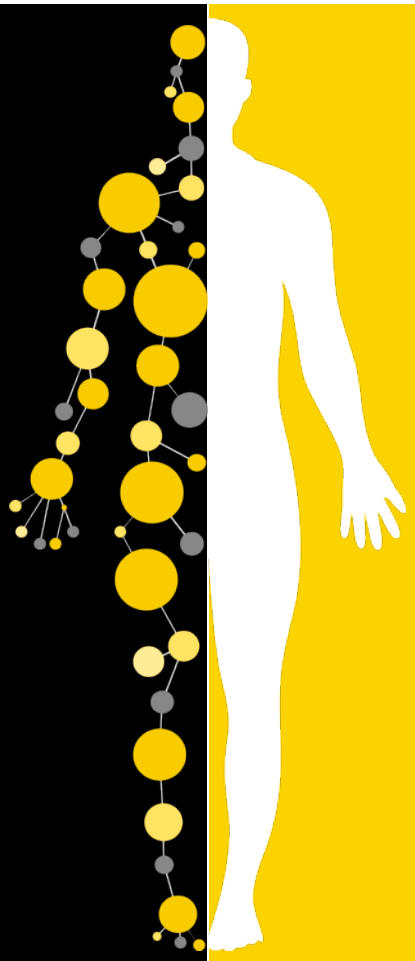


Cyber Security ≠ Data Theft Prevention

LemonFish® TECHNOLOGIES

July 2015

Contents

Summary

After two years of mega data breaches, distinguished by the release of sensitive information insiders have stolen and frequently released to the public, organizations are rebuilding their defenses. These defenses are focused on improved cyber security tools and procedures, concentrating on the patterns of data moving on the physical network. Cyber security solutions, however, only analyze the data in a superficial manner, ignoring the critical significance the data's content carries and its importance to an organization's mission, thereby failing to consider the consequences when an insider performs data theft. To tackle the appearance of an organization's sensitive data or ideas on the "dark web", and other aspects of the growing data breach problem, a new type of data theft detection is emerging. This new capability applies a fundamentally deeper level of analysis to the organization's internal data and relevant public data. This capability provides entirely new insights and defenses to an organization's security team.

Data Theft = Insider Threats + Data Breaches

This paper focuses on the data theft problem, which we define as a combination of both an insider threat and a data breach problem. While not all insider threats are related to data theft, and not all data breaches are caused by insiders, the intersection of the two creates a particularly vulnerable area where traditional cyber security fails to protect the organization. Protection against insiders exploiting or performing data breaches requires focusing on identifying important data, protecting it, detecting its loss, and tracking data exposure.

Insider Threats

There is no official definition of "insider threat". However, in their 2014 report, the National Cybersecurity and Communications Integration Center (NCCIC) said:

> *An insider threat is generally defined as a current or former employee, contractor or other business partner who has or had authorized access to an organization's network, system, or data and intentionally misused that access to negatively affect the confidentiality, integrity, or availability of the organization[']s information or information systems.*

Carnegie Mellon University's Computer Emergency Response Team (CERT) Insider Threat Center contributed to this definition and has the broadest acceptance across public and private sectors. CERT focuses on insider threats exhibiting the following patterns of behavior: intellectual property (IP) theft, IT sabotage, fraud, espionage, and accidental insider threats. However, CERT's definition is limited to information systems.

In conversations with security professionals at Fortune 2000 companies, this definition has been expanded to address everything from workplace violence, to bribes in Third World countries, and even employee turnover. To avoid any ambiguity regarding this important concept, we use the following definition of insider threat:

> *An insider threat is one or more people with legitimate elevated access, relative to the general public, who intentionally misuse that access to negatively affect the operations of the organization that granted those privileges.*

Insider threat is a broad and growing concern, yet the prevalent cyber security toolsets still focus on external threats. Nine out of ten security professionals feel that their organization is at greater risk from an insider attack than previously, and 34 percent feel very vulnerable. Less than 50 percent of organizations feel they have appropriate controls to counter insider attacks, which contributes to their sense of vulnerability.

