proofpoint

Big Data Solutions to Enterprise Data Security Challenges

A Proofpoint White Paper

Contents

Executive Summary 3

Introduction 4

The Evolution of Inbound Security Threats 4

- Winning the Recipient's Trust 5
- Advanced Persistent Threats 5
- The Difficulties of Threat Detection 6

Big Data: A Revolution in Data Analysis 7

Proofpoint Security–as–a–Service 8

- Proofpoint Threat Detection and Big Data 9
- Proofpoint Anomalytics in Action 10
- URL Clicktime Defense Service 10

Conclusion 11

Executive Summary

Enterprises today are exposed to an ever-broadening range of IT security threats, from basic annoyances such as auto-emailed viruses, to targeted phishing-style attacks that trick employees into clicking on dangerous links that install malware, steal credentials, or in some other way jeopardize the security of the enterprise.

What's particularly troubling is that the trendline is in the wrong direction, towards these low-volume, highly targeted, and very dangerous attacks. Traditional IT security approaches that rely on pattern-matching—including whitelist/blacklist comparisons, content scanning, and reputation systems—are all-too-often incapable of protecting enterprises from these customized attacks. And when these attacks succeed, they often lead to the theft of an enterprise's most valuable intellectual property, including customer records, product designs, business plans, and more.

Fortunately, there is a new approach to threat detection and remediation. Instead of pattern-matching against "known bad" email, Proofpoint's security-as-a-service platform applies Big Data analysis techniques to continuously analyze billions of email messages and ever-changing patterns of communication, enabling Proofpoint's security-as-a-service platform to detect anomalous behavior – effectively finding deviations from "known good" email flow –and block inbound security attacks in real time. Using Big Data techniques, Proofpoint's new Targeted Attack Protection service delivers comprehensive security against even the most elusive inbound threats, such as phishing attacks designed to steal corporate data. Proofpoint's Big Data solution gives enterprises the data security they need to withstand today's highly customized inbound attacks.

Introduction

For email administrators and IT security teams surveying recent developments in enterprise IT, two major transformations stand out.

First, inbound security threats are becoming much more sophisticated and dangerous. Phishing attacks and other highly targeted, low-volume attacks are taking the place of conventional high-volume spam. Why are attackers changing their tactics? Because their goals have changed, as well. No longer content merely to disrupt IT operations or to persuade naïve users to buy counterfeit goods, hackers and criminal syndicates are now designing attacks to gain access to an organization's most valuable intellectual property. The new strategy and tactics are working—attackers are catching many large enterprises and mid-size companies off guard.

Second, computer science itself is making dramatic advances, particularly in the area of data storage and data analysis. Because of the proliferation of Web technologies and mobile devices, enterprises find themselves managing more data than ever before. To make sense of this data deluge, software developers are leveraging "Big Data" analysis technologies such as Apache Cassandra and Apache Hadoop. Enterprises can apply this technology to analyze customer trends and manufacturing yields—and Proofpoint (NASDAQ:PFPT), a leading enterprise security vendor—can use it to detect and thwart the new forms of cyber attacks being wielded against enterprises today.

The paper will consider both these new developments—new, highly targeted inbound security threats, and the ability of Proofpoint's security-as-a-service platform to leverage Big Data techniques for detecting and stopping these attacks with unprecedented speed and accuracy.

Let's start by examining the changing nature of inbound security threats.

The Evolution of Inbound Security Threats

First, some good news: spam volumes are declining. They dropped 68% year-over-year between February 2011 and February 2012.[1] In a year-end report in 2011, IBM estimated that volumes had returned to their 2008 levels. Other estimates suggest far more dramatic declines: from an all-time daily high of 225 billion messages down to 25 billion—a drop of nearly 10x. There's no doubt that law enforcement's take-down of two major spam botnets—Rustock and McColo—has contributed to these lower volumes.[2] But other forces are at play as well. And that's the bad news.

