FLOWTRAQ®

WHITE PAPER

WHAT HAPPENED?

ENSURING YOU HAVE THE DATA YOU NEED FOR EFFECTIVE FORENSICS AFTER A DATA BREACH

Over the past ten years there have been more than 75 data breaches in which a million or more records were compromised[1]. As a result, data leakage is on everyone's radar these days. But regardless of how much money, effort and time companies are spending trying to protect their organizations' networks, completely eliminating the possibility of a data breach would mean locking up data so no one — good or bad — could access it.

But we can do better. Companies often focus on defending themselves from malicious hackers on the *outside* when a huge percent of data exfiltration happens from the *inside* out. Think of the previously trusted disgruntled employee who is secretly stealing documents, or the executive who accidentally leaves his or her laptop on the plane, or the careless employee who lets a trade secret slip to the wrong person. Because we're all human, there will always be some risk of data leakage. And one day, sooner or later, your organization will need to perform forensics after a data breach of its own.

This white paper focuses on the aftermath of an inevitable data breach and some often-overlooked steps you can take today to make sure you can respond quickly and get up and running again.

DISCOVERING A DATA BREACH

Organizations often fail to discover they've fallen victim to a data breach until weeks or more after the attack happened. And more often than not, someone on the outside discovers it first. When a breach occurs, time is of the essence because you must immediately determine if data is still leaking or whether you are unknowingly participating in an attack on another party. Both situations must be shut down rapidly, and the damage must be quickly assessed. You need to be able to answer the following questions: What happened? How bad is the damage? And who else is affected?

The challenge is that investigating security breaches can be very time-consuming. In fact, security analysts spend more time cleaning up *after* attacks than they do preventing them. Like detective work, you must try to ascertain what happened based on the evidence at hand. However, more often than not, companies don't know where to start looking. That's where your network data comes in.

Typical investigations focus on log- and host-based evidence, which really only tells you half the story. But by digging deeper and considering forensics evidence in *data in motion* (which is not often readily available), you can expedite your investigation and save your analysts many hours of sleuthing. This white paper details the critical nature of this process.

[1] Bloomberg News, Privacy Rights Clearinghouse, Breach Level Index, Feb. 5, 2015

TWO TYPES OF DATA REQUIRE DIFFERENT ANALYSIS TOOLS

After a data breach you need to collect and analyze your data immediately to begin forensics and determine what happened. Before you begin, it's important to understand the two different types of data — data at rest and data in motion — and the types of tools that enable you to analyze each.

FORENSICS INTO DATA AT REST

Data at rest is stored — in a database, on a hard drive or in the cloud — and is essentially static. Typical forensic investigations only focus on evidence of this type. For this type of data you need to determine what the perpetrators *could* have gotten their hands on. What was the nature of the attack (data leakage, data compromise, malware, a malicious link, etc.)? How widespread is the damage? Which systems and servers are affected? To answer these questions, there are a number of tools that organizations typically rely on:

Security information event management (SIEM) systems

After a data breach, it's common to start your investigative work by analyzing the logs collected by your SIEM systems. These logs may also be available on your specific host, if not exported to the SIEM, and can indicate suspicious behavior such as:

Login/failed login: thousands of failed logins and one successful one on hosts, databases and other informational assets — as well as an audit trail of who logged in on which computer when.

Virus scanner: evidence of viruses and other on-the-host compromises can indicate a possible compromise using well-known malware.

Application software failure: may indicate a more deeply rooted problem. The power of this type of evidence is often ignored. Applications or server software that typically runs smoothly, but seems to have failed inexplicably can be second-hand evidence of attacker interference. Know what to look for.

Host-based investigation after the fact

Typical next steps involve investigation of the affected hosts or servers. Be sure to:

- Identify which data was available on the systems that suffered the compromise.
- Run a vulnerability scan of the affected system, or manually search through logs of a previous vulnerability scan of those affected systems. Additionally, look for trace evidence on the host in the form of running processes or files left behind.
- When the option exists, perform full forensic analysis of all disk-based evidence of your host server using process-driven host forensic tools. While this is very time-consuming, it's extremely thorough.

Typically the above investigative approaches are the first — and unfortunately often the only — forensic steps taken, and are only part of the puzzle. They are limited by the fact that they only look at data at rest — not data in motion — so it's similar to looking at only half the evidence in a crime scene. You will not be able to tell how widespread the issue is, how much data was taken, and if there are any other affected systems — or it will take you a very long time to realize any of this.

When you're only looking at half the story, trying to get answers is much harder and takes much longer. For the other half of the story, you must evaluate your data in motion, which can help you resolve the issue much more quickly.

FORENSICS INTO DATA IN MOTION

Data in motion travels over the network. Any evidence of data communication forms the second half of the picture you need to quickly evaluate the extent of a compromise or data leakage. For data in motion you need to determine exactly what and how much data has possibly left your network, when it left, and where it went. Specifically: What *did* the perpetrators get their hands on? How much did they steal? When did the breach first happen? How long has it been going on? Where did the data come from? And more importantly where has it been sent? To answer these questions you can start by looking at the following:

Log investigation into previous data movement

Intrusion prevention system (IPS) or firewall logs may indicate misbehaving systems of a deeper-rooted cause, but they are just the tip of the iceberg when it comes to your forensics investigation. Newer network behavior intelligence tools support the logging of traffic pattern anomalies to standard syslog output, and sometimes even offer deeper forensic evidence into traffic flows.