Furthermore, the 2014 US State of Cybercrime Survey showed that almost one third (32%) of respondents said insider crimes are more costly or damaging than incidents perpetrated by outsiders. In another survey, this one sponsored by the SANS Institute, almost one-fifth (19%) of respondents believed that the potential loss from an insider attack would total more than $5 million, and over half had no idea what losses from insider threats might total.

Data Breaches

There are almost as many definitions of data breach as there are definitions of insider threat. For the purposes of this paper we use the following definition:

> *A data breach is the unauthorized disclosure of sensitive information to a party that is not authorized to have the information.*

The main differentiator among types of data breaches is the regulatory framework that guards Personally Identifiable Information (PII), Protected Health Information (PHI), and Payment Card Industry (PCI) data. The regulations associated with such data sets compel companies to monitor their security and custody to maintain compliance with Federal, state, and local laws. The financial impact and prevalence of data breaches over the past two years has highlighted the need for companies to better secure their regulated data. The personal impact of data breaches also prompts individuals to protest the lack of information security.

According to the IBM/Ponemon 2015 Cost of Data Breach Study, the average consolidated cost of a data breach increased 23% to $3.8 million between 2013 and 2015. According to Juniper Research, data breaches are expected to cost two trillion dollars by 2019. Most of this data is Personally Identifiable Information (PII), such as name, social security number, date and place of birth, although it may also include other information that can be linked to a specific individual, such as medical, educational, financial, and employment information. IBM claims one billion such records were leaked in 2014 alone. IBM's statistics are mainly from major breaches and do not include individual disclosures or small-scale events. A recent data breach at the US Government's Office of Personnel Management (OPM) included at least 21.5 million records that include PII for people in sensitive government positions.

Insider Data Breaches

While the most notorious examples of data breaches have been attributed to outside hackers, such as Anonymous or aggressive military states, insiders continue to be the leading cause of data breaches. This is precisely why organizations should put more resources and assets towards protecting themselves against insider threats. According to a survey of the Information Security Community on LinkedIn, data leaks stemming from insiders were the most concerning (63 percent of respondents). The concern about data leaks caused by insiders ranked higher than IP theft, fraud, espionage, or even IT sabotage. As shown in Figure 1, a Forrester survey shows three out of the top four ways in which companies were breached over the past 12 months were due to insiders inadvertently or maliciously misusing data, or losing a corporate asset.

