Understanding Web 2.0

Web 2.0 is more about how use of the Internet is changing than about a new version of web technologies. The Internet is becoming more of a platform for existing technologies, many of them collaborative, and a perpetual beta site for new technologies. So-called “mashups” of existing technologies, combined with growing numbers of knowledgeable users, and the proliferation of how-to data, is opening doors to threats that didn’t previously exist. Web-based communities and hosted services such as social networking sites, wikis, and blogs, aim to facilitate creativity, collaboration, and sharing among users; but the very openness of these sites makes them vulnerable to new, multi-layered malware attacks.

"The web 2.0 tag bubble map” made by ace designer Markus Angermeier, shown in Figure 1, gives an idea of the technologies, ideas, and uses that comprise web 2.0.
The Internet as a Platform

Web 2.0 thrives on network effects: databases that get richer the more people interact with them, applications that are smarter the more people use them, marketing that is driven by user stories and experiences, and applications that interact with each other to form a broader computing platform.

What’s causing this change? Consider the following raw demographic and technological drivers:

- One billion people around the globe now have access to the Internet
- Mobile devices outnumber desktop computers by a factor of two
- Nearly 50 percent of all U.S. Internet access is via always-on broadband connections

Combine these drivers with the fundamental laws of social networks and lessons from the web’s first decade, and:

- In the first quarter of 2006, MySpace.com signed up 280,000 new users each day and had the second most Internet traffic
- By the second quarter of 2006, 50 million blogs were created – new ones were added at a rate of two per second
- In 2005, eBay conducted 8 billion API-based web services transactions

These trends manifest themselves under a variety of guises, names, and technologies, notably: client-side processing (AJAX/RIAs), XML syndication (Blogs and Wikis), mashups and shared content, social networks and user participation.

Client-Side Processing (AJAX and RIAs)

Asynchronous JavaScript and XML (AJAX) is arguably the most popular member of a class of technologies that enable rich Internet applications (RIAs) by performing functions and processing data through seamless background integration. RIAs create the visual effect of an interactive web page, though in actuality the traditional Hypertext Transfer Protocol (HTTP) GETs and POSTs are occurring behind the scenes. The goal of RIAs is to give the user an experience equal to or better than that of a desktop application. One way RIAs accomplish this is to make a single-page interface more intuitive by providing contextual controls wherever the opportunity exists. So hovering a mouse over a word may trigger a pop-up definition, or hovering over a stock ticker symbol may bring up its current trading data. AJAX has become popular now that most browsers support the four key underlying technologies – JavaScript scripting language, Cascading Style Sheets (CSS) to dynamically apply styles, the Document Object Model (DOM) that defines object properties of the web page, and the XMLHttpRequest (XHR) object that makes dynamic HTTP calls “behind the scenes” of the page. Because of the full browser support and addressability of objects within it, developers have embraced AJAX as a way to spice up the browser window and improve the user experience.

XML Syndication: Blogs, Wikis, and RSS

Blogs, wikis, and many web sites support XML file formats that provide syndication of content. RSS (Real Simple Syndication) and Atom are two of the most popular formats. RSS and Atom files provide the necessary information for software running on the desktop, called “newsreaders” or “aggregators,” to automatically recognize whether a particular web site or grouped set of web pages has been updated or otherwise changed in some way. Typically, a user subscribes to web content by inserting the feed name (a Uniform Resource Locator [URL]) in the newsreader. The newsreader monitors the feeds and identifies new content either through a scheduled or manually triggered process.
Mashups and Shared Content

Mashups combine data from multiple sites into a single user experience. While there are many ways to capture and process information from other sites, the preferred web 2.0 method is through a set of application platform interfaces (APIs). For example, Amazon and Google both have APIs for other sites to customize in their mashups [see iGoogle].

Sharing information is not new to the web; some might even say that is the reason the web exists. Many web sites use available Hypertext Markup Language (HTML) capabilities such as image source references and iFrames to retrieve content from third-party web sites and incorporate it into a page. Ad servers that provide online content for publications are a common example of this use. APIs simply extend that model.

However, there is a significant architectural difference between traditional links such as those in online magazines that connect to ad servers and the way public APIs are being used by mashups: the actual retrieval of the information shifts from the client side to the server side. When a browser issues a request for a web page, the browser interprets the content as it is downloaded. When it sees an external link such as those associated with advertisements, the browser issues a request directly to that site to retrieve the data.

In contrast, a mashup issues its request via a server-side “proxy” or callback. The user still downloads the web page, but in order to deal with “same origin” requirements in JavaScript and other programming languages, the data retrieval must return to the original web site. A proxied process in the background then retrieves the information and presents it back to the user. This means that several sources are trafficking data without explicit user direction.

Social Networks and User Participation

User participation in content creation is the name of the game on the Internet today. Web sites that facilitate and encourage users to create their own content act as distributors to other like-minded individuals. MySpace, Facebook, Flickr, and Wikipedia are examples of the user participation phenomenon. On such sites, users can:

- Update web pages with commentary via standard HTML forms (blogs, wikis, social networks, and forums)
- Upload files for viewing and download by others (résumés, photos, audio podcasts, and videos)

This user participation is reflected in the design of social networking sites: They operate as structured environments that allow users to manage their own web pages – often with a number of configuration options for look and feel as well as various types of transaction processing.

