State of the Web 2015:
Vulnerability Report

March 2015

Motivation

In February 2015, security researchers reported that Forbes.com had been hacked. The duration of the exposure, known as a "watering hole attack," is unknown – at least days, maybe weeks or months. Visitors to the site, which included people at major defense and financial organizations, were infected by merely going to the site, without clicking on any links. Researchers noted that "watering hole attacks are insidious because it wouldn't occur to anyone that these sites could be infected."

Insidious indeed. The attackers exploited a vulnerability in WordPress, the popular Web publishing software used by Forbes and millions of other organizations, to insert their malicious code for delivery via the "trusted" Forbes.com site. The attack was eventually detected and the malicious code was removed from the Forbes.com site. But the full extent of damage from this one incident will likely never be revealed and probably continues today.

The Forbes incident is merely the latest reminder that even seemingly innocuous activities on the Web – like browsing to the home page of a well-known and trusted site like Forbes.com - are fraught with risk. This report is intended to help quantify the scope of that risk as organizations struggle to balance their cyber security policies and protections against the needs of their employees for access to the Web and its resources.

Background

Despite enterprise spending of more than $70 billion on cyber security tools in 2014, malware continues to evade even the latest security technologies. In fact, successful attacks are on the rise in number and severity. In 2014 businesses lost nearly $400 billion as a result of cyber crime. In many cases a company's own employees put the business at risk, often unintentionally, by browsing to a trusted website or clicking on a link in an email that brought them to a compromised site. With today's increasingly sophisticated malware, simply navigating to a compromised website or opening a document can unleash a whole slew of malware onto a user's computer. Once an endpoint has been compromised an attack can quickly spread to other systems both within outside the user's organization.

"A WordPress vulnerability allowed Forbes.com to be used to deliver malware"

The vast majority of malware infections are delivered via Web browsing and email. But eliminating or even limiting access to the Web is simply not an option for the growing number of organizations that rely on knowledge workers who need access to the Web to do their job.

Organizations expect their security teams to keep their data and systems safe while also providing employees with access to the Web. A major factor compounding the security practitioner's challenge is the fact that other companies' websites, which they don't control, constitute a major source of threats. The scope of the problem is daunting: There are over one billion websites on the Internet, with more than 100,000 new sites coming online daily. One study reported that over 70 percent of Web domains exist for just a single day. And as the Forbes.com incident shows, the notion of a "trusted" site is often illusory, because a vulnerable site cannot ever be trusted. So how broad is the risk from vulnerable sites?

Methodology

With these questions in mind, in January 2015 Menlo Security scanned the Alexa top one million sites to see which sites were vulnerable and/or compromised. In total, the team scanned over 1.75 million URLs representing over 750,000 unique domains. Each URL was analyzed as follows:

- The URL was checked against third party classification systems to determine the site category (e.g. eCommerce, gaming, news, etc.) and to see if it appeared on lists of known malicious sites;
- The IP address to which each URL resolved was checked against an IP reputation database to determine if it was associated with a spam network, botnet or other bad actor;
- A Web request was issued to each URL, and the response received was finger-printed to identify the software running the website, including the Web server software (e.g. Apache, IIS), the content management system (e.g. WordPress, Drupal) and the application framework (e.g. PHP), etc. The fingerprint was then checked against the CVE List (http://cve.mitre.org) to identify sites that were running vulnerable, unpatched software.

Key Findings

The research results show that a significant percentage of the Web is either currently compromised or at risk:

- One in three of the top one million Alexa domains are "risky" – meaning that they're either already compromised or running vulnerable software and therefore at risk of compromise by groups or individuals planning the next attack.
- More than one in twenty sites, or six percent, were identified by 3rd-party domain classification services as serving malware, spam or are part of a botnet.
- Over one-fifth (21%) of sites were running software with known vulnerabilities.
- Of the 2.5 percent of sites that were "uncategorized," a significant proportion (16%) were running vulnerable services.