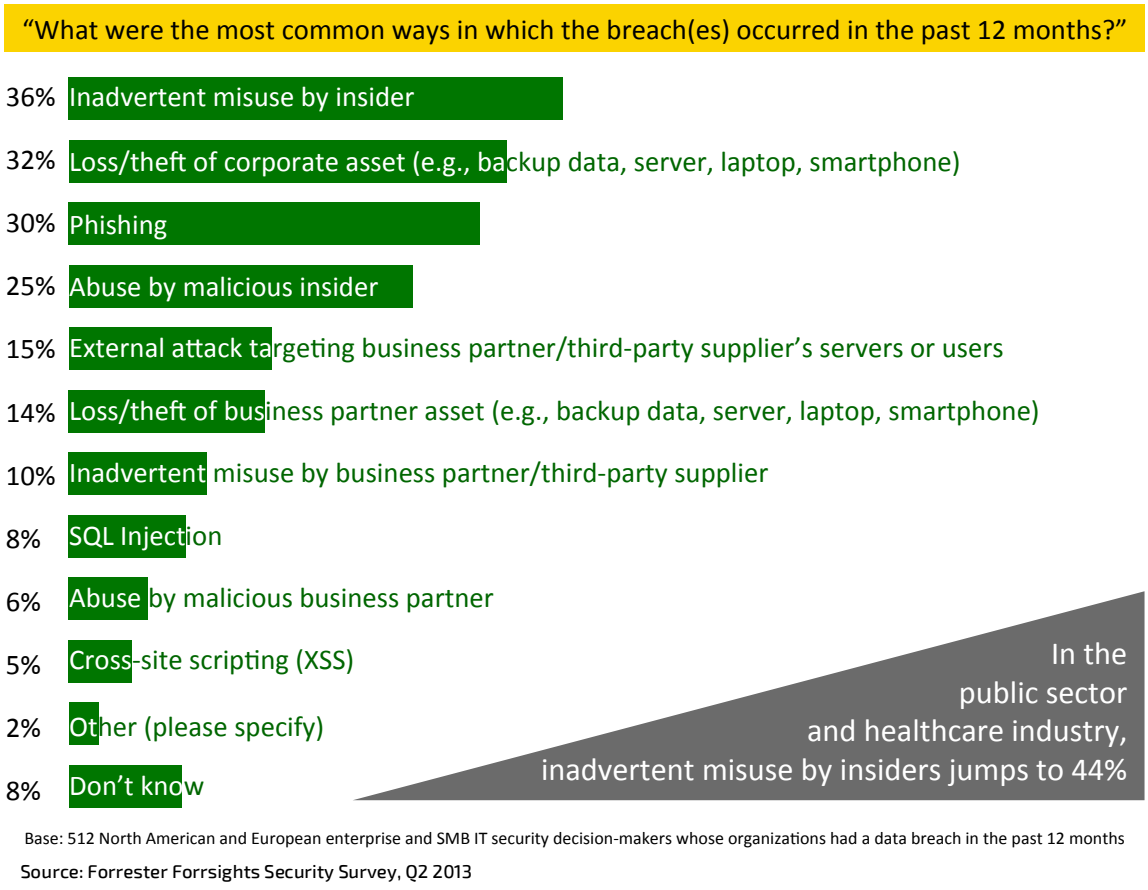


Figure 1: Common Sources of Data Breach[1]

As a function of both insider threat and data breach, data theft may be the most common and costly type of threat companies face. The Organization for Economic Cooperation and Development estimates that data breaches cost $250 billion annually. A survey by ASIS

[1] A single breach may occur in more than one category.

International, a security-industry body, estimated the annual value of stolen corporate intellectual property within the US at $300 billion. Another estimated the cost at over $1 trillion worldwide. While not all of this theft can be attributed to insiders, the sheer magnitude of the theft is undeniable. Beyond the direct financial costs, the indirect costs associated with the loss of confidential or sensitive information is high and difficult to estimate. These indirect costs include the disruption of business operations, decreases in competitive advantage, reputation damage, and exposure to criminal and civil litigation.

As shown in Figure 2, data theft targets both trade secrets and regulated data. This includes one-time theft of an email contact list and frequent and enduring exfiltration of terabytes of proprietary information. However, data theft does not include the misuse of publicly available intellectual property, such as patents, copyrights, or trademarks. While trademark, copyright, or patent infringement is a crime, it does not require an insider, nor is the issue one of stealing the data since it is readily available from public sources.

Figure 2: Defining Data Theft

Figure 2 also shows the relationship of data theft to regulated data, such as Personally Identifiable Information (PII), Protected Health Information (PHI), and Payment Card Industry Data Security Standard (PCI DSS). Loss of these types of data may not directly affect an organization's mission, but the impact of lost customer confidence, share value, or regulatory penalties can have substantial consequences. Due to its value to attackers, customer data is most vulnerable to insider attacks (57 percent), closely followed by intellectual property (54 percent), and financial data (52 percent).

Insider Data Theft = Opportunity + Capability + Intent

Insider data theft can be discussed and viewed in legal terms. Legally, a threat requires the opportunity, capability, and intent to do harm. External threats expend most of their efforts to obtain an opportunity to cause harm. Insiders, by their very nature, already have the opportunity and probable capability, and can quickly become a threat when their intent changes from benign to malicious. Since intent requires evidence of a person's internal state of mind, as intelligence analysis and courtroom cases have shown, it is difficult to either discover or prove.