Spammers, criminal syndicates, and other hostile entities are switching their attention from conventional high-volume spam to carefully crafted, low-volume attacks that are more sophisticated, devious, and potentially costly. Instead of peddling counterfeit drugs or luxury goods, they're launching phishing attacks, distributing email with fake links or malware attachments (such as keyloggers and rootkits) that enable criminals at remote locations to surreptitiously siphon or "exfiltrate" valuable business data from an enterprise network. Phishing might capture the login credentials for a bank account. Or the login credentials to an internal database of customer records. Or the SharePoint credentials to a document repository with industrial secrets such as product designs or medical patents.

[1] See the March 2012 Proofpoint Threat Report for details.
[2] http://www.theemailadmin.com/2012/03/declining-volumes-means-spam-in-transition/

Confidential data like this can be worth millions of dollars on the black market. It promises a much bigger payoff than hawking sugar pills branded as Viagra. And it's of much greater interest to the organizations perpetrating these attacks: foreign governments, industrial spies, and criminal syndicates working in the black market for stolen intellectual property.

Winning the Recipient's Trust

The key to perpetrating one of these low-volume attacks is winning the email recipient's trust and luring him or her to click on a link. The link often leads to a bogus Web site that collects login credentials or installs malware. Once malware is installed on the recipient's system, it might leapfrog onto other systems on the internal network, working its way to the most valuable data repositories.

How do phishing messages win the recipient's trust? Attackers typically use either of two approaches: forging corporate identities and social engineering.

Forging Corporate Identities

Attackers might send a phishing message with graphics and text designed to look like an official message from a trustworthy source such as the recipient's bank. Because the message comes bearing a corporate logo and an official-sounding message about a promotion or an account error, the recipient is fooled into assuming the message is legitimate and clicks on one of its links. That link might lead to a fake site—again branded with the artwork of a real institution—where the recipient logs in, inadvertently delivering his or her login credentials to fraudsters. Or the link might lead to a malware-infected site that downloads a rootkit or other malware onto the recipient's computer.[3]

Social Engineering

The other type of attack relies on social engineering. A message is more believable if it appears to be from a friend or colleague, is addressed to friends or colleagues, and refers to special topics that the recipients are genuinely interested in. For example, if a manufacturing manager receives a message that appears to be from one of her vendors warning about a parts shortage, the manager is likely to click on any links that message contains. Another successful ploy: emailing a malware-infected spreadsheet in the guise of financial reports, HR reports, or Web statistics. Some of these messages appear to be sent through social media sites such as Facebook, thus combining both the tactics of forged corporate branding and social engineering. Social networking sites such as Facebook provide attackers with a wealth of information for putting together these bogus messages. Through Facebook profiles and status updates, through LinkedIn profiles and tweets, business users regularly broadcast their interests, their whereabouts, and the identities of their friends, giving attackers a wealth of information to draw upon.

Advanced Persistent Threats

Another change in the threat landscape is the rise of the Advanced Persistent Threat. Installing malware or stealing credentials isn't necessarily part of a quick, hit-and-run-style attack. Instead it might be the first step in a stealthy long-term attack that persists for months or years.

"We surveyed the field, did our homework, had all the vendors in, made our decision grid and Proofpoint came out on top, as it had in our two earlier evaluations."

— **Kreigh Tomazewski,**
Sr. Technical Support Specialist
Postmaster, Alticor Inc.
parent company of Amway

[3] The growing popularity of this type of attack is reflected in the changing role of graphics in spam messages. Five years ago, images would have depicted the cheap drugs or luxury goods being marketed. Today, they're much more likely to be corporate logos and other graphics that are part of an organization's brand. Source: IBM X-Force 2011 Trend and Risk Report.

