Trusted Internet Forensics: design of a network forensics appliance

extend abstract

D. Bruschi M. Monga E. Rosti
Dip. di Informatica e Comunicazione
Università degli Studi di Milano
{*bruschi, monga, rosti*}*@dico.unimi.it*

***Abstract*— With the spreading of cyber-crime, computer forensics has emerged as a new discipline in the system security arena. Some work is being done towards the definition of methodologies for the collection of digital evidences from storage devices that can withstand legal analysis in court. On the contrary, the collection of network evidences that allows for a selection of the traffic and guarantees legal admissibility is still an open field.**

In this paper we present the architecture of TIF, Trusted Internet Forensics, a network appliance that collects data from the network for forensics purposes. Such an appliance relies on a Trusted Computing Platform in order to allow for the verification of the computational chain so that the data collected could be used as evidence in court.

I. Introduction

Internet and information and communication technology are the innovations that have most influenced our society in the last decade. The way several human activities are performed has changed with advantages in terms of efficiency, pervasiveness, service-user mobility, etc. Unfortunately, such technologies have also introduced new problems when they are used improperly, opening the door to the so called *cyber-crime*. Computer crimes, i.e., illegal, dishonest, or unauthorized processing and/or transmission of data, such as computer intrusions, and computer related crimes, i.e., traditional crimes that can take advantage of ICT technologies to improve it effectiveness, are different types of cyber-crime. Fighting cyber-crime has become a significant portion of the activities carried out by law enforcement agencies. Computers and their storage media are the object of investigative analysis, known as Computer Forensics, aimed at characterizing and interpreting the digital evidences present on the devices.

Computer Forensics as a discipline aims at defining methodologies and developing tools in order to support investigative activities regarding computer systems. Most of the research work carried out so far has focused on storage systems. However, since a considerable percentage of cyber-crimes are pursued by exploiting computer networks, a significant amount of evidential data come from the network. In particular TCP/IP networks are the most common scenario for cyber-crime.

On-the-fly traffic inspection is a common activity in network intrusion detection, but for forensics purposes traffic should be recorded. While archiving data on disks or other mass storage devices for forensic analysis is not much different from their normal usage, archiving network data for forensic analysis may be difficult. The reasons are the amount of data generated by networks whose bandwidth commonly ranges in the hundreds or thousands of megabits per second, and the possibly sensitive nature of part of such data, e.g., personal or banking information. In particular, in some cases, it may not even be legal to record information unless there is a compelling reason or court order. Furthermore, only the information relevant to the crime case under investigation should be processed, avoiding the collection of unrelated data of unaware users.

In this paper we present the architecture of TIF, Trusted Internet Forensics, a network device that collects data from the network for forensics purposes. Data selection during the collection phase, respect of the privacy of users not involved in the investigation, and admissibility of the collected data in court are among the features of TIF. Data mining and Trusted Computing Platforms [1], [2] are the key components that allow us to provide the features above. Additional cryptographic functionalities guarantee the protection of user data that is not relevant for the investigation. The use of a Trusted Computing Platforms as the base for our appliance allows the legal actors to verify the computational chain thus fostering transparency and cross-validation of investigative actions.

The rest of the paper is organized as follows. Section II reviews the main concepts of Trusted Computing Platforms. Section III describes the TIF architecture and the problems of building a network forensics appliance. Section IV concludes the paper.

II. Trusted Computing Platforms

A variety of security solutions have been proposed that aim at improving the security of a platform by reducing its exposure to attacks, e.g., perimeter defenses, strong authentication, antivirus programs, OS memory protection. The common characteristic of all such system is that they do not modify the target system in any way. Even though such tools have been shown to be quite effective in protecting computer systems, they have limited capacity of blocking new forms of attacks, as they generally rely on characterization of known ones. One of the most relevant contribution in such a direction is the research on Trusted Computing Platforms (TCP) characterized by the property of either computing according to its "initial

specifications" or promptly detecting any unauthorized modification of its components. Various proposal of TCP have recently appeared in the literature (see [1], [2]).

In this paper we will investigate the solution proposed by the Trusted Computing Group (TCG) [3], a multi-vendor consortium formerly known as the Trusted Computing Platform Alliance (TCPA). TCG has proposed several specifications for systems that, by using a modified BIOS and a supplemental chip hardwired on the motherboard, can systematically verify each software component cryptographically, as firstly proposed by Arbaugh et al. in [4]. The basic building block of a trusted platform is the Trusted Platform Module (TPM). It is supposed to verify the integrity of the system and grant access to protected resources only to trusted components. By using such a platform it would be possible to show *irrefutably* which component has computed the given output. This

III. Building a network forensics appliance

A simple Internet network forensics appliance could be based on a standard Intel-based PC and capture packets with an Ethernet interface running in promiscuous mode. Basically, the system analyzes the packets in memory, performs rudimentary data analysis and reduction, and writes selected results to disk or to a database over the network.