Why Does Web 2.0 Matter?

Web 2.0 technologies have invaded the workplace leading to a dramatic increase in malware and data leak vulnerabilities.

Web 2.0 sites such as blogs, wikis, and RSS feeds have complex content that require code to be executed on the user’s browser, which in turn allows hackers to embed malicious code that is then automatically executed.

Web 2.0 client-side technology (applications run on a user’s local computer connecting to a server) has developed far more quickly than security technology, which traditionally protects the server operating system rather than the client side (depending on the user’s security implementations for protection).
Web 2.0 sites often use data from different sources, such as a retail store site using Google maps to display locations, this makes it more difficult for security systems to validate the integrity of the code. Hackers are also embedding spam and malicious code into other types of content, such as instant messaging, shared video content, and business documents like PDF and Excel files.

Everyone who depends on the Internet must learn and face the challenges of web 2.0 because it is changing the way everyone, enterprises and individuals, use the Internet:

**Internet 2.0**
- 42% of office workers 18-29 discuss work-related issues on social networking sites (blogs)
- 50% of IT managers indicate that 30% of bandwidth is social networking
- SaaS (software-as-a-service) usage is steadily increasing

**Intranet 2.0**
- Forrester predicts that web 2.0 intranets will be a $4.6 billion industry within 5 years

**Extranet 2.0**
- Nearly half of all web developers are already using AJAX
- In 2007 more than 30% of large companies had a web 2.0 business initiative
- 66% of companies indicate web 2.0 is essential to maintaining their company’s market position

**Web 2.0 in the Workplace – Enterprise 2.0**

Internet use in the workplace has become ubiquitous. Employees browse dozens of websites a day for research purposes, competitive analyses, and pertinent information. Many useful tools and Internet sites have enabled people to become more efficient at accomplishing multiple tasks in a shorter amount of time. The average worker spends nearly 2.5 hours a day online (eMarketer). Outsourced applications and services such as Salesforce.com and Outlook Web Access are common in the enterprise. Enterprise WAN application delivery networks (ADNs) deliver sensitive data, sometimes in real-time. See Figure 2.

In September 2007, Forrester Consulting conducted a survey of 153 businesses with 1,000 or more employees. They found that organizations, whether officially sanctioned or not, use web 2.0 applications extensively. The future workplace will include web 2.0-inspired applications such as RSS, blogs, RIAs, tagging, virtual worlds, and wikis, according to a recent report by Forrester detailing the “The Seven Tenets of the Information Workplace.”
AJAX and RIAs

Because RIAs improve the way people find and manipulate content, complete transactions, and consume multimedia content, these technologies are ideal for improving the user experience for information workers. Moving forward, RIA technologies like Adobe Flash and Flex, Adobe Integrated Runtime (AIR), AJAX, the Curl RIA Platform, Laszlo Systems OpenLaszlo and Webtop, Microsoft Silverlight, Nexaweb’s Enterprise Web 2.0 Suite, Oracle WebCenter, and Sun JavaFX will be used to augment or even replace traditional enterprise portals and Microsoft Office as Information Workplace front-ends.

Blogs, Wikis, and Feeds

Blogs are valuable knowledge management and communication tools in companies, quickly communicating information such as project updates, research, and product and industry news both inside and outside the business. But blogs aren’t entering through the CIO’s office; they often first appear in companies as the convenient records of engineering or design projects.

Wikis are revolutionizing collaboration within the enterprise much as email has revolutionized communications. Wikis may be used for knowledge management, document management, project management, documentation, scheduling, meetings, directories, and more. Unlike most previous collaboration tools, wikis are simple enough to use without special training or a large degree of technical knowledge. An estimated 33% of enterprises are already using wikis and another 32% plan to do so within 2 years (Economist Intelligence Unit survey, January 2007).

RSS feeds are available for a wide variety of web sites; many can be useful to professionals. In the workplace, professionals subscribe to a newsreader or aggregator client software or (if the feed is hosted) a web application that aggregates syndicated web content such as news headlines, blogs, podcasts, and vlogs in a single location for easy viewing. Many companies that discovered utility in blogs and wikis are realizing that RSS is necessary to push that content to users.
Social Networks
Social networking technology allows users to keep in touch and share data with colleagues in an increasingly rich environment. The ‘millennials’ – generally considered those between 16 and 24 years of age – use sites like MySpace, Xanga, and FaceBook the way that many older people use the telephone, email, and coffee shops. Social networking sites are a way to communicate and the millennials are totally comfortable in the online environment; they “chat” and “blog” and post pictures of parties and parents and everything else on the Internet. The line between public information and private information is rapidly blurring.

Hackers Have Changed
Perhaps the single most relevant change is in the hackers themselves. Criminal organizations have realized how powerful the Internet is for committing online fraud, and they’ve invested huge resources. They’ve learned to take advantage of the Internet’s weaknesses and to exploit these for their own profit. This ‘new’ fraud includes identity theft, credit card fraud, theft in online banking, illegal pharmacies, scams, cyber squatting, fraud in online auctions, and malicious code.