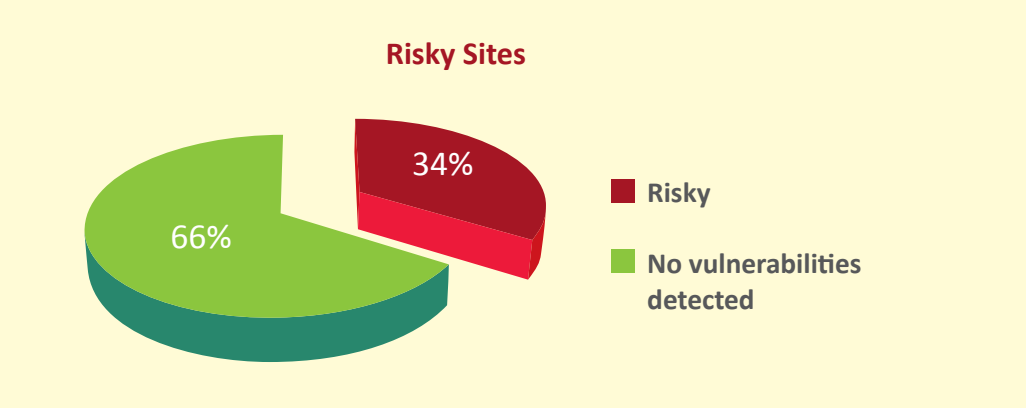

Figure 1: One-third of the Alexa top one million sites are compromised or vulnerable to compromise

Alexa Top 1Million Domains KEY FINDINGS:

1 in 3 | *are risky*

1 in 5 | *run vulnerable software*

1 in 20 | *serve malware, spam, or botnets*

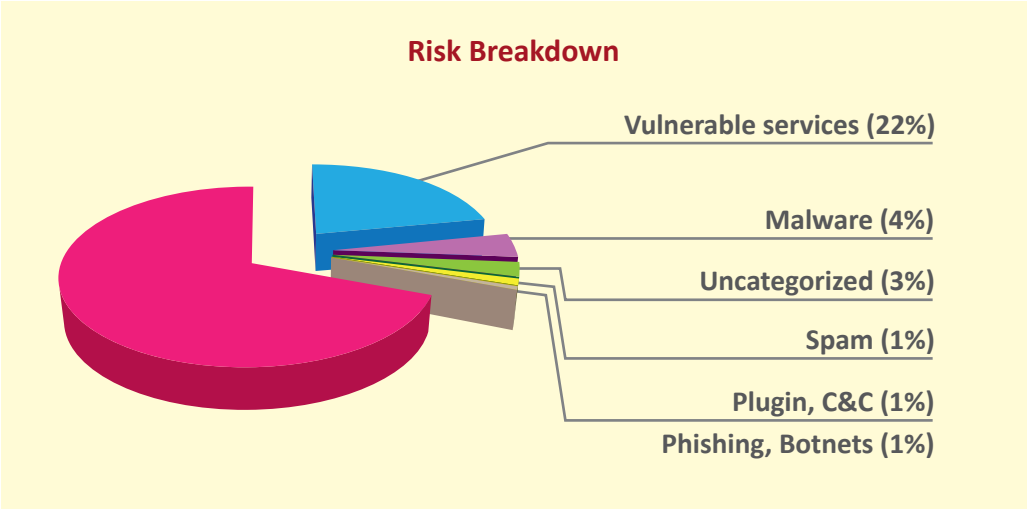

Figure 2: More than 20 percent of the top one million sites are running software with known vulnerabilities

Website infrastructure can be compromised via multiple vectors. Over one in ten sites are running a vulnerable version of the PHP application framework. Another eight percent are running vulnerable Web server software (Apache-4% and IIS-4%). Vulnerable content management systems are present on two percent of sites, split roughly equally between Drupal and WordPress. It's worth noting that no special or invasive means were needed to determine if a site was running a vulnerable service: Information regarding a site's underlying software infrastructure is routinely returned to any browser that makes a Web request. Attackers need no more than a standard browser to find vulnerable sites to exploit.

