Generalizing sources of live network evidence

by Bruce J. Nikkel

nikkel@digitalforensics.ch

September 2, 2005

Abstract

This paper suggests combining the capture of network traffic and the collection of data from remote network services into a more general acquisition category of *live network evidence sources*. These two evidence sources exhibit many similarities, collected data share the same basic characteristics, and the acquisition architectures used for collection are very similar. When viewed from a more abstract perspective they can be described in the same terms. The OSI model's layered approach to networking can be used to help bring these two branches of network evidence together, organizing and reducing the complexity found in live network acquisition. The concept of an acquisition window is also introduced as a fundamental variable in live network acquisition.

Keywords: Network Forensics, Live Network Evidence, Live Network Acquisition, Live Network Sources, Cyber Forensics, NFAT

Contents

1 Introduction and Motivation

1.1 Introduction

There are many different sources of network forensic evidence. Two evidence sources in particular exhibit striking similarities, both in the characteristics of the data collected, and in the sources from which the data was acquired. The capture of live traffic from a network (for example, packet sniffing on an Ethernet segment) is considered a source of network evidence. The collection of data from remote network services (often beyond the control or jurisdiction of the investigator) such as websites, ftp servers, peer-to-peer networks, etc. is also considered a source of network evidence. From an abstract perspective, this evidence and the acquisition tools used to acquire it, can be viewed together in general terms.

A closer look at the data retrieved from both sources shows the same characteristics. For example, they both have the same issues with evidence authenticity and reliability. From an acquisition tool perspective they exhibit similarities when viewed within the context of a layered network model and differ largely by the layer at which the collection tools operate. They also share several fundamental differences with traditional storage media forensics. This paper looks at these similarities and proposes combining these two evidence sources into a single category of *live network evidence sources*.

The layered approach of the OSI model is significant in viewing these two areas as one. A layered view of the acquisition process makes it easier to understand the generalization and to visualize the similarities. It also makes it easier to incorporate some of the apparent differences into the generalized view (for example, active vs passive collection of data).

1.2 Motivation

There are several factors which drive the concepts outlined in this paper. One motivating factor is redundancy. Currently, network traffic capture and remote network service collection are viewed as two separate areas from an acquisition standpoint. However, merging these together into one general category may help to reduce parallel research and development work.

Another motivating factor is the reduction of complexity regarding live network acquisition. Network application protocols are becoming more complex as application logic is wrapped or encapsulated into increasingly higher and more abstracted application layers. Bringing in the OSI layered model

helps to simplify and organize a number of areas related to live network acquisition.

Much of the work in the area of network forensics is based on the TCP/IP protocol suite and other Internet protocols. Creating a general set of definitions for live network acquisition will allow this area of network forensics to become independent of any particular network technology. The concepts will then be applicable to other networks, even to network technologies not yet invented.

Finally, at a much higher level, the need for formalizing network forensics in general has provided some motivation for this work. Evidence is no longer confined to local physical media, and is increasingly found on remote servers or captured from live networks. This creates new problems which are not addressed in current acquisition methodologies. In the future, as the widespread availability of high speed network bandwidth increases and network users go mobile, a significant portion of digital evidence will be gathered from live network sources. Research needs to be done to ensure systematic and proper collection of this evidence.

2 Background

It is useful to show where the concepts in this paper sit in relation to other areas of digital forensic research, especially that of network forensics. From this network forensics landscape, two areas are identified which have much in common and are the basis for this paper. An overview of the impact and influence of other areas of forensic acquisition is also shown. The significance of the OSI model in network forensics is discussed, in particular, where it can be used to simplify data acquisition.

2.1 Network forensics

There are a number of areas which have been considered in the literature to be part of network forensics. A few of the more prominent examples are as follows:

- the analysis of IDS and firewall logs as evidence
- the back tracking of network packets and TCP connections
- the analysis of network related artifacts on forensically acquired harddisks
- the analysis of logs generated by network services and network applications

- the capture and analysis of network traffic using sniffers and NFAT[1] devices
- Collecting data from remote network services

This paper deals directly with combining the last two items of this list in a generalized way.

Traditional computer forensics provides generally accepted principles for media acquisition, some of which can be applied to live network evidence sources. The acquisition of storage media has been well documented in the past and a number of accepted principles[1][2][3] have been produced. The essence of these standards can be applied to the collection of live network evidence, but some difficulties remain. For example, the requirement for verifying the integrity of an acquired image assumes that a complete original exists for on-demand comparison. This is often not the case with live network sources. Also, the requirement for absolute write-blocking during acquisition has difficulty applying in a general way when collecting evidence from some live network sources.