Criminal hackers have cost United States businesses an estimated $67.2 billion a year, according to the US Federal Bureau of Investigation (FBI).

More Criminal Intent for Profit over Fame
A report released by the Georgia Tech Information Security Centre (GTISC), one of the leading independent security research institutes, warns that 2008 is likely to see a wide range of new threats to information security. “The increasing involvement of organized crime syndicates in online theft is leading to more sophisticated hackers who are motivated by financial gain, rather than personal reasons,” according to the report, leading to more sophisticated attacks that often combine different techniques and look to exploit developing technologies that are not as well protected as existing systems.

These issues include outside hackers; employees attempting to access unauthorized material; and denial of service attacks (DoS), protocol-based attacks that attempt to overwhelm company systems. Hackers may try to guess passwords, intercept unencrypted communications, bypass application security with attacks that exploit a weakness in the software, find ‘backdoor’ weaknesses including unprotected accounts or admin accounts that have the default passwords, and more.

More Money, Time, and Expertise
Hacking has achieved enterprise status. Malware kits are sold – with management consoles and regular updates – over the web. Professional programmers are paid to discover and exploit vulnerabilities. A single good vulnerability in popular software such as Internet Explorer can be sold for as much as $10,000.

There are numerous websites devoted to distributing, usually for a fee, tips on spamming and propagating malware. Newbie hackers have a name of their own: script kiddies.

More Short-Burst Attacks
New websites can be thrown up for the purposes of botnet recruitment (discussed below) or phishing attacks [criminal masquerades of legitimate e-commerce] and torn down before any database can categorize them. Roving DNS host IDs and temporary websites are used to keep one step ahead of URL categorizing engines and reputation tracking host databases.
Botnets – A Hidden Economy
Bots are little programs that are installed silently without user intervention; often through drive-by downloads operating secretly on an otherwise "good" website. A botnet is a network of computers on which bots have been installed, usually managed remotely from a Command & Control (C&C) server; however, more and more through peer to peer management to be more resilient. The main purpose of botnets is to use hijacked computers for fraudulent online activity; they are managed by a criminal, a group of criminals, or an organized crime syndicate and are often for rent.

Once a set of computers has been compromised, they can be involved in many kinds of online criminal activity, including identity theft, unsolicited commercial emails, scams, and massive attacks. It is estimated that more than 6 million infected computers worldwide are connected to some type of botnet. Most owners of infected computers do not know that their machines have been compromised. Well known botnets are STORM, MegaD, and MayDay. The eventual weakness of a botnet is 'exposure' allowing security researchers more samples to analyze and then develop defenses.

Bots gained their current status as a result of several factors. Perhaps the most important is that bots leverage the work of others. Several bot families are considered open source projects, developed collaboratively and refined by many.

But even more importantly, bot developers piggyback on the work done by well-intentioned security researchers. Most cybercriminals do not have the skills to discover and exploit software vulnerabilities. But when such vulnerabilities are made public in an effort to raise awareness, bot authors incorporate the work into new versions of their threats.

According to Commtouch, more than 85% of spam messages and nearly 100% of malware messages are sent from zombie machines (a system infected with a bot). As of early 2008, Google Message Security had tracked a 62% increase in the daily number of unique IP addresses that are blocked by its network compared to early 2007. This is a clear indication of the growth of botnets.

According to Postini statistics:

- More than one million internet protocol (IP) addresses are now coordinating spam and malware attacks each day.
- More than 50,000 infected computers are attacking at any particular point in time.
- Spammers are constantly changing IP addresses to evade detection.

Organizations Are Not Prepared For Web 2.0 Threats
Miscreants on the Internet manipulate and defraud users and organizations alike. Web 2.0 technologies introduce risks along with benefits; the core notion of a flexible and open architecture invites exploitation and compromise. This is nothing new. But the web’s components are becoming more distributed and interdependent than ever before, and they are often at or beyond the outer bounds of control for the individual enterprise or user.

Want Benefits of Web 2.0 But Worried About the Impact
Numerous web 2.0 applications are useful for business purposes. At the same time, however, the extensive use of web 2.0 applications leads to increased non-business web usage and bandwidth consumption, which in turn leads to decreased employee productivity. Many businesses believe that taking away access to social networking and rich media sites will visibly increase employee productivity. In a survey of network managers and security managers, more than 70% wanted social networking sites banned.
According to Forrester research, 47% of organizations reported that malware and viruses caused significant problems, while 42% reported problems due to Trojan software. Twenty-one percent experienced critical business disruption from viruses and 16% with Trojans.

Organizations also listed data leakage as a major source of problem. Thirty-three percent reported significant problems as a result of data leaks, and 18% experienced critical business disruptions.

Web Gateways Lack Adequate Protection
Forrester research also reveals a visible discrepancy between how prepared businesses perceive themselves to be and how prepared they actually are to deal with these threats. Nearly 97% consider themselves prepared for web-borne threats, with 68% conceding room for improvement. However, when asked how often they experience malware attacks, 79% reported more than infrequent occurrences of malware, with viruses and spyware leading the pack. Perhaps more astoundingly, 46% of the organizations surveyed reported that they spent more than $25,000 in the last fiscal year exclusively on malware cleanup.

