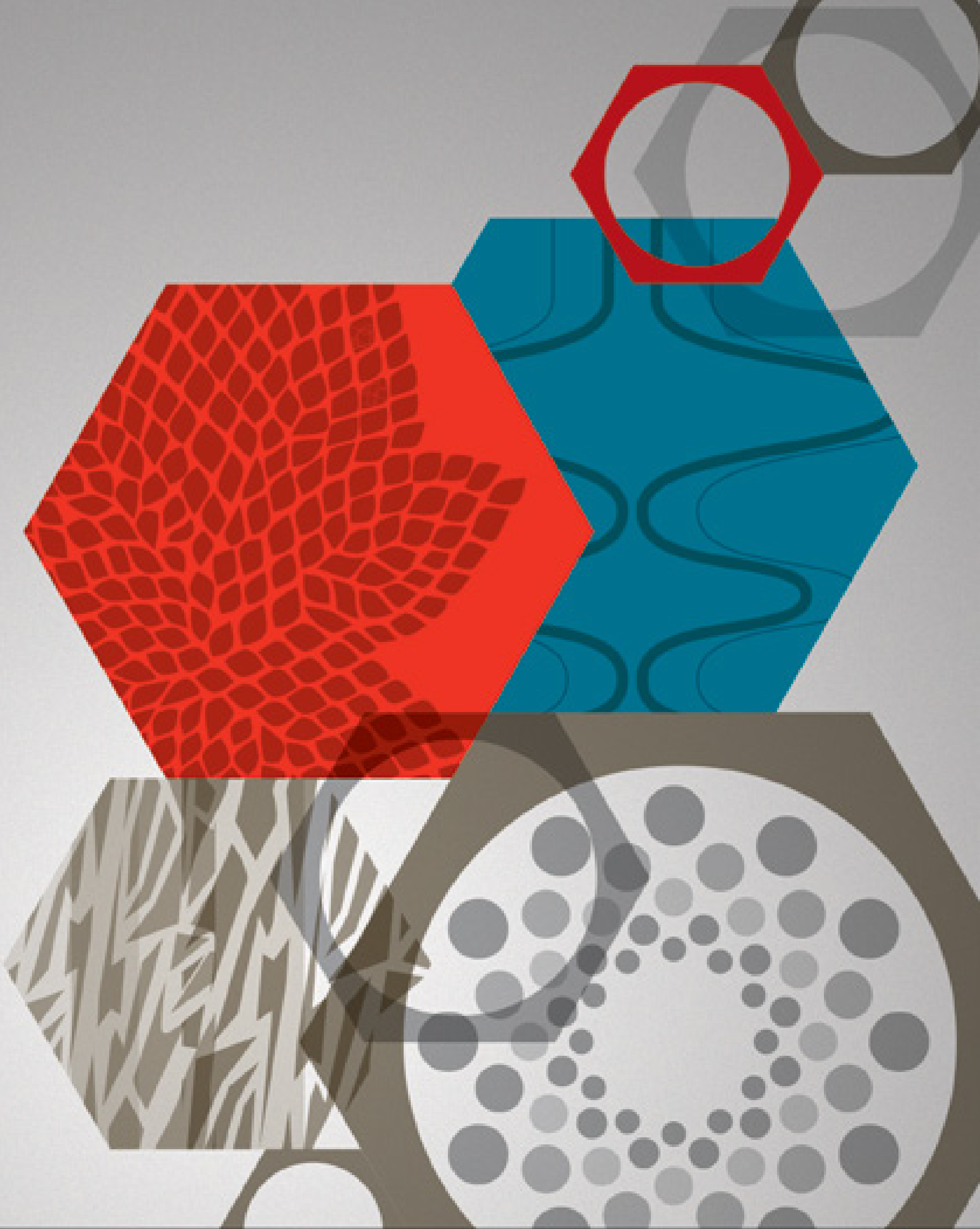


# ANATOMY OF A PHISH

## Exploring The New Security Threats From Social Attacks, And What To Do About Them

# CONTENTS

# OVERVIEW

“We have met the enemy and he is us.” That remark may just be the right capsule description of data security trends in the coming years. We are used to thinking of hackers as the other people, the ones outside the organization trying to break in. But in the case of phishing attacks, we—employees, managers and even a few CEOs—share a good deal of the blame for letting the hacker into our organizations. We’ve swallowed the bait—often a deceptive email or text message—and effectively opened the door for hackers and cyber thieves.