Opportunity

According to a 2013 poll by Vormetric, 73% of organizations are failing to limit insider access to sensitive data. In order to reduce the number of potential vulnerabilities (the attack surface), a cyber security engineer should limit the access any component has to the rest of the system. If we think of insiders as components of a computer system, which they are, giving them unlimited access dramatically increases the human aspect of the "attack surface".

In cyber security, a "zero-day" is a vulnerability that is unknown to the general public or, more importantly, to the users and developers of a computer system. Much like the traditional use of the term "zero-day", insiders willing to use their elevated privileges to steal data are also a vulnerability that is unknown to the general public. In essence, malicious insiders are a zero-day. Given Vormetric's poll, 73% of organizations are creating a situation where any single employee can become a zero-day with the potential for unlimited damage – an enormous attack surface.

In the case where sensitive information is exfiltrated, the result is often identical whether the exploitation is by a traditional, or insider, zero-day. The only difference between the two types of zero-day vulnerabilities is the way in which the vulnerability is discovered by the thief, and the way the thief performs the attack. External hackers are well aware of the tools that help them find vulnerabilities, such as pen' testers and fuzzers. Insiders, on the other hand, are aware of ways to access information, copy it, and take it out of the organization's control. Two-thirds of companies use perimeter-focused network defenses to deny access to sensitive data, but the tools are focused on external threats, not insider threats.

Capability

Privileged users, such as managers and senior executives, not just system administrators, pose the biggest threat because they possess the greatest opportunity and capability to steal data. Although dated, a 2011 CERT report stated that of the people involved in the insider threat cases, 54% held non-technical positions, 41% held technical positions, and 5% held both. Insiders do not need to write code or employ malware in order to steal data. By virtue of their access, less sophisticated capabilities are sufficient for insiders. Although organizations may have Identity and Access Management (IAM), Security Incident and Event Management (SIEM), and Data Loss Prevention (DLP) tools, past case

history shows insiders avoid detection by each of these systems and steal data because the theft looks like a legitimate part of their daily activity.

To strengthen their defenses against hostile data access, organizations spend most of their cyber security resources controlling network access points. Yet, major advancements in Internet-scale technology continue to enable widespread computing distribution and collaboration, providing threats with many more opportunities to bypass these controls and gain access within an organization's network. In cyber security terminology, the "attack surface" grows larger every year.

Another security approach, termed "defense-in-depth", advocates layer-upon-layer of defenses. The controls in this approach must, by necessity, grow more permissive within the deeper layers of the defense, making it far less effective against insider threats because insiders cannot be treated like external parties. An insider requires extensive permissions to work effectively and efficiently; therefore, databases and file servers are inherently more vulnerable to data theft by an insider, and this is where the majority of sensitive data resides.

Case history shows insiders and outsiders generally resort to similar limited vectors to exfiltrate data:

- Portable/Removable media
- Email
- Cloud Storage
- Network access
- Reconstituted network traffic
- Exploited memory
- Printed handouts

Cyber security professionals focus on external threats to differentiate between legitimate and nefarious access. For insiders with legitimate access, these methods fail because most, if not all, of these vectors resemble legitimate, appropriate, or required means for an insider to conduct their work. This emphasizes the need to focus on analysis of the sensitive data, rather than an analysis of the patterns of computer or network activity, especially where the insider is involved.

Intent

The most common motivation for data theft is financial gain. This includes selling the data, or directly competing with a former employer. While other motives (espionage, hacktivism, retribution, etc.) do occur, insiders are rarely motivated by ideology. This means stolen data is likely to appear for sale online, indirectly when a person starts their own company or works with competitors. This observation, although it does not help prevent it, does help to quickly identify data theft.

Given a set of sensitive documents, a data loss detection system can find similar documents within the organization and across the Internet, to include hidden areas like

The Onion Router (TOR) network. Discovery of trade secrets or regulatory data can be used to identify the source of the leakage.

Selling Data

The size of a company and its industry predict the frequency and cost of a data breach. Healthcare companies are breached more than retailers. Retailers are breached more than financial institutions. According to one source, 72% of financial institutions have experienced a case of data theft by an employee in the last 12 months, the monetary value of the information being one of the main reasons for the theft.