An Advanced Persistent Threat (APT) is an organization that launches stealthy, long-term attack usually against a specific business or government agent with the purpose of stealing information.[4] An APT might launch a phishing attack against an organization with a few highly targeted email messages. When a recipient clicks on a link it might install malware such as a key-logger or another piece of malware that remains dormant for several days or weeks. The malware might monitor system or network activity to discuss a particular system to target. Then it might surreptitiously target that system—usually not to bring it down but rather to steal its information. It might then "exfiltrate" that data over a secure FTP or HTTPS channel.

The goal of the attack is not to cause havoc by bringing down servers or networks. Rather, the goal is the theft of intellectual property, such as product designs or customer records. No longer are customer records and credit card data at risk: some APTs are run by foreign governments interested in trade secrets, military secrets, and other forms of espionage.

Perhaps the most famous APT attack was the infiltration of the IT security vendor RSA, which the company revealed in March 2011. Attackers breached the RSA's network security measures and managed to steal perhaps the company's valuable intellectual property: data about RSA SecurID two-factor authentication.[5]

But the criminals who infiltrated RSA infiltrated other organizations as well—around 760 organizations according to one source—and began quietly extracting confidential data from some of these organizations as long ago as 2010.[6] These attacks were launched and controlled by 338 command-and-control networks, the majority of which were based in or around Beijing, China. Almost certainly, the hackers behind these attacks were not the same criminal syndicates based in Russian and Eastern Europe who have been responsible for the majority of spam over the past decade or so.

In another attack now known as "Night Dragon," hackers working regular business hours in China surreptitiously stole information from executives at leading energy companies over a period of four years. Every aspect of the attack was stealthy: no taunting messages appeared on screens, no servers strained under surges of email. Instead the attack relied on an occasional, almost impossible-to-detect trickle of highly privileged information leaving a few dozen laptops in encrypted tunnels terminating at servers managed by parties unknown. The contents likely included sensitive information about unannounced discoveries of natural resources, development plans, and financial returns.[7]

Former counterterrorism czar Richard Clarke believes every major company in the U.S. has been attacked by data thieves.[8] The risk posted by Advanced Persistent Threats is real.

The Difficulties of Threat Detection

The challenge with phishing and APT attacks—especially with attacks that rely on social engineering—is that they're extremely difficult to detect. Most spam filters rely on telltale attributes, such as high message volumes, senders' domains known to be bad, virus or malware signatures, or spam keywords. None of these attributes apply to low-volume

[4] http://www.schneier.com/blog/archives/2011/11/advanced_persis.html
[5] http://www.eweek.com/c/a/Security/RSA-Warns-SecurID-Customers-of-Data-Breach-395221/
[6] http://krebsonsecurity.com/2011/10/who-else-was-hit-by-the-rsa-attackers/
[7] http://www.networkworld.com/news/2011/021011-night-dragon-attacks-from-china.html
[8] http://www.smithsonianmag.com/history-archaeology/Richard-Clarke-on-Who-Was-Behind-the-Stuxnet-Attack.html?c=y&page=4

phishing attacks. The messages appear to be arriving from a reliable source—perhaps even from the recipient's own domain. The messages aren't part of a deluge of messages coming from a botnet. They don't contain spam keywords like "Viagra," "pharmaceuticals," or " discount." In most of their particulars, these messages are identical to the legitimate messages that friends and colleagues exchange every day. There's no blacklist of known signatures for these attacks; they're individually crafted. In fact, their elusiveness derives from their individuality.

Worse, the URLs in the email may in fact be benign at the time of delivery. Many attackers "pulse" malware, loading it only for a few minutes each day, or post "site busy / try again later" plain HTML pages, only converting to credential attacks periodically. No conventional anti-SPAM or protection system will stop such an inbound email, as the email isn't "bad" yet. It only "becomes bad" after passing the gateway and entering the enterprise.

How can enterprise data security solutions stop highly targeted attacks they've never seen before? How can they tell a legitimate message from a colleague from a nearly identical message purporting to be from the same colleague? How can they detect an email that "becomes bad" at some random point in the future?