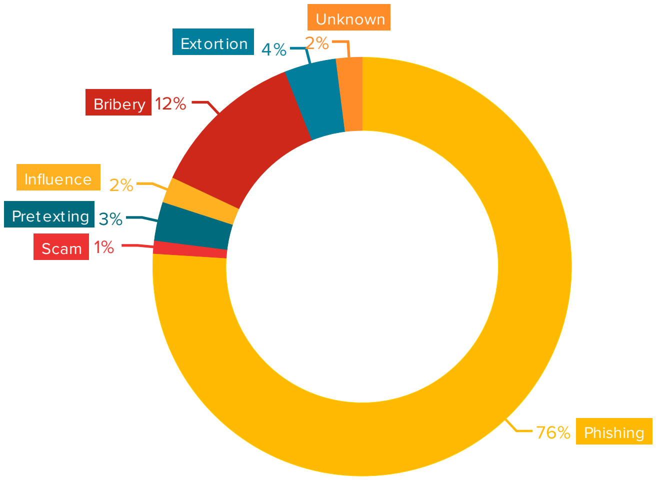
77% of all social attacks are based on phishing (Verizon 2013 DBIR)

# HOW BAD A PROBLEM IS PHISHING?

According to experts, phishing and other social attacks are on the rise. Verizon—yes, the wireless carrier—says that social attacks spiked in 2012. In their Data Breach Investigations Report (DBIR), Verizon's annual survey of hacking, attacks where victims have been tricked into revealing information, or socially engineered attacks, jumped by 52%. Phishing represents the largest component of social attacks with pretexting and old-fashioned bribing lagging far behind.

Reports and statistics are of course useful in seeing the big picture. Fortunately, the details behind some well-publicized social attacks at major US companies have surfaced. When looking at actual case files, it becomes very clear—we'll see this later—that social attacks take advantage of a typical employee's broad access to sensitive documents and files.

While many in IT may assume hackers crave elevated permissions—root access—the truth on the ground is that just ordinary access rights are all that's required to find troves of social security and credit card numbers, passwords, email addresses, and other personally identifiable information or PII.



# WHERE ARE HACKERS FINDING THIS SENSITIVE AND, OFTEN, EASILY MONETIZEABLE INFORMATION?

It's scattered throughout an organization's file system—in the data files employees are creating, copying, and editing as they go about their work day.

Wherever there is digital collaboration, there is human-generated data—90% of the data created in the next decade will be freeform, unstructured data. Most of that is human-generated data contained in documents and emails and spreadsheets and presentations that tend to be the most valuable and the most sensitive data you have. Often this data has improper permissions that are far broader than they need to be.

In this eBook, we'll be looking at a phishing incident involving a large US company. We've obscured some of the facts to protect identities, but all the significant details are based on actual events.

Let's just call this hypothetical company American Electronic Media or AEM. It's a new media organization—news-oriented web sites, game apps, social sites, web TV—with 10,000 employees, 5 large data centers, and 10 million subscribers for its services. AEM was recently the victim of a very successful phishing attack that resulted in the exposure of 30,000 customer records.

And the attack all started with an email. But not an ordinary email, one that was engineered to manipulate the victim to do something he or she wouldn't do knowingly. The email is the bait, and the unfortunate employee is called by hackers the phish.

# PHISHING LESSONS

The best way to understand a phishing attack is to step back, and first consider the common URL—you know, the standard dub-dub-dub dot kind. We expect brand names to show up in these URLs: for example, clicking on www.ibm.com or www.microsoft.com or www.fedex.com/us will take us to the official sites for IBM, Microsoft, and FedEx USA. But what about a URL such as www fedex3-dispatch.com—if that URL were in an email, one with official FedEx branding logos, and in which the message said that your package is ready, would you still click on it?