Methods of acquiring or analyzing remote computers using forensic servlets, boot CDs, or PXE net booting are commonly known as remote forensics. Much practical work has been done in this area[4][5][6][7]. It is based on the premise that the remote machine is owned or controlled by the investigating organization, and allows special authorized investigative access for acquisition or analysis. Remote forensics in this respect is outside the scope of this paper and considered here to be an enhancement of traditional computer forensics. However, the work regarding authentication and encryption of network connections, together with the integrity of the target machine can be useful when researching live network sources.

Live system forensics and remote forensics both address issues of volatile acquisition. They also face many of the same challenges as network forensics. Live system forensics also deals with the issue of collecting evidence from a volatile environment where an original source does not exist to verify the authenticity of the acquired copy. Basic principles for memory acquisition have been proposed[8] which are based on the NIST requirements with some modifications, such as dropping the requirement for verifying integrity against the original source.

Using Intrusion Detection Systems as a source of evidence has been documented by Sommer[9] establishing a link between the IDS and network

[1]Network Forensic Analysis Tool

forensics communities. The IDS community has done much work in the areas of sensor placement, high volume data collection, dealing with errors, and filtering. All of this work can be useful in researching various aspects of data collection from live network sources.

2.2 Captured traffic as evidence

A significant overview on the subject of acquiring and analyzing captured network traffic as evidence has been documented by Casey[10]. Work on distributed NFAT systems[11][12] has also been done to address issues with evidence collection on larger networks. The construction of NFAT systems has borrowed heavily from IDS research areas. Of particular interest here is the acquisition process and the characteristics of the evidence collected.

2.3 Collecting evidence from remote network services

Work related to the forensic acquisition of data from remote network services such as websites, ftp servers, etc. is of significant interest in this paper. These servers may or may not be under the investigator's control or jurisdiction, but could still be regarded as a source of evidence. Foundational work in this area has already been done quite early on by Sommer[13]. The method suggests using screenshots and video captures of an end user application as evidence of remote system activity. It also introduces some key tests to determine authentication, accuracy, and completeness. The importance of continued research in this area has been recently highlighted[14].

2.4 The significance of the OSI model

Layers and abstraction help to organize and simplify the understanding of concepts which might otherwise be unnecessarily detailed and complex. The importance of using layers in digital forensics has been discussed in the literature[15][16].

The layered OSI (Open Systems Interconnect) network model[17] was created to simplify the complexities of network protocols and technologies by abstracting them from each other at various layers. This model has had a profound effect on the networking field. It has helped to increase interoperability among network vendors, and helped ease the understanding of networking technologies among IT professionals.

This model has been used extensively in organizing and simplifying the development of network forensics[18]. It can also be used to help solve certain issues relating to the acquisition of live network sources, in particular, issues related to network write-blocking, filtering, and locating evidence.

Most of the work done in network forensics today is based on the TCP/IP protocol and Internet technologies. Applying the OSI layered model to this area will allow these ideas to transfer easily to other network technologies.

3 Similarities between traffic capture and remote collection

3.1 Shared differences with traditional forensics

Network traffic capture and remote network service collection both share some basic differences with traditional computer forensics. Traditional computer forensics makes assumptions about data acquisition which do not necessarily apply to newer digital forensic areas such as live network forensics. For example:

- the evidence source is in the custody or control of the investigator during acquisition
- a complete image of the evidence source can be made in a forensically sound manner
- the evidence source is static (non-changing, fixed size)
- copies of the evidence can be verified on demand, against the original

These assumptions are based on a model where hardware and media are physically seized before data is acquired. However, when digital evidence is collected from remote network services or network traffic, seizing hardware may be neither feasible nor sensible.

A number of fundamental properties of evidence are laid out in[19], some of which include integrity, authenticity, and reproducibility. These properties are certainly desirable, but the nature of live network sources make them difficult to fully achieve. Evidence collected from live network sources has certain properties not found in evidence acquired using traditional storage media forensics. Many of these have been described previously[13][10]. These characteristics show the contrast between traditional evidence and live network evidence. They also highlight the similarities between data captured from live network sources and data collected from remote network services.

3.2 Network evidence sources

To illustrate how traffic capture and remote collection are fundamentally the same, it is useful to describe them in more abstract terms.