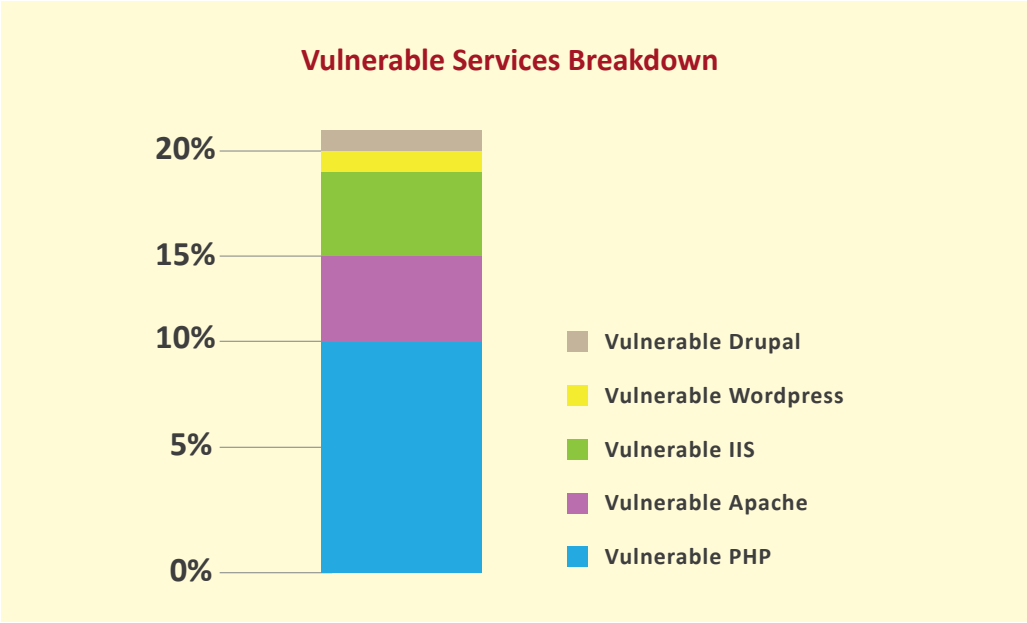

Figure 3: Many sites are running vulnerable versions of core Web infrastructure software and finding them is easy.

"Attackers need no more than a standard browser to find vulnerable sites to exploit"

Who can you trust?

We expected that sites with questionable content would have vulnerable software, and thus weren't surprised to find that sites hosting content labeled as Hate and Intolerance, Violence and Child Abuse exhibited vulnerability rates of 25-35 percent. More surprising was the vulnerability rates of sites that are typically trusted, such as Transportation (20%), Health and Medicine (20%), Computers and Technology (18%) and Business (18%). Given their prevalence in the Alexa one million, categories that are typically allowed by Web filtering policies, like Computers & Technology, Shopping and Personal sites represent the three top contributors to vulnerable sites by number. This further reinforces the notion that the type of site is not necessarily a reliable indication of the likelihood of it being compromised.

Uncategorized sites present an additional challenge. They represent 2.5 percent of total sites and show a 16 percent vulnerability rate. The obvious response is to block employee access to uncategorized sites. However, doing so can have significant implications when legitimate sites are blocked, thereby preventing employees from doing their jobs and generating requests for blocked sites to be re-categorized and unblocked. Of course, allowing uncontrolled access to uncategorized sites poses an even greater risk.