The aforementioned FedEx URL is not official, and would instead lead a customer to the hacker's trap: any personal information the phished victim enters into the online form on the fedex3-dispatch site would go right to the cyber thieves.

**PHISH BAIT FROM PHISHTANK:**
**THE DOMAIN NAME TELLS YOU ALL YOU NEED TO KNOW.**

"**real**" website: www.varonis.com
"**bait**" website: www.varonis.com.12.usa.newyork.com

The something dot com or something dot org part provides the clue that this is not FedEx's site. The something word immediately before the dot org or com is the domain name the company registers with ICANN—the organization that records all Internet domains. Microsoft, FedEx, and other major brands understand the value of having their brand in the domain and will register their own with ICANN. FYI: FedEx uses fedex.com for all its key customer-facing sites.

As a side note, brands often register sub-domains so that even legitimate URLs can look long and strange. For example, in this hypothetical URL, www.customer-us-portal12.fedex.com, customer-us-portal12 is a subdomain of fedex. The less confusing way to do this is for brands to not use sub-domains at all, but instead set up a path to the right of the domain, say,  www.fedex.com/custmer-su-portal12. In this URL, note that only the brand name is positioned to the left of the .com or .org part, making it far easier for trained users to determine the link's authenticity.

Hackers count on the fact that many web users are not technically knowledgeable enough to understand the underlying structure of URLs, so it's easy for them to make believable "forgeries". They merely need to cook up a faked address that looks enough like a real URL to fool an average user into clicking. Usually a variation of a brand name in the domain, say micros0ft.com or 1bm.com will work. But that may not even be necessary.

The Anti-Phishing Working Group or APWG is an industry-based coalition of e-commerce companies and ISPs that tracks and analyzes phishing attacks. According to them, the domain name itself usually does not matter to phishers! "Domain name of any meaning, or no meaning at all, in any TLD [top-level domain], will usually do," APWG noted in a report. "Instead, phishers often place brand names in subdomains or subdirectories. This puts the misleading string somewhere in the URL, where potential victims may see it and click..."

In other words, even very weird looking URLs will work, as long as the brand name (or some variation) is in the address. For example, here's a fake PayPal address, found in PhishTank's registry of phish URLs : http://service.paypal.creainfo.net/Login/websc/update.php. This one has received its share of clicks before it was reported—note that domain name, createinfo.net, is far removed from PayPal.

At its core, phishing attacks exploit the trust that consumers have in a brand. If they see it anywhere in the URL or email contents, they'll make assumptions about authenticity. The more information the hackers have about the phish, the more powerful the bait becomes—we say the email is social engineered to attract a specific individual. These "spear phishing" attacks move beyond the standard bait's minimal details, and may include the full name of the victim, home address, phone numbers, and possibly names of coworkers.

In the phish email that was used in the attack against AEM, the hackers exploited well known loopholes in SMTP, the email protocol. They "spoofed" the sender's address to make it appear as if it were coming from FedEx. A more technically savvy user would have expanded the email address in the From: field and checked the domain name's validity.

In the case of AEM, the employee, an analyst in the website production department, took the bait. As he normally signs for many FedEx shipments, he noticed his name was mentioned in the email and followed the email's directions for handling an undelivered package. Instead of asking him to click on a link, though, the email instructed him to open a FedEx invoice attached to the mail. By the way, FedEx would never send an email with an invoice.

The analyst hadn't, in this case, clicked on a fake URL, but just opened a document. Did he do anything wrong? We'll answer that a little later. Spoiler alert: he did indeed.

# TRAVELS WITH HUMAN-GENERATED DATA

Here's the short version of how data migrates and travels throughout an organization: data is collected from multiple sources (sales transactions, purchasing systems, banking, etc.), organized into centralized databases and then distributed to the appropriate employees and departments through application-specific systems.

