Cost of a Data Breach

Time to Detection Saves Real Money

Many know the adage, "time is money." It is a sensible guide for anyone with persistent stress in their professional (and personal) lives. The IBM/Ponemon Institute *2018 Cost of a Data Breach Study* is a must-read guide for chief information security officers (CISOs), chief privacy officers (CPOs), data protection officers, risk managers and boards to read carefully and plan actions and policies to manage their cyberrisk and concomitant costs when a breach occurs.[1] The report has recast the old saying as, "time is real money."

While the report points out the high cost of data breaches, there are some in the cybersecurity community who admit to not knowing if or when breach activity is happening within their networks. Some keep their heads in the sand and do not prioritize detection as a means of knowing. Recent privacy regulations require timely reporting of a data breach when known to the data controller. This may lead some to think that avoiding knowledge of a breach is a good strategy to avoid the costs associated with reporting, but this "ostrich strategy" is not only unwise and unethical, it is far costlier than meets the eye. Enterprises that invest in faster detection of data breaches can substantially reduce losses.

Costs Are Staggering

In the United States, the cost of a breach, on average, is a staggering US $3.86 million, up 6.4 percent from the prior year's analysis.[2] These statistics should be of real concern. And this is just the tangible costs of a breach. This amount does not account for the costs of regulatory fines. Hence, the numbers seen in this report must be considered a lower bound. Other potential costs include expenses associated with litigation, fines for regulatory noncompliance and lost revenue due to customers who leave as a result of a breach. With all the investment in cybersecurity products and professional services, why have the costs of data breaches not decreased?

Each new data privacy regulation, such as the EU General Data Protection Regulation (GDPR) or the US state of California Consumer Privacy Act (CCPA), calls for "appropriate security measures" to avoid fines, but by relying on tools that are easily susceptible to compromise via employee error and stolen credentials, enterprises are failing to protect themselves adequately.

The most vexing aspect of these new regulations is that the liability for data lies with the originating organization. This means that an organization is responsible for the security of that data regardless of with whom the data are shared and how the data are handled once they are outside of the originating organization's control. The contractual obligations of a third-party provider to secure data once it has access may not be worth the proverbial paper on

Salvatore J. Stolfo, Ph.D.
Is a tenured professor at Columbia University (New York City, New York, USA). He has been teaching computer science since 1979. He is the cofounder and chief technology officer of Allure Security (*http://www.alluresecurity.com/*), a DARPA-funded cybersecurity startup specializing in data protection and the detection of data breaches. Stolfo has been granted more than 75 patents and has published 230 papers and books in the areas of parallel computing, artificial intelligence knowledge-based systems, data mining, computer security and intrusion detection systems. His research has been supported by numerous US government agencies, including DARPA, the US National Science Foundation (NSF), Office of Naval Research (ONR), the US National Security Agency (NSA), Intelligence Advanced Research Projects Activity (IARPA), the US Air Force Office of Scientific Research (AFOSR), Army Research Office (ARO), the US National Institute of Standards and Technology (NIST), and the US Department of Homeland Security (DHS).

which it was printed. GDPR, for example, pierces any such agreements and puts the entire onus on the originating organization as the responsible party for any data losses, to the tune of up to 4 percent of total revenues or €20 million per violation. Perhaps "reasonable" best practices are not good enough. Perhaps cybersecurity technology is not being used to tackle the right problem.

Better Detection and Response

The average mean time to identify a breach is now 197 days.[3] This is a slight improvement over the previous year, but still unacceptably long. Imagine living with a thief who is taking everything of value for 197 days.

All data breaches, no matter the size and depth, result in money lost due to time and money spent during the investigation process, reputation management, lost customers, fines, etc. Of most interest and most important in the report's analysis is the revelation that "Companies that contained a breach in less than 30 days saved over (US) $1 million vs. those that took more than 30 days to resolve."[4] This is a staggering statistic. How do those organizations detect a breach and mitigate it so quickly? Why has everyone not followed suit? How might breaches or attempted exfiltration be detected before it occurs?

Consider the typical behavior of attackers. They follow a tried and true methodology: reconnaissance to locate a target and points of entry, followed by initial entry (typically by spear phishing users within their chosen target). Once they succeed at stealing a legitimate user's credentials, they set up a foothold to establish control for long periods of time. They move laterally, searching for other credentials, servers, logs, files and documents they desire. Documents and data are bundled and exfiltrated using popular protocols and third-party sites that serve as staging servers from which they can download their stolen goods. Having established a foothold allows for long-term data exfiltration at a pace that suits their needs.

Most deception approaches miss the mark on early detection. A variety of deception technology companies tout honeynets as early detectors. This actually is unlikely to be the case. Honeynets are installed alongside operational servers, and attackers would be attracted to those honeynets only if they are led to them. Care is taken to prevent ordinary users from connecting to honeynets; otherwise, false alarms will adversely impact business. And, by the time an attacker has fallen into a honeynet, he or she must already have searched through the operational network and likely already stolen his or her quarry. This approach is ineffective for detecting and responding to data loss. There are emerging sensor technologies that offer a better way and are deployed directly within an operational environment. For example, data loss sensors can be automatically generated, highly believable decoy documents with embedded beacons that are strategically placed in folders, directories or third-party shares to entice attackers to open them and alert security teams to early-stage breach activity.