Captured network traffic can be abstractly described as *the preserved communication between multiple nodes on a network*. These nodes could be either local or remote. Communication is simply observed passing over the network at the point where the collection device is attached.

Data collected from remote network services can be abstractly described as *the preserved communication between two nodes on the network*. The nodes in this case are the collection/client machine itself and the remote server with which it has contact. The (induced) communication is observed by the collection/client machine while it is attached to the network.

When viewed in these abstract terms, they are both seen as acquisition methods concerned with preserving communication between nodes on the network. They both operate at various network layers to perform acquisition of data. From this perspective, what they are collecting is essentially the same.

3.3 Network evidence characteristics

The characteristics of evidence collected from these two sources is also quite similar. Evidence from these live network sources will only be available for acquisition during a finite length of time. This could be a few milliseconds for network traffic, a few minutes for dynamically updated websites, or even much longer depending on the nature of the remote network service.

Evidence collected during acquisition may come from a source outside the investigator's control or jurisdiction. For example, remote network services which are hosted in foreign countries, or network traffic in transit between two remote networks.

The integrity and authenticity of data collected from live network sources is difficult to verify. Data on the Internet is very easy to manipulate and fabricate. Finding the real origin of a particular piece of evidence, or ensuring that it wasn't manipulated in transit could be difficult especially when collecting evidence from an untrusted source such as the Internet. Here are just a few examples of activity which might influence or impact the authenticity and integrity of the evidence collected[2]:

- TCP relaying/proxying
- onion routing
- anonymous remailing

[2]The first four of these are not necessarily malicious and can have legitimate uses.

- web anonymizers
- IP spoofing
- email spoofing
- compromised third party machines
- session hijacking
- DNS cache poisoning
- other man-in-the-middle attacks

This list is not exhaustive, but serves to illustrate some of the real threats to the authenticity and integrity of evidence collected from live network sources.

The integrity of evidence *after* collection is also an issue for both traffic capture and data collection from remote network services. In many cases acquiring data from live network sources puts the collection device and its tools at risk. The collection device may be on a hostile network and must withstand malicious activity directed against it. Collection tools may interact directly with malicious remote network services. While this issue also exists in storage media forensics (virus infection for example), the risk is significantly higher when collecting data from live network sources. A system or tool compromise can put the integrity of previously collected evidence into question. Even promiscuous devices using one-way Ethernet cables have been shown vulnerable to attack by exploiting vulnerabilities in the capturing libraries (libpcap) or in packet analysis and processing code.

3.4 Acquisition architecture and tools

Network traffic capture and remote network service collection both use a similar architecture for acquiring data[3]. Both depend on a device attached to a particular network, and both run tools on that device to perform data collection. With network traffic capture, this is typically a stripped down operating system with tools using the pcap interface to promiscuously capture network packets. With remote network service collection, this is an operating system with client side software installed to perform a client/server interaction with a remote server to retrieve data.

[3]For simplicity, complex scalable and distributed architectures are not discussed here.

The main similarity here is that they both acquire data by operating on a particular network layer of a locally attached network. One tool may access the link layer, another tool may accesses the application layer. If this is generalized, we can envision a set of individual tools allowing network forensic acquisition at any layer of a network by a collection device.

4 A closer look at the differences

Collecting data from network servers seems fundamentally different from capturing network traffic. One method collects remote network data, the other collects local network data. They both seem to use different methods for collection. One uses a client/server model to actively request remote server data, the other uses promiscuous packet sniffing to passively capture data on the local network interface. However, when viewed from an evidence perspective, and a layered acquisition perspective, these differences can be overcome, and generalized definitions of live network evidence sources become feasible.

This section takes a closer look at the notions of 'passive vs active' and 'static vs dynamic' and shows how these concepts tend to lose significance, or change their meaning, when viewed in more general terms.

4.1 Revisiting 'passive vs active'

It seems obvious that live network evidence collection should fall into categories of either active or passive acquisition. Collecting evidence from remote servers clearly seems to be active as opposed to the passive sniffing of network packets.

This 'passive' definition works well for technologies such as Ethernet segments, but starts to have problems when applied more generally to other network technologies or network layers. The difficulty begins when a particular network layer of some technology requires a certain amount of active setup before passive collection can begin (PPP negotiation for example), or when the very act of being part of a network requires some sort of active participation (token passing for example). At higher layers, passive collection might also be desirable for certain network applications or services. However, passively monitoring any remote TCP-based application or service requires an initial, active TCP connection setup or handshake, as well as actively sending acknowledgment packets to maintain the connection.