Theft of intellectual property is hard to quantify because a variety of factors affect the value of the data. For example, healthcare information and credit card information are easy to sell because they are easily used to commit fraud. According to the New York Times, patient medical records sold for $251 at auction, while credit card records were selling for only 33 cents. This is a downward trend for the value of credit card information over the past few years. In 2013, after the Target breach, credit card numbers were selling for $20 to $100 each. This dramatic reduction may be the result of better fraud prevention at major credit card companies, which makes stolen credit cards harder to use.

Competitive Advantage

Insider knowledge is a powerful profit motive for data theft because it is marketable as a competitive advantage to industry rivals, or can be used as the core of the insider's own startup enterprises. In other cases, individuals are poached by competing companies and asked to take sensitive information when they leave. According to a survey by Symantec, half of employees who left or lost their jobs in the last 12 months kept confidential corporate data. According to a global survey by Symantec, 40 percent planned to use this data in their new job.

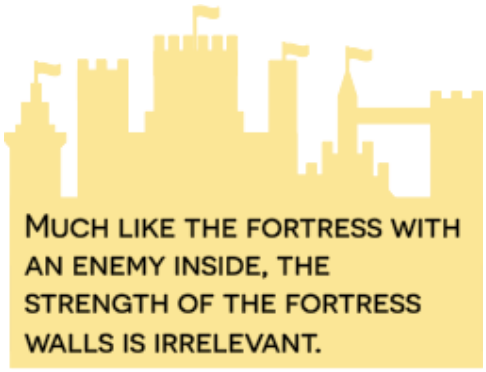
One example occurred in April 2015. Korean company Kolon Industries Inc. pled guilty to criminal charges that it conspired with former DuPont Co. employees to steal trade secrets relating to Kevlar bulletproof vests. Kolon agreed to pay $360 million to resolve the criminal and civil cases. In another case, new hires at Fitbit allegedly downloaded data about Jawbone's current and projected business plans, products and technology. Confidential information was transferred from Fitbit using USB drives and personal email accounts. Finally, Lyft filed a legal complaint alleging its former Chief Operating Officer (COO) downloaded more than 1,400 files and folders to his personal Dropbox account before leaving the company. The company claimed some of these documents contained extremely sensitive and confidential information, such as Lyft's roadmap for 2014, financial plans through 2016, growth data, and international expansion plans.

Network Security ≠ Data Security

The network is less important than the data. For an IT organization, this may sound heretical, but a well-defended network only indirectly defends the data. For example, contrast the loss of key intellectual property with the relatively trivial losses incurred by a

network outage, or the loss of a typical laptop. For a network-based company, a successful denial of service lasting several hours is far less expensive than the loss of trust and lawsuits resulting from stolen credit cards. The network will recover; there is no way to recover lost data.

If the loss of the data is more costly than the loss of network control, the defense of the data should be an organization's priority. After comparing the impact of losing control of the data to losing control of the network, it should be obvious that organizations need to change their focus and view the defense of their network as merely one, indirect means of defending their data.



For example, many organizations fail to identify the information requiring protection, only assuming it resides on their network, and the network is protected. If we assume a data focus, this view of data security is irrational. More generally, a company should perform surveillance for data theft. Regardless of the organization's confidence in their network security, if the data leaks, computer security has failed. Much like the fortress with an enemy inside, the strength of the fortress walls is irrelevant.

Insider Data Theft Detection = f(Cyber Security, Data Analytics)

Cyber security is only part of insider data theft detection. External threats have grown more and more sophisticated throughout the 50-year history of cyber security. This growing threat has led to better tools in areas like virus detection, complex event detection, and penetration testing. Tools, such as smarter static and dynamic software analyses that identify vulnerabilities, have also grown more adaptable. Despite these advances, new exploits are continually being discovered, and the last two years have been the biggest on record for data breaches.