The longer version has the official sounding name of Data Life Cycle Management. In this model, the latest data—usually transactional information—is scrubbed and normalized and then fed into data warehouses. At this point, the separate data streams are correlated and combined to make them more useful for decision making purposes. Finally, the data is carved into subject-oriented slices, and then distributed to, for example, separate applications for sales projection, cost analysis, or human resource planning.

But there's more than record-oriented, structured data in any organization.

There's terabytes of unstructured data that's created by employees—call it human-generated data. It's the data that's formed from thousands of daily work interactions and collaborations within any organization.

In a study conducted by the International Data Corporation's (IDC), analysts estimated that the amount of digital information created and replicated will grow 39% from 2012 through 2020, and more than 90% of the data created in the next decade will be unstructured data.

What's the life cycle of this human-generated data? There are some similarities to its structured cousins. Some of the data can originate from centralized databases—think of records from CRM and other ERP systems that are exported or downloaded by employees into readable files. These files can contain sensitive information and even highly granular data on individual suppliers, employees, or customers.

But then the collaborative aspect kicks in as knowledge workers contribute their ideas, analysis, and generate a cascade of new content—often containing references and excerpts of the human-readable versions of the structured data. Instead of being streamed into small sets of dedicated systems, human-generated data is ultimately dispersed throughout the organization—in presentations and documents scattered through file systems, intranets, and email.

At AEM, something very much like the above was happening. In fact, employees in the financial planning department had downloaded a spreadsheet of subscriber information for one of the company's online services. No one's quite sure who actually downloaded the data set, which had monthly revenue numbers for a specific US city but which also contained names, addresses, account and password information for thousands of subscribers. This spreadsheet was emailed to other departments, and at some point, a copy was placed in a file within a shared folder— in a directory that had very loose permissions.

# REAL WORLD PHISHING

While phishing is more associated with ordinary consumers—often to steal a single credit card or social security number—it has also been used with great success against companies. Why would hackers and cyber criminals target employees within large organizations?

One obvious reason is “that’s where the data is.” From a single attack, a hacker can gather large amounts of customer information (including PII) as well as confidential corporate information and other intellectual property. There’s another subtler reason why these corporate phishing expeditions are successful. The hackers can more finely engineer the attacks by making reasonable assumptions about the potential phish based on where he or she works.

Perhaps the employee is in an area of a company that is used to send and receive packages through FedEx, DHL, or another rapid delivery service. A crafty attacker sends multiple phish mails to a small group of employees requesting that they call the office because of a missed delivery. The email also asks the customer to download a copy of the invoice before making the call. While this attack can work against a random consumer, as in the case of AEM, it’s far more likely to yield better results in a corporate environment and in a specific department.

The goal of the phish mail attack is to get the employee to click, download, and then open the attachment—which often appears as a zip file.

When the zip file is opened, the phish has been hooked. The employee has actually launched malicious software or malware that is stealthily monitoring and collecting internal data. There are a few variations on the malware—for example, single purpose key loggers and RAM or screen scrapers are designed to monitor employs computer interactions and then send them back to the attacker. But there is an especially malicious types of malware, known as Advanced Persistent Threats or APTs, which is designed to remain undetected while scooping data under the direct control of the hackers.

How do these APTs work in practice? We’ll sketch out one such APT, known as WEBC2—the name will make sense in a second:

1. After the malware attachment is opened, it sends HTTP requests—just like a browser— to special web servers set up by the hackers. To networking and other periphery monitoring software, the packet exchange between this APT malware and the hacker’s web server appear as ordinary user browsing traffic.

2. At the web server site, the hackers have cleverly inserted secret commands into the HTML code— placed within special HTML tags—which have no meaning as a graphic object for the browser. Background: web pages are generally sprinkled with comments, mostly used by developers to document their creations. If you examine a page’s raw HTML, you’ll see something like *<!—some explanatory text--!>*, which are ignored by Firefox, IE, Chrome and other browsers.

