboration of many individuals at many different institutions. Achieving this collaboration and establishing trust were among the main challenges of the endeavor.

It was not uncommon for a large site to invest thousands of man-hours on handling this incident. One organization might find compromised hosts from hundreds of other organizations. When our incident response team contacted the other incident responders and system administrators, they gave them details on the compromised machines and offered our help with the investigations. Of course, the responses ran the gamut, from people completely unwilling even to acknowledge what was told to them to people openly asking for help and readily sharing data. However, most people were reluctant to cooperate too much. Usually they would only answer questions as to whether or not a particular machine had also attacked them, or perhaps would share high-level network data, like network flows, with our team. Nevertheless, even the limited traffic data we were able to obtain helped us better understand the scope and overall structure of the attack.

Reasons for the reluctance included legal issues, privacy concerns, concerns about leaking sensitive information, and a general inability to establish trust and secure communication channels. In fact, most communication was an ad hoc mixture consisting primarily of phone calls and PGP-encrypted e-mails. Luckily, there were already existing relationships with several other