"BY THE TIME AN ATTACKER HAS FALLEN INTO A HONEYNET, HE OR SHE MUST ALREADY HAVE SEARCHED THROUGH THE OPERATIONAL NETWORK AND LIKELY ALREADY STOLEN HIS OR HER QUARRY."

Studying the Life Cycle of Masqueraders: DARPA's Experiment

The lack of large-scale, real-world research data has hindered the development of effective intrusion detection systems (IDSs) that can stop an attack early in its life cycle. Most organizations that experience these types of attacks prefer not to announce them publicly out of liability and confidentiality concerns. Seventy-two percent of the insider incidents that occurred at the surveyed institutions were handled internally without legal action or the involvement of law enforcement. Another 13 percent of the insider incidents were handled internally with some legal action.[5] Announcing such attacks may also have market-share implications. For the same reasons, breach victims are even less likely to share real-world data that could be used to study such attacks with the research community.

The study of masquerade attacks, a class of attacks in which an outside adversary illegitimately poses as, or assumes the identity of, a legitimate user, suffers similarly from the scarcity of real-world data, despite their significance. Thirty-five percent of executives and law enforcement officials experienced unauthorized access and use of their information, systems and networks.[6] Masquerade attacks were second in the top-five list of electronic crimes perpetrated by outsiders after viruses, worms and other malicious attack vectors.

In an effort to generate more real-world research data, the Defense Advanced Research Projects Agency (DARPA) initiated the Active Authentication program several years ago with the mission to further explore data loss detection approaches. The key driver was to understand the vulnerabilities of the last mile of an attack, with the hope of improving security in this area. The focus of this research was to protect the data itself within its native setting—namely, business documents such as Word documents, PDFs and spreadsheets. Researchers were determined to provide an effective means of early detection of masqueraders.

The key insight from this experiment: Masqueraders, much like thieves breaking into a home, must gather information about the environment they just entered. They must learn more about the home, search for valuables to steal and package data to exfiltrate. This early-stage activity is key to detecting their nefarious activity. Therefore, placing sensors that act as trip wires within the folders and directories that are most likely to be searched as part of an attack serves as a detection mechanism.

The efficacy of the use of sensors embedded in documents as a means to detect data loss was proven in a scientific study sponsored by DARPA. The study measured true and false positives and true negatives. In the study, 39 individual masqueraders, all selected because of their sophisticated, deep knowledge of computer science and systems in general, were granted access to a system as if they had already succeeded in stealing the necessary credentials. The masqueraders were then told that their job was to steal sensitive information in the system leveraging the credentials they were provided. The study followed prescribed and statistically valid methodology whereby the masqueraders were given a scenario of what kind of information to steal, but not how to steal it. They were left to their own devices to find and exfiltrate their quarry.[7]

In this study, the strategic use of sensors to detect when documents had been accessed was successful in:

- Detecting 98 percent of masqueraders
- Generating only one false positive per week of operation
- Achieving detection within 10 minutes

Figure 1 details the measured time to detection for each masquerader. None had sufficient time to succeed in exfiltrating the files they acquired and were prepared to bundle and exfiltrate. This is solid evidence of the efficacy of the concept of applying sensors to documents as a way to dramatically reduce data losses and the associated costs. Considering the average dwell time of an adversary, detecting irregular behavior of a masquerader within 10 minutes is a significant improvement.

> “MOST ORGANIZATIONS ALREADY RELY ON NETWORK AND ENDPOINT SENSORS TO COLLECT CRITICAL SECURITY DATA, SO WHY NOT RELY ON SENSORS AT THE SOURCE OF THE DATA ITSELF?”

Early detection of unsanctioned access to sensitive data, such as the scenario in this DARPA research, can spare organizations the pain and cost of a breach. There is a 28 percent chance an enterprise will suffer a breach within the next two years.[8] This can result in a US $3.6 million loss based on the

Figure 1—Minutes Elapsed Since Masquerade Attack Start

average cost of a data breach. Using these numbers as a benchmark, detecting a breach within 30 days or less (rather than the average 197 days) saves the organization US $1 million in costs. Most organizations already rely on network and endpoint sensors to collect critical security data, so why not rely on sensors at the source of the data itself? If time to detection is the key to saving organizations significant amounts of money, then the security industry must examine new technologies and approaches to improve this metric. The use of data loss sensors is a proven method to catch masqueraders quickly and early in the attack life cycle.

Endnotes

1 IBM and Ponemon Institute, *2018 Cost of a Data Breach Study*, USA, 2018, *https://www.ibm.com/security/data-breach*
2 *Ibid.*
3 *Ibid.*
4 *Ibid.*
5 *CSO Magazine*, "2010 Cybersecurity Watch Survey—Survey Results," 25 January 2010, *https://resources.sei.cmu.edu/asset_files/News/2010_100_001_53454.pdf*
6 *Ibid.*
7 Salem, M. B.; S. Stolfo; *On the Design and Execution of Cyber-Security User Studies: Methodology, Challenges, and Lessons Learned*, Columbia University, New York, USA, *http://ids.cs.columbia.edu/sites/default/files/CSET_2011_0.pdf*
8 *Op cit* IBM and Ponemon