Threat Focus on Email/Lack of Productivity
Traditionally, enterprises have focused on filtering email for viruses or spam and/or limiting employee use of the web in an effort to combat loss of productivity. However, with the rise of web 2.0 technologies neither of these strategies is adequate today. Malware and spam arrive in the form of browser "drive-bys" even when the website is a legitimate one. Data leakage, often inadvertently, is becoming a greater issue that productivity loss.

How Security Challenges Change with Web 2.0
The underlying functionality of web 2.0 includes, at minimum, the technologies associated with client-side processing, syndication, mashups, and user participation. All of these technologies have a role to play in web 2.0 and introduce a set of risks that must be evaluated.

To accomplish their objectives, attackers use a myriad of techniques in the web 2.0 environment. They upload malware, insert malicious code, traverse domain security restrictions, abuse rendering engines, and propagate worms (SPI Dynamics in Atlanta).

Web 2.0 Creates:
-> Compromised websites – zero-day attacks, third party widgets, auto-executing code
-> More malware access:
  • Client-side processing – more surface areas for attack
  • Syndication – greater transparency for attackers
  • Mashups – complex trust scenarios
  • User Participation – erosion of traditional boundaries
-> More avenues for data leakage – wikis, blogs, trackbacks (described further in this document), emails, instant messaging
Compromised Websites

Google has identified more than three million unique URLs on more than 180,000 websites that automatically install malware on visitors’ machines – spam often is used to drive traffic to these sites simply for the purpose of installing malware for later use.

Even though major virus outbreaks are rarely reported these days, highly regarded websites continue to be hacked. For example:

- In April 2008, malware injection attacks were reported in the UN, UK, USA Today, Target and Walmart websites.
- The Dolphin’s Stadium website was hacked before the 2007 Super Bowl. The website redirected thousands of visitors to phishing and malware sites.
- The Better Business Bureau links sponsored by Google infected users with keyloggers that recorded and stole banking information.
- Tom’s Hardware – a popular technology website – infected users over the course of several days early in 2007 with Trojans that stole personal information from their computer (A More Secure Approach to Dynamic Web Threats: A Frost and Sullivan White Paper sponsored by Postini).

Most organizations today use static URL filtering and anti-virus signature scanning at their web gateway as their primary means to combat web threats. This approach is clearly not sufficient as evidenced by the continued plague of malware.

The type of malware that makes it past static URL filtering at web gateways is most likely of the “zero-day” variety injected into popular websites that are allowed in policies for user access. Signature scanning cannot detect zero-day attacks as, by definition, there are no signatures available for zero-day threats, or the attack uses a custom encryption wrapper to blind the web gateway and only uncloaks itself once on the client system.

The only hope to catch zero-day attacks is to participate in a web-based honey grid ecosystem that detects hidden malware; and employ on-the-fly, dynamic detection capabilities such as behavioral and heuristics-based detection. Without that, many attacks will go undetected.

Consider the following:

- The idea of “good” websites and “bad” websites that can be blocked via URL is no longer viable: As many as 70% of web-based infections come from legitimate sites injected with HTML iframes (sometimes known as malframes). Because these are categorized as “good” websites, they by-pass firewalls, static URL filtering, and reputation scores.
- Fast-flux services constantly change DNS records every few minutes, or 1000s of sub-domains hide the real site, making hostID (the host part of an Internet address) reputation databases useless to mitigate web threats.
- Polymorphic malware (viruses that change with each infection) make traditional, signature-based virus detection less and less effective.
- Predictable traffic increases to legitimate sites, such as sports sites before the Super Bowl, present opportunities for “drive-by” infection if the site is injected with malware.

A website is only as safe as the last transaction or request. Injection attacks can change a popular and trusted website into a malware distribution point in seconds. For web gateway security, this means that every request needs to be analyzed for threats as the trust or reputation model used for email spam filtering does not extend into the web domain. Not analyzing content because a website is noted as having a positive reputation is a huge mistake for web gateway security.
Malware Access

"Web 2.0 technologies, like client-side AJAX, XML syndication, mashups, and social networks create many new input vectors for attackers: they upload malware directly into social networks, insert malicious mobile code (MMC) into HTML documents, and continue to build techniques for traversing domain security settings." [SPI Dynamics in Atlanta]

New Client-Side Technologies

The idea of the Internet as a platform, specifically a beta platform, encourages the deployment of beta software. New technologies (AJAX, Flex/Flash, XAML, REST, XML, JSON, plugins) are more susceptible, may be less well protected, and less well understood by existing anti-malware.

The popularity of AJAX creates more risk, simply because more people (and therefore more attackers) are familiar with it and develop with it. The following list describes how the four technologies work together to expand the attack surface and make attacks more likely to succeed:

- JavaScript (like other scripting languages) provides the “active ingredient” processing capability. A web page that contains JavaScript code is downloaded into a browser, where the code executes automatically. This code runs in the domain of the web page and retains full privileges as if it’s operating on behalf of the user (which, unfortunately, is not always the case).