This dilemma begins to sort itself out when the attribute of 'passive' is used to describe the acquisition at individual layers of the OSI model instead

of the network connection as a whole. With this approach, tools can make capture and collection exhibit passive characteristics at arbitrary network layers. Packet capturing from an Ethernet segment can be made passive starting with the link layer. Packet capturing over a PPP connection can be made passive starting with the network layer. Collecting data from a tcp based application can be made passive starting with the session layer. This can be extended to higher application layers whenever the abstract notion of 'passive' collection is needed. It is also interesting to note that if a particular layer is made passive, all layers above it will also be passive.

This abstract definition of 'passive' collection impacts the previous definition of 'active' collection. The original distinction described one as silently observing, and the other as actively requesting evidential data. With this layered view collection might be considered both passive at one layer, and active at another. This can be solved by letting the terms 'passive' and 'active' apply to individual network layers, and then using the terms 'observed' and 'induced' to describe the high level nature of collection. This brings back the original distinction between the two, without disturbing the layered approach applied to active/passive acquisition.

4.2 Revisiting 'static vs dynamic'

Another difference between data traversing networks and data available from remote network services is the length of time the data is available. Data traversing a network may exist for a few milliseconds, while data on a website could exist for years. Here, the website appears to be a static evidence source, while the network is clearly a dynamic evidence source.

This difference can be generalized by applying the concept of a variable acquisition time window to data collection. Both sources will have acquisition windows, but of different sizes. A ftp site may have an acquisition window of weeks, months or even years. A dynamic website might have an acquisition window of a few minutes or several hours. Network packets would have an acquisition window of milliseconds. However, for every live evidence source, there exists a finite, non-zero time window during which the evidential data can be observed or acquired.

Traffic capture and remote collection become similar in that they share the same attribute of having acquisition windows for data, even though the size of those windows may differ. In light of this, everything can be categorized as dynamic, and the 'static vs dynamic' argument disappears.

5 Forming some general definitions

The similarities discussed in the previous sections show the striking parallel between captured network traffic and remote network service collection. From these similarities, general definitions can be formed which apply to both areas, reducing redundancy and complexity from the current understanding of network forensic evidence sources. These generalized definitions are as follows:

- A *collection device* can be defined as a device which has access to all OSI layers of an attached network for the purpose of acquiring evidential data.
- *Collection tools* can be defined as a set of tools which run on the collection device. A collection tool acquires evidential data from one or more OSI layers.
- A *live network evidence source* can be defined as any network or network node, accessible by a collection device, which can provide data of evidential value at one or more OSI layers.
- *Live network evidence* can be defined as evidence found in data acquired from live network sources by collection tools on a collection device.
- *Live network acquisition* can be defined as the process of collecting evidence from live network sources using collection tools.
- An *acquisition window* can be defined as the finite time period during which a certain amount of live network evidence can be observed or collected.

The requirement for a collection device to have access to all protocol layers becomes important during the acquisition of data. It allows tools to acquire data at any layer. This allows network traffic capture at the link layer, as well as collecting data from network services at the application layers.

Making a distinction between a network and a network node is unnecessary when defining live network evidence sources, since all traffic captured from a network ultimately comes from various network nodes.

The acquisition window is significant. Not only does it help remove the difference between static and dynamic evidence sources, it also defines a theoretical time period during which acquired evidence might be verified against an original.

6 Examples

Having proposed some general definitions of live network evidence sources it is useful to show some examples.

6.1 Live network evidence sources

There are many different types of networks and network services which can be considered live network evidence sources. Here are some examples.

- local Ethernet segments
- Bluetooth piconets
- 802.11 (Wifi) hotspots
- database servers
- DNS and whois servers
- websites
- ftp servers
- peer to peer networks
- chat servers
- Usenet servers
- network routing tables
- responses from SNMP servers
- reply messages from SOAP servlets

From these examples, it is clear that live network evidence collection can happen at many network layers, from the link layer, up to the application layer. The generalized definitions encompass all of these layers.

6.2 Live network evidence

Here are some examples of information found at various network layers. This information could be useful as evidence or simply to aid in furthering an investigation.