Depending on the complexity of the network and the amount of data to collect at the points one wishes to monitor, the precision of tapping may vary considerably. In particular, if one is trying to record every packet that is transmitted across a fully-loaded gigabit link, building a suitable capture platform and disk farm may be challenging because of the performance of the network card and protocol stack.

A different yet not less difficult problem comes from the use of the collected data in court. An objective and non-refutable assessment of the quality of the network dump for the judge and the other parties is hard to obtain. Furthermore, such an assessment is required also for all the tools that have analyzed or filtered the data, both for the discarded and the stored data. Even if the communication was encrypted, it is common to perform the so called "traffic analysis", inspecting every IP packet containing destination and sender address. By examining the flow of packets over time, it is possible to infer when a user is working, whom they communicate with, the Web sites they visit, etc. However, the quality of reconstruction relies entirely on the correlation tool one is using. Thus, it is crucial to be able to show their logic during the trial.

The main goals of the Trusted Internet Forensics appliance are the collection and analysis of network traffic, minimizing the resources that are needed to perform the task, and the generation of trace files that could be used as digital evidences in court. Such a device must also guarantee the protection of the privacy of all the network users not directly involved in an investigation.

The architecture of the TIF appliance is illustrated in Figure 1. All the system components are part of a trusted computing platform. This allows all the law enforcement agencies and other actors involved in cyber-crime investigation and trial to verify the computational chain between the data captured from the network and their image on the storage media. The data collected through the network card is passed to a *Tapping Quality Control* module which is responsible for the quality of the network dump. The data is then passed to the filter that selects the data to be stored and discards the rest, the *Intelligent Tapping Filter* in Figure 1. Finally, the data is stored on disk by the *Secure Data Storing* module, which is responsible for the cryptographic transformations that add confidentiality, integrity, and availability, required for digital evidences.

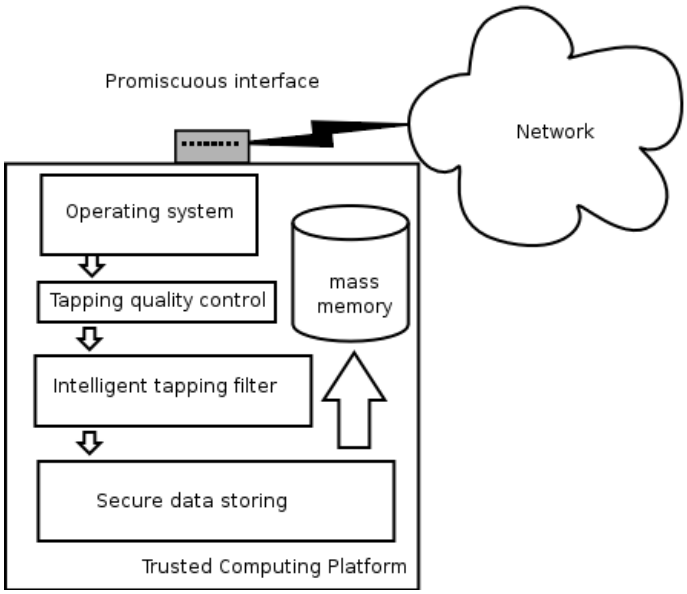

Fig. 1. The architecture of the Trusted Internet Forensics appliance.

In order to extract meaningful information from mainly unstructured data and to improve the response time of forensic analysis, the architecture may be extended with a data mining module that could take advantage of the most adequate strategies, such as association rules, neural nets, decision trees, temporal Markov chains, and temporal rules or combinations thereof. Such a module is responsible for the analysis of heterogeneous traffic data (texts, images, videos) and anomalous activities over networks and systems.

IV. Conclusions

In this paper we have presented the architecture of a TCP based appliance to collect network data for forensics purposes. The TCP allows for the verification of the computational chain, which could be a requirement for court admissibility of the digital evidences collected by the appliance. Specific modules to control the quality of the data gathering process and the cryptographic functionalities of the storing process fulfill the requirements of quality data selection and collection and confidentiality, integrity, and availability of the stored data. The detailed design of the proposed architecture is currently under development.

References

[1] P. England, B. Lampson, J. Manferdelli, M. Peinado, and B. Willman, "A Trusted Open Platform," *Computer*, vol. 36, no. 7, pp. 55–62, 2003.

[2] Tal Garfinkel, Ben Pfaff, Jim Chow, Mendel Rosenblum, and Dan Boneh, "Terra: a virtual machine-based platform for trusted computing," in *SOSP '03: Proceedings of the nineteenth ACM symposium on Operating systems principles*. ACM Press, 2003, pp. 193–206.

[3] *Trusted Computing Group*. http://www.trustedcomputinggroup.org.

[4] W. A. Arbaugh, D. J. Farber, and J. M. Smith, "A secure and reliable bootstrap architecture," in *SP '97: Proceedings of the 1997 IEEE Symposium on Security and Privacy*. IEEE Computer Society, 1997, pp. 65–71.