3. These hidden commands are known as WEBC2 backdoor or Web “Command and Control”. The commands are read by the malware software, instructing it to perform a few basic actions. Depending on the WEBC2 variant, the commands can include launching an interactive command shell, uploading and download files, listing processes and DLLs, and pinging hosts on the local network. Of all these, the ability to launch a shell remotely is the most powerful. Effectively, the hacker has admin level access to Windows or Linux commands: they can navigate through a directory structure, search for text patterns, and open and edit files.

4. Any information that the hackers “exfiltrate” or remove from the victim’s system is encrypted and compressed before being POSTed as a Web request. As far as an organization’s networking monitoring software is concerned, the exfiltrated data looks like a browser-based form that’s been filled out and submitted to a legitimate web site.

5. Finally, the hackers, acting as the control agents, are remotely refining their WEBC2 commands, updating them as they learn more about the targeted system.

Back at AEM, the attachment that the analyst in the website production department opened up contained an APT, very similar to the WEBC2-style malware. The remote hackers then began a month long campaign to ‘stake out’ their new system. They were not in any hurry, knowing they could be very sure that their activities would not be noticed. Even if the IT department enabled system-level auditing, the hackers’ commands to navigate the file directory hierarchies, list files, and view files contents would likely appear as an average profile of an AEM website analyst and would not raise any red flags.

# SECURITY WALLS VS. SECURITY CAMERAS

Over the last few years, a new consensus has been forming about how to handle the disruptive security threat posed by phishing attacks. Traditional networking and virus-detection software focuses on, so to speak, locking the door. But with phishing attacks, internal employees effectively let the hackers in. And once in, APT malware is designed to elude detection. RSA, the leading data security and encryption company, noted in a recent report that the "definition of successful defense has to change from keeping attackers out' to 'sometimes attackers are going to get in'. Their advice is to detect them as early as possible and minimize the damage."

As a CTO, CSO, or IT security executive your operational assumption should be that your organization has already been compromised.

In fact, you'll be more secure if you assume someone will always get around your security walls. So you need to take a cue from the world of physical security by installing the digital equivalent of a security camera.

Just like a security guard scanning video feeds, your IT security resources should be focused on spotting unusual activity or, according to the RSA report, "to detect them [APTs and malware] as early as possible and minimize the damage." System admins should forget about trying to do this monitoring and surveillance work using standard methods—manually looking at system or app logs and then guesstimating. Humans just can't read and process all the logs quickly enough to make a difference.

The report recommends a real-time analytics engine that maps out trends on user and system activity based on lots of data points, and then alerts the appropriate IT security staff after finding anomalies.

Meanwhile AEM hackers have been controlling the WEBC2 malware for over two months. Since the malware inherits the credentials of the AEM analyst, the remote hackers are "limited" to those parts of the file system for which the analysts have read or view permissions. But that turns out to be more than enough. The hackers eventually discover the shared folded with the spread sheet containing personal data of thousands of AEM subscribers and exfiltrate the file. They've hit the jackpot and quickly close down their operation—delete files, and take down the C2 server.

How long did it take for AEM to discover the exposure of customer records? Overall, there isn't good news on this front. The Verizon DBIR talks about the average delay between data exposure and detection being measured in months. For AEM, the breach was discovered 6 weeks later after some of its subscribers complained about their accounts being hacked and then paying for services they didn't order. Only after extensive analysis by AEM's security team were they able to make the connection between the phish mail from FedEx and the malware attachment.

# FIVE ANTI-PHISHING STEPS

When faced with new security threats, IT security have traditionally gone on the offensive:  purchasing more monitoring components for the periphery of the corporate network. But APTs, such as WEBC2, suggest a far different approach. Organizations now need to develop better defensive or 'Plan B' capabilities. The goal is not to prevent entry but to limit the damages of hackers who are already in your system.