Throughout the short history of cyber security, people have emerged as the most vulnerable part of a computer system, especially if they have privileged access. From Kevin Mitnick's "social engineering" in the 1980s to Edward Snowden in 2015, individuals with privileged access to computer systems represent the greatest risk. With the insider, there is no division between inside and outside to help technical defenses differentiate between malicious and benign activity.

Shortly after computers communicated with each other in the 1980s, people interfered with their operation. The first computer virus (i.e., Brain) was transmitted by removable media. The first wide area networks arrived, along with the first network malware (e.g., Morris Worm), and Robert Morris (now a tenured professor at MIT) was the first person convicted of a computer crime.

During the 1990s cyber security matured and the cyber security industry emerged, along with increasing malicious activity on a rapidly growing Internet. Signature-based malware detection, rule-based firewalls, rule-based internal detection, and weekly patches became the norm. Standard cyber security procedures (e.g., patch Tuesday, firewalls, encryption at rest, encryption in transit) began to emerge. The 1990s also saw the first use of computers by state and criminal organizations to obtain data. The use of software vulnerabilities and social engineering became common by the end of the decade.

In the 2000s, criminal enterprise became the primary source of malicious cyber activity, while states secretly built offensive cyber capabilities. Rather than breaking computers to demonstrate technical expertise, data was ransomed, banks were extorted, and personal information was used for blackmail. Signature and rule-based security systems remained the norm, and "big data" integration and visualization emerged to provide better awareness of cyber threats and vulnerabilities. While not ubiquitous, the DoD and industry both recognized the need for standardized "security in-depth" (e.g., CIS Critical Security Controls).

Today in the 2010s, criminal and state organizations routinely employ offensive cyber operations to gather intelligence and further their financial and political objectives. In response to these external threats, the cyber battlefield sees organizations waging minute-by-minute skirmishes to prevent "script kiddies", organized crime, and militarized cyber attackers from taking the data out of their network. New tools are emerging that focus more accurate, adaptable, and informative analytics on existing and rapidly emerging threats. As cyber security professionals realize data does not equal understanding, we should expect a movement toward smarter sensors closer to the data sources rather than the expensive movement of big data to central repositories (thereby continuing the emerging "small data" movement).