Big Data: A Revolution in Data Analysis

To learn how Proofpoint detects and stops these stealthy low-volume attacks, let's examine the other major IT development mentioned earlier: Big Data.

The phrase "Big Data" refers to the storage and analysis of data sets so large that they would be difficult or impossible to manage with traditional database technology. New Big Data technologies for caching data and distributing analysis workloads across hundreds or thousands of commodity processors enable organizations of all kinds to process unprecedented amounts of data quickly and affordably.

Where does all this data come from? The volume of business data is exploding as more business operations move online, trackable Web transactions replace offline or antiquated client-server transactions, and the number of end user devices multiplies into the billions. McKinsey estimates that by 2009, the typical organization with 1,000 or more employees had 200 TB of data.[9] In 2010 global businesses stored more than 7 exabytes (millions of terabytes) of new data on hard drives, while consumers added 6 exabytes of new data to their personal computers and peripherals.[10] Web-scale applications like Twitter generate tens of TBs of data every day.[11] In some industries, data volumes are growing exponentially every few years.

It would be unwieldy and prohibitively expensive to manage TBs of changing data in conventional databases. To manage this data, organizations of all sizes are turning to new data storage and analysis technologies, many of which were created internally by companies like Facebook and Yahoo! specifically to address the problem of collecting, storing, and process enormous volumes of Web data. The Cassandra data store, for example, was originally developed in-house by Facebook to manage its mailbox search feature.

[9] http://www.mckinsey.com/Insights/MGI/Research/Technology_and_Innovation/Big_data_The_next_frontier_for_innovation
[10] McKinsey Global Institute, cited in the McKinsey publication listed above.
[11] http://wikibon.org/wiki/v/Transcript:Kevin_Weil,_Twitter_Analytics_Lead._at_HadoopWorld_2010,_on_Open_Source_tools

One of the most popular Big Data technologies is Hadoop, an open source data management solution that enables organizations to store and analyze vast amounts of both structured and complex data. Originally developed by a programmer named Doug Cutting and now an open source project managed by the Apache Foundation, Hadoop comprises the Hadoop Distributed File System—a scalable, distributed file system designed to large numbers of commodity servers—and a parallel distributed processing framework that uses an approach developed by Google called MapReduce. MapReduce performs "domain decomposition": it divides and distributes a large problem set across tens, hundreds, or thousands of commodity processors, then collects the distributed answers and reduces them to a single, coherent result.

Proofpoint is now using Hadoop and other Big Data technologies to address the challenges of detecting and defeating inbound IT attacks, such as phishing attacks and other activities launched by Advanced Persistent Threats.

Proofpoint Security-as-a-Service

A quick introduction to Proofpoint: Proofpoint is a pioneering security-as-a-service vendor that enables large and mid-sized organizations worldwide to defend, protect, archive and govern their most sensitive data. Proofpoint's security-as-a-service platform comprises an integrated suite of on-demand data protection solutions, including threat protection, regulatory compliance, archiving and governance, and secure communication.

To address today's rapidly changing threat landscape, Proofpoint solutions are built on a flexible, cloud-based platform and leverage a number of proprietary technologies, including Big Data analytics, machine learning, deep content inspection, secure storage and advanced encryption.

The Proofpoint security-as-a-service platform addresses enterprise IT security by protecting data as it flows into and out of the enterprise through on-premise and cloud-based email, instant messaging, social media and other web-based applications. In addition, the Proofpoint platform securely archives these communications for compliance and eDiscovery.

Proofpoint solves four important problems for mid-sized companies and large enterprises:

- Keeping malicious content out;
- Preventing the theft or inadvertent loss of sensitive information and, in turn, ensuring compliance with regulatory data protection mandates;
- Collecting, retaining, governing and discovering sensitive data for compliance and litigation support; and
- Securely sharing sensitive data with customers, partners and suppliers.