Fortunately, there are some very straightforward actions that can drastically reduce the threat and liabilities involved with social attacks. Here are five to get you started:

## EMPLOYEE TRAINING

Employee should be trained to spot the most obvious types of phishing attacks. They should not click on a link within an email purporting to be from a bank, airline, delivery service, or credit card company. Also, employees should be trained to identify the domain names embedded within email addresses or website HTML. And they should never, ever extract a zip file from an outside address. Some companies have even run simulated phishing attacks to measure employee susceptibility—it's an idea you may want to explore.

## AUTHENTICATION

While you may not be able to prevent hackers from getting in, you can make it far more difficult for them to get easy remote access even if they have an employee's password. Two-factor authentication forces remote users to provide not only something they know, a password, but something they have or something they are (biometrics).

## FINDING ESCAPED PII

Another important preventive method is to ensure that sensitive customer information, especially PII, is only available to those who are authorized to access it, and restricted to well-defined locations. In many breaches, the exposed PII was accessible to anyone having user-level permissions—the data files were either in folders with "everyone" access or having broad group-level permissions. So a good defensive plan is to regularly scan the file system looking for PII patterns that have escaped into unauthorized folders. PII definitions vary by industry—medical and financial in particulre have their own regulations—so you'll need to review legal definitions or industry-standard compliance rules in developing search patterns.

## DATA OWNERSHIP AND FILE PERMISSIONS

Who actually owns the data? That's the question many experts say should be the cornerstone of a data loss mitigation strategy. This requires identifying your most important digital assets and determining who in your organization should be responsible. It's the true owners—typically content experts, line of business managers, key executives—who have the best knowledge about the data and, most importantly, who should have access it. It's also critical that the data owners are in any workflow arrangement involving approval decisions about new users requesting access, and should regularly review existing users to remove those who should no longer have access.

## MONITORING USER ACTIVITY WITH AN ANALYTICS ENGINE

With a phishing attack, hackers effectively take on the identity and to some extent the computer behaviors of employees. The APT malware controlled by the hackers performs typical employee computer commands. In other word, it's not easy to spot the hackers with standard techniques—reviewing audit logs and system activity. Instead, you'll need special software that learns the baseline file activity of employees and is constantly monitoring current activity—who's reviewing files in a specific directory, who's copying files, etc. It's the only way to differentiate in near real-time the hackers from legitimate users. At some point their actions will differ enough from average patterns, and the analytics engine will be there to notify IT security.

# ABOUT VARONIS

Varonis is the leader in unstructured and semi-structured data governance software. Based on patented technology and a highly accurate analytics engine, Varonis solutions give organizations total visibility and control over their data, ensuring that only the right users have access to the right data at all times from all devices, all use is monitored, and abuse is flagged.

Varonis makes digital collaboration secure, effortless and efficient so that people can create and share content easily with whom they must, and organizations can be confident their content is protected and managed efficiently.

## Free 30-day assessment:

### WITHIN HOURS OF INSTALLATION

You can instantly conduct a permissions audit: File and folder access permissions and how those map to specific users and groups. You can even generate reports.

### WITHIN A DAY OF INSTALLATION

Varonis DatAdvantage will begin to show you which users are accessing the data, and how.

### WITHIN 3 WEEKS OF INSTALLATION

Varonis DatAdvantage will actually make highly reliable recommendations about how to limit access to files and folders to just those users who need it for their jobs.

**WORLDWIDE HEADQUARTERS**
1250 Broadway, 31st Floor, New York, NY 10001 **T** 877-292-8767 **E** sales@varonis.com
**UNITED KINGDOM AND IRELAND**
Varonis UK Ltd. Warnford Court 29 Throgmorton Street London, UK EC2N 2AT **T** 020 3402 6044 **E** sales-uk@varonis.com
**WESTERN EUROPE**
Varonis France SAS 4, rue Villaret de Joyeuse 75017 Paris France **T** +33 (0)1.82.88.90.96 **E** sales-france@varonis.com
**GERMANY, AUSTRIA AND SWITZERLAND**
Varonis Deutschland GmbH, Welserstrasse 88, D – 90489 Nürnberg **T** +49 0911 893711 11 **E** sales-germany@varonis.com