- slanderous webpages, illegal files, and intellectual property abuse
- contact information for domains and IP ranges
- TLS/SSL and other cryptographic certificate info

- traffic from port scans and vulnerability scans
- routing tables, network distance (hops), ttl
- Hardware manufacturer (from the MAC address), physical distance between nodes[4]
- protocol encapsulation (raw Ethernet, 802.3, 802.2, 802.11)
- frame size, mtu, line speed, wireless signal strength and direction

There are also examples of evidence which might be found at all layers of the network model. For example, within protocol header or data fields, there may be evidence of maliciously malformed data, covert channels or protocol tunnels created with unused or unchecked areas.

7 Advantages for current and future research

Viewing live network evidence in this generalized way can be beneficial to various aspects of network forensic research. Here are some areas where a positive impact might be seen.

The need for verifying acquired evidence against an original source has proved difficult in network forensics. Having defined acquisition windows during which an original source exists could be useful in dealing with this problem.

Network write-blocking and ensuring non-modification during acquisition could also be viewed in a more general and layered way. Enforcing a read-only property at specific network layers could solve acquisition problems where network 'writes' or induced collection are required to retrieve data. This could also lead to the development of read-only or write-blocking client applications. For example, a forensic ftp client which is unable to PUT files or otherwise modify data on the remote server.

Documenting or logging the process of acquisition can be viewed in a more in-depth and standardized way across all forms of live network acquisition. For example, the collection of evidence from remote servers could be done in conjunction with local NFAT functionality to log precisely what was acquired, including lower layer protocol activity and out-of-band activity (DNS queries, ICMP, etc.).

[4]On an Ethernet segment this could be done using time domain reflectometry.

The concepts here might be extended to evidence collected through honeypots. A honeypot is simply the reverse of remote network service collection. Instead of the collection device playing the role of the network client, it would play the role of the network server.

Having a layered and generalized view of live network evidence ensures that network forensic research is applicable to networking protocols other than TCP/IP. These definitions then become transferable to other (possibly future) networking technologies.

References

[1] NIST Computer Forensics Tool Testing (CFTT) Project, Disk Imaging Tool Specification, 2001

[2] ACPO Computer Crime Group, Good Practice Guide For Computer Based Evidence, 1999

[3] US Department of Justice (CCIPS), Searching and Seizing Computers and Obtaining Electronic Evidence in Criminal Investigations, 2002

[4] Frank Adelstein, MFP: The Mobile Forensic Platform, International Journal of Digital Evidence, Spring 2003

[5] Eoghan Casey, Aaron Stanley, Tool review - remote forensic preservation and examination tools, Digital Investigation Vol 1 No 4, 2004

[6] Philip Sealy, Remote forensics, Digital Investigation Vol 1 No 4, 2004

[7] Owen O'Connor, Deploying forensic tools via PXE, Digital Investigation Vol 1 No 3, 2004

[8] Brian Carrier, Joe Grand, A hardware-based memory acquisition procedure for digital investigations, Digital Investigation Vol 1 No 1, 2004

[9] Peter Sommer, Intrusion detection systems as evidence, Computer Networks 31, 1999

[10] Eoghan Casey, Network traffic as a source of evidence: tool strengths, weaknesses, and future needs, Digital Investigation Vol 1 No 1, 2004

[11] Ren Wei, A Framework of Distributed Agent-based Network Forensics System, DFRWS 2004

[12] Yongping Tang, Thomas E. Daniels, A Simple Framework for Distributed Forensics, Proceedings of the 25th IEEE International Conference on Distributed Computing Systems Workshops, 2005

[13] Peter Sommer, Downloads, Logs, and Captures: Evidence from Cyberspace, Journal of Financial Crime, 1997

[14] Erin E. Kenneally, The Internet is the computer: The role of forensics in bridging the digital and physical divide, Digital Investigation Vol 2 No 1, 2005

[15] Brian Carrier, Defining Digital Forensic Examination and Analysis Tools Using Abstraction Layers, International Journal of Digital Evidence, Winter 2003

[16] Matthew Gerber, John Leeson, Formalization of computer input and output: the Hadley model, Digital Investigation Vol 1 No 3, 2004

[17] ITU-T Rec. X.200 (1994) — ISO/IEC 7498-1:1994, Information technology - Open systems interconnection - Basic reference model: The basic model, 1994

[18] Eoghan Casey, Digital Evidence and Computer Crime, Elsevier Academic Press, 2004

[19] Sarah Mocas, Building theoretical underpinnings for digital forensic research, Digital Investigation Vol 1 No 1, 2004

Document History

Oct 2004 - July 2005: Created original article
July 29 2005: Submitted to Elsevier
Aug 8, 2005: Accepted for publication