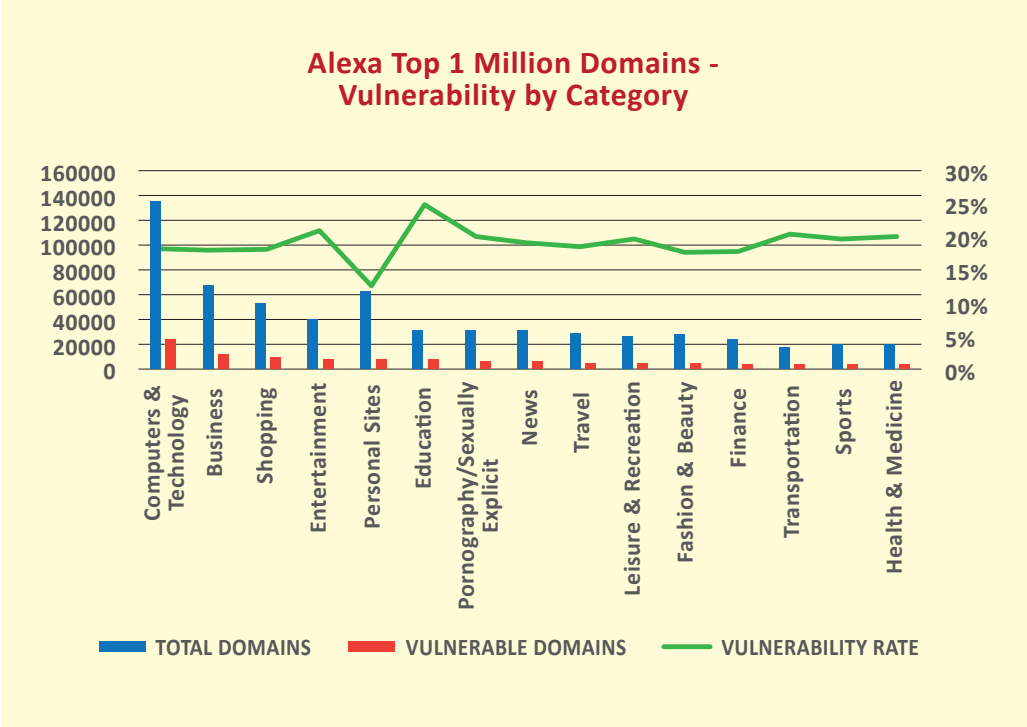

Figure 4: Trusting sites based on category is risky

"Categories typically allowed by web filtering policies represent the top three sources of vulnerable sites"

Implications and Conclusions

Like it or not, most organizations need to provide their employees with access to the Web, yet doing so opens the organizations to risk. In the face of this necessity, organizations spend billions in security defenses, most of which rely on the ability to distinguish safe sites from compromised sites and good content from bad content. In view of the number and scope of successful attacks being reported (and the many that aren't reported) this strategy is clearly not working. An analysis of the Alexa top one million websites illustrates the challenges: a sizeable portion of the Web – some six percent of domains – is already compromised, and more than one-fifth of Web domains are powered by vulnerable software that could be compromised at any time.

In light of the facts, what should organizations do to protect their employees? Restricting access to pornography, gaming and other such categories is a reasonable practice for reducing wasted time and boosting productivity; however, allowing access to seemingly safe site types such as business, education and shopping can actually expose an organization to even greater risks. Uncategorized sites add further complication as they're very likely to be vulnerable, but blocking all of them can lead to lost productivity for both employees and the IT staff that supports them.

In view of the data, the Forbes.com incident can be seen in a new light. It was not a rare and unlikely event. On the contrary, the millions of sites on the Web running vulnerable software provide a rich and fertile auger for supporting new attacks. Vulnerable sites can be compromised at any time, and thus any list of "safe" vs. "unsafe" sites becomes inaccurate the moment it's published. And when vulnerable sites are compromised and used to launch zero-day malware, existing security technologies consistently fail to detect and stop infections until after the damage has been done.

The next major attack is likely already in process – it's just a matter of time until we discover what's been lost. Organizations are trying to minimize the damage by investing in new tools that do a better job of detecting infected systems and limiting the impacts of security breaches. That's certainly a good idea. But it's also a bit like investing in bigger pumps to empty a basement being flooded by a broken water pipe. At some point, you need to address the source of the problem. The real answer to the challenge of preventing Web-based attacks will come from new tools that can completely stop all attacks before they reach their targets.

"Existing security technologies consistently fail to detect and stop infections "