- DOM and CSS turn the structure and contents of a web page into individual addressable objects that can be manipulated via programs. Standardized structures that enable social networks and other member-oriented websites, such as online banking services, are accessed via the DOM and can be used to mount an attack.

- The XHR capability hides the attack by executing HTTP GETs and POSTs on behalf of the user in the background, effectively separating the user experience from the retrieval and transmission of web content. With XHR’s ability to retrieve data in the background, the user is less likely to recognize an attack in progress and therefore less likely to successfully counter the attack.

AJAX functionality facilitates two types of attacks. First, as with any program where the processing occurs on the client, the local programs that support and parse the commands are susceptible to buffer overflows and similar attacks, which could be exploited to download and install malware such as rootkits and remote control programs that take over the client. This is true with any client-side processing.

The second, less common capability of AJAX is the opportunity for client-side code to mimic any actions that a user can perform on a website. Constraints are in place that attempt to limit access within the web domain from which the code was downloaded, but a number of cross-site scripting and cross-site request forgery techniques exist to bypass this containment and attack any web session that is currently open in the browser.

XML Syndication – Feeds for Wikis and Blogs

XML syndication introduces a new model of content retrieval to the web. Rather than requiring users to access a website by typing in a URL or clicking on a link every time they want to view the contents, the software automatically discovers new information from subscribed feeds and “pulls” it to the endpoint through an automated process. This new technology brings about a handful of new risks:

- RSS aggregators are software, too. Anytime new software is introduced onto the client, vulnerabilities are likely to come with it. These vulnerabilities may be either of two types: software coding errors (bugs) that break the intended software functionality, or emergent errors that abuse the functionality of a program.
Readers may save data locally. Because RSS readers often save their formatted files locally, the files become a resident threat to the client. These files may be opened and manipulated by other software programs or, in some cases, may be laced with malicious code that can manipulate other local content. This amounts to privilege escalation, compared to typical web pages that are configured to be self-contained with no direct access to the local system.

Browser vulnerability. Because browsers are often used to view the data, RSS provides another attack vector for scripting attacks that may be used to compromise systems. This potential holds true for both client software and online aggregators.

About Trackbacks

When a website automatically posts comments to another site it is called a trackback. Blogs and many wikis automatically post comments on a website’s page when comments are enabled. In most cases, the trackbacks and comments are used to create a discussion around a particular subject; that’s why wikis and blogs are key to developing an Internet community. However, wikis are often used in development projects, and development secrets can be inadvertently given away through trackbacks. Even secured blogs can be exploited through the trackback vulnerability.

Mashups

The key risks associated with mashups are:

- Reliance on a third party’s security posture. Because mashups connect dynamically to third-party websites, the integrity of the data and programs being accessed on those websites is either assumed or validated in real time. If the content provider is compromised, the damages may be inherited by the website host as well.

- Availability issues are transitive. If the content provider has a problem with availability due to overuse or denial of service (DoS) attacks, that problem is inherited by all other service providers with respect to their own users. Users will simply see the issue as associated with the service provider itself.

- Responsibility for attacks against the API. It remains unclear how two entities should handle a trust relationship when a website accessing an API unknowingly facilitates an attack against that API. An inadvertent attack like this is essentially indistinguishable from a cross-site scripting attack.

- Liability associated with intellectual property misuse remains unclear. With many media sites such as Flickr and YouTube facilitating the use of multimedia content, the rights associated with the display and distribution of this content elsewhere are still being determined.
SSL/HTTPS Traffic
The growth in outsourced applications and private and/or protected web use on port 443, usually blind to IT, opens the door to malware attacks using Secure Sockets Layer (SSL). This is illustrated in Figure 3.

![Diagram](image)

**Figure 3: SSL Provides a Private Link for Legitimate Applications, AND Malware, Confidential Data, Unsanctioned Traffic**

Because SSL content is encrypted, it can’t be intercepted by security systems without being decrypted and re-encrypted. Users can bring in various malware (including viruses), access forbidden sites, and leak confidential business information over an HTTPS connection. Because IT organizations often have no visibility into SSL sessions traversing their network, they are blind to any potential security threats sent over HTTPS. In addition to the security threat, encrypted traffic makes it difficult for IT to assess bandwidth usage and apply intelligent content control policies to ensure maximum user productivity.

Data Leak Vulnerabilities
Data leakage caused by removable media, spyware and malware, and the resulting regulatory compliance issues, remains at the top of enterprise IT challenges. Most data leaks and targeted attacks – inadvertent or intentional – occur at the endpoint and many of these are generated internally. Unmanaged removable media and applications can easily open the floodgates for data to escape into the wrong hands. Recent examples include:

- 320,000 sensitive files were allegedly transferred to a thumb drive by a Boeing employee and leaked to the Seattle Times.
- 8,000 Texas A&M Corpus Christi students’ personal information, including social security numbers were lost in Madagascar when a professor vacationing off the coast of Africa took the data with him on a flash drive.
- Data thieves breached the systems of credit card processor CardSystems Solutions making off with data on as many as 40 million accounts affecting various credit card brands, according to MasterCard International.
- The entire database of child benefit recipients maintained by Her Majesty’s Revenue and Customs (HMRC) department has gone missing after being posted to the National Audit Office by a junior official. (darkREADING)
- Some 4.6m customers of Japanese broadband outfit Yahoo! BB – which is run by Softbank and Yahoo! – had their personal details compromised. Criminals allegedly hacked into Softbank’s database and copied the data after obtaining a password from a former Softbank worker. The company is set to spend around four billion yen (£20 million) compensating those hit by the leak. (The Register)
Seventy-five percent of FORTUNE® 1000 companies fell victim to accidental and/or malicious data leakage. The cost to recover when corporate data was leaked or stolen was an average of nearly $5 million in 2006, 30% more than in 2005. Furthermore, malware attacks, unauthorized access to networks, lost/stolen laptops and mobile hardware, theft of proprietary information or intellectual property accounted for more than 74% of financial losses. (Lumension Security)

Blue Coat Web 2.0 Layered Security Solutions

How can one combat web 2.0 security challenges? How to defeat expert programmers with plenty of money and time to devote to their malware craft, beta software that may not be fully vetted, collaborative websites that depend on several people maintaining their own security, and new technologies that put processing at the client where numerous vulnerabilities may exist? Blue Coat can help. Blue Coat ProxyAV + ProxySG layered defenses are illustrated in Figure 4.

Web content is very dynamic; URL databases rate less than an estimated 20% of the Internet, and new attacks are coming in short time bursts in popular websites. Signature updates for threats, websites, and reputations, often take hours (and sometimes more than a day) to develop, and leave networks exposed.

Blue Coat security appliances provide Secure Web Gateway (SWG) technologies that include three types of zero-hour protection:

- WebFilter categories for “Spyware/Malware Sources” to block URLs/hosts, and “Spyware Effects/Privacy Concerns” to block calling home efforts
- Proactive anti-malware threat detection engines in ProxyAV
- A real-time web content rating service in ProxySG with WebFilter

Additionally, the Blue Coat WebFilter URL database is continuously updated by the Blue Coat WebPulse™ collaborative defense using more than 10 threat-detection engines across a honey grid of clients to uncloak attacks and analyze nearly 1 billion requests per day for hidden malware. This allows ProxySG to block outbound XHRs to malware hosts from injected scripts in popular websites, so users can still safely visit these popular websites. WebPulse acts like a neighborhood watch program. Each user request updates the WebPulse ecosystem for the benefit of all users by providing a constant flow of user requests for malware analysis, allowing new sites to be rated with cross-categorizations and reputation ratings so gateway policy decisions can advise, filter, or block web content.

For devices inside a network, layered defenses of a Secure Web Gateway provide a key advantage. A layered defense provides five important elements for protection:
1. WebPulse honey grid updates to WebFilter of malware sources hidden in popular websites; thus malware hosts can be blocked while access to popular websites with injection attacks can still be allowed.

2. Leading anti-malware threat detection engines with heuristic/behavioral analysis to detect zero-hour threats.

3. URL filtering with a real-time content rating service to limit web exposure and detect Phishing attacks.

4. Web content controls for true file type detection, active content and certificate validations, method-level controls, content filtering, and active script filters; plus a robust policy language.

5. Optional Data Leakage Prevention/Information Leakage Prevention (DLP/ILP) integration with leading DLP/ILP solutions to validate outbound data & information.

Secure Web Gateways not participating in honey grid ecosystems that continuously analyze user requests and, most importantly, simulate clients to uncloak attacks, cannot keep pace with malware injected into popular sites. Static URL lists, host reputations, and signature-based engines do not play in the Secure Web Gateway arena these days; however, that is a common profile for web gateways deployed over the past few years.

Five Secure Web Gateway solutions are discussed below:

- **Preventing Malware**
- **Dynamic URL Filtering**
- **Preventing Data Leakage**
- **Securing SSL Traffic**
- **Monitoring and Logging**

### Preventing Malware: Layered “Defense in Depth”

Malware prevention cannot be achieved by one method alone; rather, it is best to deploy many layers of defense – including the firewall and desktop-based solutions.

Blue Coat ProxySG provides a powerful toolkit for blocking mobile malicious code (MMC) and preventing malware infection encountered on the Internet. With the addition of Blue Coat WebFilter (BCWF) on-box and dynamic real-time rating (DRTR), rules can be selectively applied so that the “bad neighborhoods” of the Internet receive additional protective measures. Finally, Blue Coat ProxyAV provides heuristic/behavioral analysis and signature scanning, to augment your desktop-level AV product.

Four policies are briefly described:

- **Block Access to Known Malware Sources**
- **Detect Hidden/Cloaked File Types**
- **Validate or Remove Active Content**
- **Block MMC and Malware with ProxyAV**
**Block Access to Known Malware Sources**
Firstly, not accessing a malicious website guarantees avoiding infection by it. To that end Blue Coat WebFilter (BCWF) provides the “Spyware/Malware Sources,” “Spyware Effects/Privacy Concerns,” and “Phishing” URL categories. WebFilter is continuously updated by the WebPulse ecosystem that analyzes close to 1 billion requests per day for hidden malware injected into popular websites with the goal of blocking the malware host, not the popular website. Blocking access to these categories of websites removes the risk of infection from known bad sites. The dynamic real-time rating service of BCWF includes automatic detection of phishing kits, phishing look-alike web pages, and untrustworthy, temporary DNS-registered hostnames, to protect against new and emerging threats. The point with this method is to remove known malware/bad sites up front.