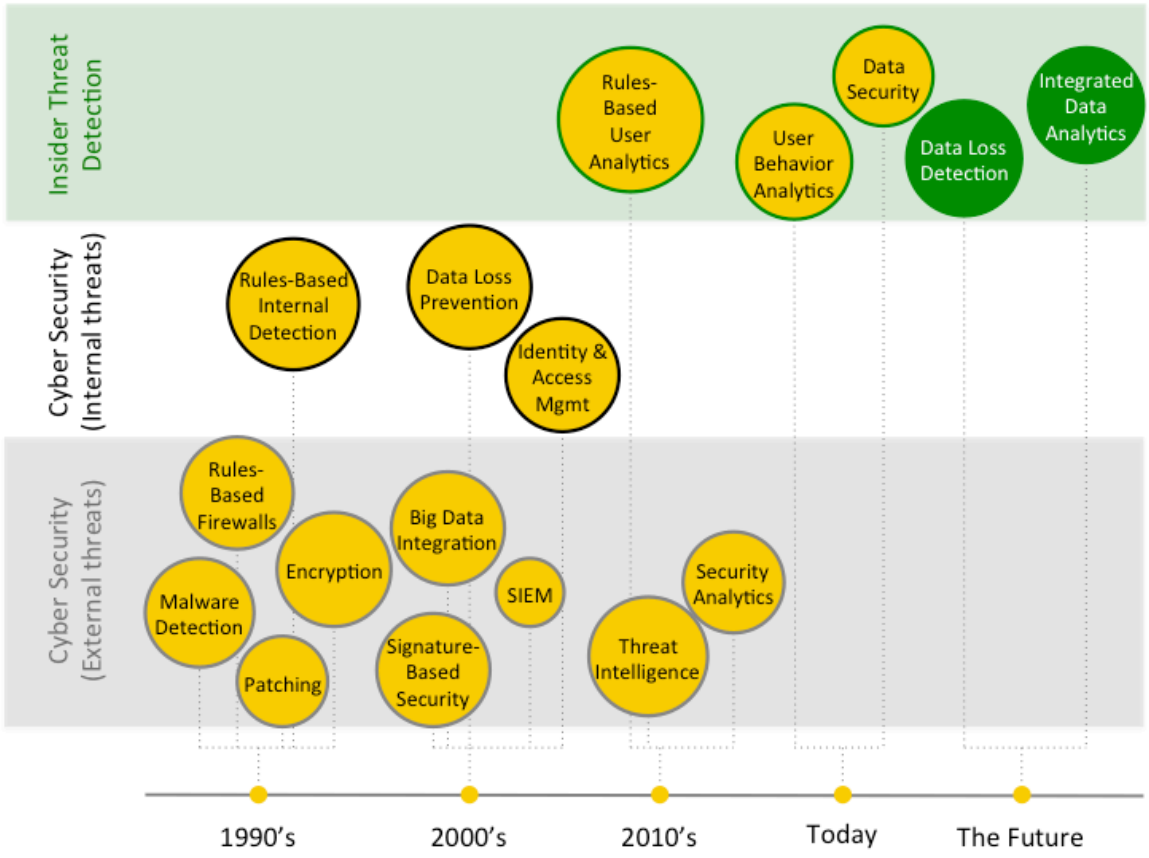


Figure 3: Cyber Security and Data Theft Timeline

As shown in the Figure 3, there are many defenses against external threats; a much smaller number have started to emerge for the insider threat. Today's defenses against the insider threat parallel the defenses used against the external threats of the 1990s, characterized by the employment of rule-based analytics much like signature-based malware detection and rules-based firewalls of the past.

The use of rule-based insider threat detection today parallels rule-based firewall security of the 1990s. Both attempt to codify normal network activity. However, "normal" is a moving target that depends on a large number of contextual variables, such as the time of day, day of the week, holiday calendar, the role of the individual, corporate events, and too many others to name. As a result, static rules create many false positives, fail to identify many abnormal events, and fail to provide sufficient context to make informed decisions once an event is flagged for review.

Data Analytics

Cyber security professionals have realized that more data does not create more security – the "small data" movement is evidence of this. In addition, static rules cannot protect against unknowns. Insider threat detection solutions must avoid shortcomings common to cyber security solutions, such as numerous false alarms and lack of context. Data theft detection must judiciously integrate and analyze events in the context of diverse data

sources. Such a system would provide a fundamentally new defense against the rapidly growing data theft problem.

Cyber security analysts do not typically possess the skills required to detect data theft.

Unfortunately, cyber security analysts do not typically possess the skills required to detect data theft. An analyst, who is probably an expert on malware, firewalls, and server configuration, is not an expert in the detection of subtle behavioral anomalies in data. Detecting the subtle differences when an insider copies data for a benign company requirement, as opposed to a malicious personal desire, requires individualized models with a context provided by non-cyber data (e.g., HR data, physical security) over extended periods of time. Building accurate behavioral models requires a background in topics like statistics, machine learning, and large-scale data analytics; topics typically studied by data scientists, rather than network administrators.

Data Loss Detection

Cyber security professionals continue to work on data loss prevention, but the task is complicated when companies fail to identify what they need to protect. This is one of the main reasons that the average time to detect a data breach exceeds 200 days. When a breach is detected, it is usually because a third party informs the company that their sensitive data appears on the Internet. The information about an organization could be considered "vulnerability intelligence" that allows criminals to identify the crown jewels of businesses and target them.

One of the most obvious keys to solving the data theft problem is recognizing not all data is equally sensitive. All data does not deserve equal protection. Unfortunately, only 17% of companies actually identify the business value of their data. While most companies know some of the documents describing their "secret sauce", companies do not generally know the location of every instance of that information within their own network, or if that information appears outside their network, even if that "secret sauce" appears in public as a result of a data breach.

Data Loss Detection should derive the core concepts from the companies "secret sauce" documents and persistently search the information across the Internet.