Full packet capture traces

Full packet capture systems can provide data about your network traffic for a limited timeframe, which is dependent on storage capacity. Unfortunately, they cannot provide a long enough history to enable useful forensics on an event that started potentially weeks earlier. In contrast, network traffic flow records, such as NetFlow, are more compactly stored and can often be forensically analyzed further in the past than full packet captures.

Email traces or email logs

Email evidence is usually unavailable in the SIEM, but traces may be available on your email server. This may include solutions that flag malicious executables in email through sandboxing. But don't delay your investigation, since most organizations typically keep email records around for at most two weeks.

Data loss prevention (DLP) evidence

DLP systems are useful for catching accidental data leakage, for instance when documents were incorrectly emailed. But they are also useful for detecting malicious, and deliberate breaches. Most attackers who deliberately exfiltrate data know easy ways around DLP systems by encrypting their data. To spot potential data loss or information security breaches, consider deploying tools that can analyze network traffic flow records to gain the highest visibility about data flows into and out of your network. Filtering and anomaly detection tools can recognize hosts initiating a connection, and more importantly, hosts receiving data outside normal thresholds.

Forensic data in flow aggregators

These tools can provide intelligence about DDoS, malware and botnets that bring network availability to a crawl, but because they are flow aggregators, they only provide a sampling of the data from recent events and are unable

to provide you with the complete data you need for a full forensics investigation. Older flow aggregators were not designed to handle the ever-increasing volume of flow traffic from routers, managed switches and other network devices, resorting to sampling instead of full-fidelity data capture. Finding the proverbial needle in the haystack becomes impossible with sampling or aggregating flow collectors.

Full-fidelity forensic flow data in tools

Full-fidelity forensic tools have 1,000 times the retention capability of full packet capture systems using the same hardware. They can provide a complete forensic history of data in motion, enabling you to determine *when* the data traveled, *where* it traveled to and *from where* and *how often* it traveled, which other tools can't. Full-fidelity forensics tools provide the right balance of detail and retention history. Since flow data is so compact, it yields sufficient detail for forensic investigations and can be a very powerful tool for quickly answering the when/what/where questions, so that you can discover a leak within minutes or hours, instead of weeks.

When specifying new flow data tools, look for technology that was designed to meet requirements of large enterprises and government sector users. Look for systems that offer features such as a high-speed database architecture that enables full recall of all network flows, behavior learning, anomaly detection algorithms, and scalability through a clustering architecture. This will allow virtually unlimited traffic volumes to be analyzed.

BE PREPARED BOTH BEFORE AND AFTER THE FACT

While security analysts rely on their SIEM systems to analyze data at rest, they also need a tool that enables them to perform forensics on data in motion. Even though leading organizations like Target, Sony and the NSA had security in place, they still had significant data leakage before anyone discovered it. And without a full-fidelity forensics tool to help them clean up afterwards, they incurred huge losses in time, money, and credibility.

Here are four things you can do right away to protect your organization from data loss:

1. Add network traffic record retention. This will require a combination of high performance, full-fidelity NetFlow analysis software and multi-core servers with sufficient data storage.

2. Identify the network resources that contain vulnerable data. Whether deployed in the cloud, internal data center or desktop, key servers can be secured by collecting, storing and analyzing flow records of network traffic to and from those devices. When unusual traffic behavior is sensed, alerts can be triggered for further immediate investigation or mitigation before confidential data can be pirated.

3. Watch network traffic patterns to and from those servers to establish both normal and unusual traffic patterns. You can watch that traffic manually, subcontract it to a managed security-as-a-service firm, or aim advanced NetFlow-based analysis software at the vulnerable servers.

4. Set alerts immediately on unusual patterns. Why wait weeks or more until someone else finds a breach? At that point, it's too late.

TWO DIFFERENT FORENSICS APPROACHES

There's more than one way to perform forensics on a data breach, but like anything, some ways are more effective than others. Consider the options:

	Scenario 1 Forensics using an SIEM system	**Scenario 2** Forensics using full-fidelity forensics tool
Day 1	Evidence shows that an employee's system was compromised	Evidence shows that an employee's system was compromised
Within a few hours	Manual investigation begins; communication from this employee's computer is analyzed; you search your server for evidence	Your network intelligence tool indicates that there were thousands of login attempts and then one successful one, as well as large downloads at an unusual time (for instance, Saturday night at midnight) By analyzing flow data you can determine which computers this employee's system communicated with, where the data was sent, and how far the breach has gone You quickly discover that the same pattern exists on several other computers in the company which have all sent documents to the same IP address You conclude that multiple employees clicked on a malicious email which gave the perpetrator access to your network You isolate the systems and immediately implement mitigation steps
Weeks later	You eventually find out that that employee has a legitimate alibi Network analysts need to start over Meanwhile, the breach may still be occurring You discover that the employee clicked on a malicious email which gave the perpetrator access to the network You isolate the system and begin mitigation	

ABOUT FLOWTRAQ

FlowTraq® provides software and services for high-performance network monitoring, analytics, security and forensics to detect a range of network behaviors, including distributed denial of service (DDoS), brute force attacks, botnets, worms, network scans and other network traffic anomalies. To request a free trial of FlowTraq, visit www.flowtraq.com/trial.

FLOWTRAQ®

16 Cavendish Court, Lebanon, New Hampshire 03766
Phone +1 (603) 727-4477 | Email info@flowtraq.com