**Detect Hidden/Cloaked File Types**
The defense in depth strategy continues with detailed tests to prevent misrepresentation of true file types, or container mismatch. Too often today, malicious executable content is misrepresented as safe file types such as “JPG” or “GIF”. Blocking this makes use of policy tests comparing the claimed file type to the actual initial data in the files.

**Validate or Remove Active Content**
Mobile Malicious Code (MMC) is not an “executable” in the traditional “exe” file sense, rather it exploits vulnerabilities in the browser (or other client application) software through malicious JavaScript, VBScript, Flash, or ActiveX modules. Protection against these can take several forms from stripping all active-content from pages, to selectively “de-fanging” malicious code methods, and/or signature/heuristic scanning.

The safest option that still allows access to web pages is sanitizing the HTML to remove all active-content; however, this has significant impact on today’s interactive web 2.0 sites. Due to the risk of over-blocking, this option should be applied in conjunction with BCWF to only occur on the riskiest, least business-oriented sites. Any exceptions can then be handled by white listing.

**Block MMC and Malware with ProxyAV Proactive Detection Engines**
Assuming all the above levels of defense are applied, a significant reduction in malware can be achieved with ProxySG and WebFilter alone; however, no solution is complete without leveraging the billions of dollars spent each year by the antivirus industry on heuristic/behavioral based detections. This is available via the Blue Coat ProxyAV with leading anti-malware engines: Sophos, Kaspersky, McAfee, Trend Micro, Symantec, and Panda.

Blue Coat recommends that all objects be subjected to such scanning regardless of their website of origin, or file type. The ProxySG + ProxyAV solution makes this workable via the “scan once, serve many” technology. Each time a browser request is received, the ProxySG checks its object store for a cached copy, if one is found that was analyzed with the most recent AV-heuristics engine version, it can be delivered immediately. If the object in cache was analyzed before an update to the AV-heuristics engine, then the object is re-analyzed before delivery to the user. For non-cacheable objects, a finger print cache is kept to avoid analyzing the same file on frequent requests. Once an AV-heuristics engine update occurs, the finger print cache for non-cacheable objects restarts.
Dynamic URL Filtering with WebPulse and WebFilter

Recent research shows that on average, one in ten websites host some form of malware. New phishing and malware-hosting sites emerge in great numbers every day. These malicious websites are becoming increasingly short-lived to avoid being blacklisted. Without some form of near real-time website reputation, pre-classified URL lists simply cannot keep up with this fast-changing threat landscape.

WebPulse Ecosystem URL Categorizing

The Blue Coat WebFilter solution provides a real-time rating service of websites to cover the estimated 70% of websites on the Internet (plus new ones each day) that are unrated. Each user request updates the ecosystem for all Blue Coat customers as the WebPulse cloud service dynamically rates unrated websites and analyzes all requests for malware.

The WebPulse ecosystem is driven by over 100M user requests per day that make Blue Coat WebFilter relevant, accurate, and dynamic for URL filtering. Real-time ratings cover 98% of objectionable content sites in multiple languages, plus phishing kit detection, and the ability to search deep into translation services, image searches, and cached search engine content, for an accurate rating not found with static URL filtering solutions. WebPulse is a very large honey grid with a high volume of user requests combining client requests from the K9 family Internet protection, enterprise ProxyClient remote users, enterprise ProxySG deployments, and a tremendous volume of requests from ISP deployments of ProxySG with WebFilter. No other URL filtering system provides real-time ratings and is driven by user requests for unmatched filtering protection of web 2.0 content.

Figure 5, below, illustrates the WebPulse Ecosystem.
**WebFilter: Filtering & Blocking Unwanted Web Content**

Accurate coverage is just as important as relevant coverage in selecting a URL filtering solution. To ensure the greatest accuracy, each site in the WebFilter database is classified into multiple categories, as appropriate. This allows WebFilter customers to define a virtually unlimited number of “cross-categories” via ProxySG, to fit their specific filtering requirements (for example, Motor Sports, Sports Betting, etc.).

The Blue Coat ProxySG supports up to three “on-proxy” content filtering lists at once. This allows customers to run a commercial grade URL list like WebFilter, perhaps the IWF (Internet Watch Foundation) sealed child pornography list, and a custom allow/deny URL list for their company.

Blue Coat WebFilter on ProxySG is automatically updated and configurable by administrators. Allow/deny lists can be created on Blue Coat ProxySG, plus new categories, overrides, and exceptions. Administrators can define flexible policies using header request elements not provided by content filtering solutions. With the rise of image search engines, these header controls complement URL filtering to optionally ensure SafeSearch modes are always enabled, no matter what the user selects. Blue Coat can also apply policy to embedded URLs commonly missed by traditional filtering platforms, such as translation sites, archives, and image searches. Blue Coat can apply policy based on the category the embedded content is in, since that is the actual “destination” URL.