A new service included in the Proofpoint platform, Proofpoint Targeted Attack Protection™ deploys an array of advanced technologies including Big Data analysis techniques, URL interception, and malware sandboxing to provide unprecedented protection that follows messages and users wherever they go – whether they're behind the corporate firewall or off the corporate network, on mobile devices, or public terminals.

ACID vs. BASE

One difference between Big Data technologies and traditional databases is that traditional databases adhere to a model known as ACID, which stands for Atomicity (a transaction is "all or nothing"), Consistency (transactions invariably occur according to rules), Isolation (transactions must not interfere with one another), and Durability (transactions are stored even in the event of IT outages).

To achieve the scalability and flexibility needed for vast data sets, Big Data technologies like Hadoop dispense with the rigor and reliability of ACID in exchange for a more flexible model that scales dramatically while proving to be reliable enough for applications like social network updates and network analysis and threat detection. BASE—the name of this alternative model—stands for Basically Available, Soft state, Eventual consistency. At any given moment, a Big Data store might not have perfectly consistent data caches, but the caches will become consistent eventually, and they will do so quickly enough to assure that IT operations are effective and reliable.

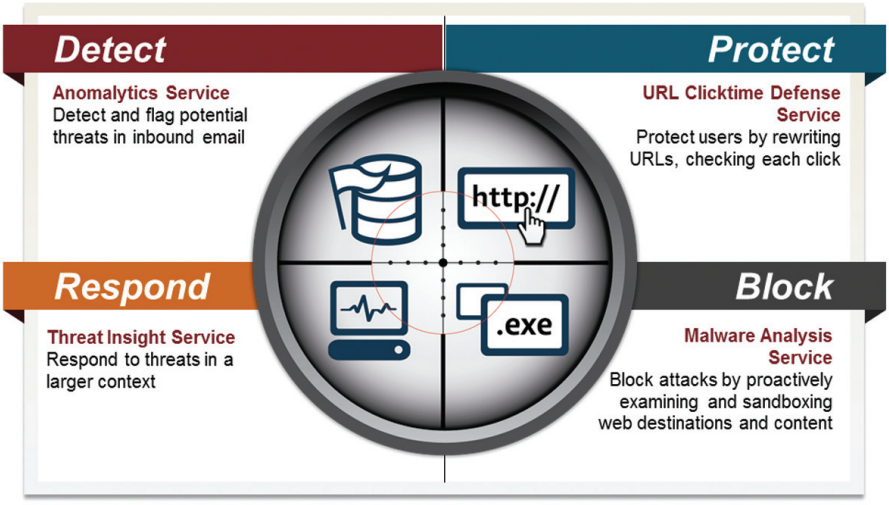
By switching to a BASE data model and completing SQL queries with NOSQL (Not Only SQL) queries, Big Data applications can effectively process huge data stores and gets timely answers to questions that would have been impossible to ask only a few years ago.

Proofpoint Threat Detection and Big Data

To detect and stop inbound attacks, Proofpoint applies Big Data analysis techniques to scrutinize millions of email messages every day. Through continuous in-depth analysis that involves patented machine-learning techniques, Proofpoint is able to identify circles of trust for individual users, noting whom they usually communicate with, how many messages they typically send, and when and where they usually communicate, and what type of content they usually convey. Through this continuous, real-time analysis, Proofpoint models "normal flow" for every individually protected mailbox, group, and organization.

Once it has established norms for users, groups, and organizations, Proofpoint can detect message anomalies in real time and stops low-volume phishing attacks even before they strike. Without relying on whitelists ("senders or content known to be good") or blacklists ("senders or content known to be bad") —which are unavailable for this type of attack— and instead focusing on anomalies ("things that are different from the known good"), Proofpoint defeats low-volume attacks and protects enterprise networks safe from Advanced Persistent Threats.

Because it's so useful at detecting low-volume phishing attacks through anomalies, Proofpoint calls this Big Data analysis of communication norms and anomalies, Anomalytics.[12] Proofpoint Anomalytics is now available as part of Proofpoint's Targeted Attack Protection suite.