Data Loss Detection, the converse of data loss prevention, leverages internal expert information about critical digital assets within the organization. Data Loss Detection should derive the core concepts from the companies "secret sauce" documents and persistently search for that information across the Internet. As we described earlier, most insider data theft results in an online sale or is taken to a competitor, where it appears as part of that new company's product, so availability on public parts of the Internet is common. Data Loss Detection identifies the appearance of trade secrets or regulated data, whether online or through a competitor, and alerts the organization to the theft.

If companies were able to decrease the time it takes to handle data theft (detection, response, forensics, damage control, and mitigation), they could dramatically reduce the costs of a data breach. Performing this process rapidly may also allow the company to attribute the theft to an insider and recover costs. The current average of 200 days before the first step occurs (detection) makes preventative measures infeasible.

Integrated Data Analytics

Insider threat detection can be dramatically improved by applying analytics to an integrated view of data. Companies should leverage their investments in point-solutions, such as external social media and public record information, and best-in-breed data analytics to identify insider threats. Determining if a large data transfer to a USB drive constitutes a data theft requires more than simply examining one source of data. By itself, a single data transfer appears benign. When combined with factors found in other data sources spread across an organization's point solutions, such as a poor performance review, a sudden change to much shorter work days, and the appearance of the data from the USB drive on Pastebin.com, the additional context paints a more alarming picture. To better address data theft, organizations need to look beyond cyber security tools to other types of data. By combining data analytics with multiple data sources, data breaches can be detected earlier and handled swiftly. This prevents unnecessary litigation, financial expenditures, and loss of confidence by customers, stockholders, and the public at large.

BY COMBINING DATA ANALYTICS WITH MULTIPLE DATA SOURCES, DATA BREACHES CAN BE DETECTED EARLIER AND HANDLED SWIFTLY.

As Figure 4 shows, data theft is part of a larger problem that encapsulates the many manifestations of insider threats. Each one of these areas requires analytic solutions that help organizations enforce policies, address security concerns, identify anomalies, and make decisions to mitigate the largest risks and costs. Only integrated data analytics applied to identify the presence of sensitive data within internal and external data can accomplish this.

With the possible exception of IT Sabotage, cyber security has little to do with any of these problems. The detection of these problems requires a mathematical mindset, intelligence analysis expertise, and the engineering to bring them together. Organizations need to know where to focus their limited attention. They do not want to be inundated with false alarms devoid of context.

Figure 4: Manifestations of Insider Threat

How To Win

Data theft is an issue complicated by subtle human behavior and technological constraints. However, data theft is also a major contributor to data breaches and the resulting impact on a company's bottom line, so it must be addressed as a high priority problem. By focusing on the data itself, rather than the network and computers, technology can be applied to detect data theft.

Cyber security plays only one of many parts in *Data Loss Detection*. Addressing the broader problems requires *Integrated Data Analytics* and the application of statistics, machine learning, large-scale data analytics, and intelligence analysis. These topics are typically studied by data scientists, rather than network administrators.

Organizations must recognize that data-oriented solutions are an appropriate means for Data Loss Detection. Investing in solutions to detect when they have lost custody of their important and sensitive data, tying disparate data together, and building Integrated Data Analytics to address the insider threat are imperative in today's information economy. As long as organizations rely on traditional, external-facing cyber security, they will continue to be vulnerable to the malicious insider, and lose control of their most sensitive data. There is an alternative, and the choice is clear.

About Us

At LemonFish Technologies, we build analytics solutions using best-in-breed machine-learning techniques, deep domain expertise, and leverage the systems and data in which your company has already invested. Simply put, we eat data for breakfast and detect your insider threats and data breaches by dinner. While other companies target your people using behavioral analytics, we protect individuals' privacy while keeping your enterprise safe. We believe the data itself will set your company free. Combining an organization's internal data with external public data, our analytics detect data theft and fraud, helping your company address insider threats, regulatory compliance, and data breaches. We're LemonFish. Let's find your lemons and fish them out.

Contact LemonFish at www.lemon.fish