Proofpoint Targeted Attack Protection Cycle

[12] http://wwwyoutube.com/watch?v=vZxk35UgL50&feature=player_embedded

Proofpoint Anomalytics in Action

Let's take a closer look at how Proofpoint Anomalytics detects and stops attacks.

Imagine a large enterprise with hundreds of thousands of employees. Every day, this organization will send and receive millions of email messages. Simply scanning those messages for conventional threats like malicious attachments is an enormous task. To protect that enterprise against low volume phishing attacks, the Proofpoint platform needs to understand what's normal for that organization. This requires analyzing traffic patterns in aggregate—company-wide, for example, what's a typical email volume per day, per hour, and at a specific time of day—as well as traffic patterns for individual users.

If on a typical business day a finance manager typically receives 100 messages and typically sends 20 messages, all during business hours, and those messages are typically sent to only 1–3 users, then Proofpoint can immediately notice something unusual about that same user sending hundreds of messages in a five-minute burst, particularly if that user is sending messages to recipients he or she doesn't normally send to (for example, a remote engineering organization), and those messages are being sent in the middle of the night. This anomalous behavior is a sign that almost certainly an attack is under way. Anomalytics policies configured by the IT team determine whether such messages should be simply deleted or quarantined for future review by IT security specialists.

This sort of analysis—scrutinizing hundreds of attributes of millions of email messages in time to take action—would have been extraordinarily resource-intensive and expensive before the advent of Big Data technologies.

Because Proofpoint provides security services for a larger number of domains—approximately 2,400 enterprise customers as of this writing, including many Fortune 100 companies—Proofpoint customers benefit from the "network effects" of a cloud-based system. Intelligence about the attacks affecting one customer can be applied to other customers. For example, upon detecting an attack from a domain registered with a dubious registrar in the past 24 hours, the Proofpoint platform can immediately can block or quarantine traffic from that domain for all its other customers. Through Big Data analytics and machine learning, the Proofpoint system gets smarter as it works, delivering ever-rising degrees of protection.

URL Clicktime Defense Service

To defend against spear phishing and advanced persistent threats, an IT defense solution must be able to distinguish potentially hostile URLs from safe URLs included in email.

To escape detection by A/V filters, the phishing messages used in targeted attacks often don't contain any malware attachments at all. Instead, the messages encourage the recipient to click on a link – and then the resulting Web page either attempts to download malware (a drive-by download) or trick the user into entering credential information (username and password, and/or private information such as Social Security number).

Detecting hostile links becomes even more difficult when an attacker compromises a legitimate domain and injects malware only periodically—"pulsing" it, as described earlier.

Another danger comes from users working at home or other remote locations beyond the protection of a corporate firewall or proxy. These users might click on a link in a targeted

attack, install malware or give away credentials, and end up compromising IT security. Enterprises need email link protection beyond the WAN–limited security delivered by conventional Web gateways; they need comprehensive email link protection that protects users even when they're accessing email on untrusted devices and networks.

To deliver this protection, Proofpoint has developed technology that ensures users never reach a URL before it's checked.

Conclusion

Inbound security attacks are more sophisticated and dangerous than ever before. They threaten not just an organization's daily operations but also its most prized intellectual capital.

By applying advanced Big Data techniques, Proofpoint provides its customers with comprehensive 24/7 protection against phishing, data leaks, and other cyber attacks from Advanced Persistent Threats. Through a commitment to leading–edge technology, and a new approach to email–borne threat detection in Proofpoint Targeted Attack Protection, Proofpoint gives mid–sized companies and large enterprises the best possible protection against the inbound IT threats of today and tomorrow.

For more information about the Proofpoint security–as–a–service platform, please visit **www.proofpoint.com or call +1 (408) 517–4710.**

proofpoint™

Proofpoint, Inc.
892 Ross Drive, Sunnyvale, CA 94089
Tel: +1 408 517 4710
www.proofpoint.com