The Blue Coat WebFilter database contains more than 15 million ratings, primarily of domains, directories, and IP addresses, which in aggregate covers more than one billion web pages, published in more than 50 languages, organized into more than 80 highly useful categories, and updated constantly by user requests into the WebPulse ecosystem.

Unlike other vendors, Blue Coat does not rely on site-mining processes such as web crawlers to find new websites and applications. Instead, Blue Coat leverages its large URL research ecosystem of WebFilter customers and off-proxy OEM deployments to refer sites for categorization. These customers feed nearly 1 billion requests each day into the Blue Coat WebPulse automated rating architecture, which uses a combination of proprietary, patent-pending classification software and human inspection, to categorize sites and review classifications. WebPulse also uses a large honey grid of clients with more than 10 threat detection engines to detect malware hidden in popular websites to continuously update WebFilter of malware hosts.

The relevance and accuracy of the WebFilter database is driven only by the actual web usage patterns of customers. As a result, WebFilter has an average 92-95% coverage rate of requested sites in enterprise deployments, coupled with an extremely low rate of false-positives.

**Preventing Data Leakage with ProxySG**

Blue Coat ProxySG supports multiple Data Leak Prevention/Information Leak Prevention (DLP/ILP) solutions from leading vendors in the DLP/ILP niche. This allows customers their choice for an overall DLP/ILP solution knowing the integration is certified, and DLP checks can be made in web traffic analyzed by ProxySG. The Blue Coat ProxySG supports standard ICAP (RFC 3507) and certified implementations are currently available for the following vendors:

DLP/ILP solutions certified and supported:

- Blue Coat Data Loss Prevention
- Code Green Networks
- Vericept Network Monitor/Prevention
- Vontu Prevent
- Reconnex iGuard
- Port Authority/Websense
Inspecting SSL Traffic

The Blue Coat ProxySG provides HTTPS (or SSL/TLS) traffic inspection at multiple levels to reduce web threats and data leakage. HTTPS sessions can be optimized based on destination, user, group, or other policy variables such as personal medical website use, with complete privacy through the web gateway. HTTPS transactions with financial sites can involve server certificate validation only, while leaving the communication private. Or HTTPS sessions for business can be inspected for content leaving the organization, and malware and threats entering the organization’s network. Also, HTTPS server certificate validation checks are useful to block proxy avoidance tools or malware trying to call home.

Monitoring and Logging

Along with web security solutions, companies need to capture and analyze highly valuable audit log data produced by their IT systems to measure the effectiveness of their web security policies, to help substantiate compliance, and to respond quickly to security and forensics incidents.

Industry experts agree on the importance of log management as a cornerstone of any organization’s risk management, compliance, and privacy protection strategies. Log data can provide a complete real-time and historical record of access, activity, and configuration changes for applications, servers, and network devices. It can also be used to aid security and privacy policy validation. In addition, IT managers can use log data to receive early warning of potential security and performance problems and can mine log data for root-cause analysis to aid in system recovery and damage cleanup after security or performance incidents.

Blue Coat Reporter analyzes comprehensive log files from Blue Coat ProxySG in over 150 pre-defined reports, including malware, spyware, IM, P2P, and popular sites. Beyond URL filtering, Reporter provides visibility to web content, performance, threats, and trending, over defined time periods. With comprehensive, policy-enabled logging, Blue Coat Reporter with ProxySG provides the advantage of capturing data regarding all user web activities. Reporter quickly processes robust log data, providing easy-to-view reports for department managers, HR managers, security specialists, and network administrators. Blue Coat provides the ultimate architecture for complete web visibility and control.
Conclusion

Web 2.0 is here to stay; some experts are already talking about web 3.0. To take advantage of the benefits and still keep the enterprise safe is the challenge. Popular websites are now trusted targets for injection attacks of scripts that call on hidden malware host servers often using obfuscation techniques to hide themselves. Cloud services with large honey grids are now creating hybrid web gateways that also use real-time web content rating services and proactive threat detection engines to stop malware downloads. web 2.0 content is constantly changing with a two-way publishing model; real-time analysis is a requirement.

Blue Coat offers a powerful, proxy-based architecture that accelerates good web traffic and stops malware from entering your organization. It’s the solution you need to:

- Prevent malware, spyware, viruses, and other threats from reaching desktops and remote devices
- Categorize web traffic for policy controls, plus apply real-time ratings to new websites and content
- Inspect all web communications including SSL traffic, plus control IM, P2P, and Streaming Media traffic
- Control web content: validate active content, true file types, plus leverage blended policies to content-filter web objects from suspicious or unrated websites
- Integrate leading DLP solutions to inspect web gateway traffic for data leakage
- Allow users to safely view popular websites while blocking hidden injected malware downloads

Blue Coat security appliances provide zero-hour protection with the ProxyAV proactive anti-malware engine and ProxySG with Blue Coat WebFilter real-time web content ratings. The Blue Coat WebPulse ecosystem analyzes nearly 1 billion requests per day to continuously update WebFilter with new malware sources often hidden in popular websites. Blue Coat layered defenses for malware provides enterprise protection not seen from competitors. Blue Coat offers the widest choice of anti-malware and URL filtering solutions of any secure web gateway vendor.
Sources

In addition to the quoted sources, the following provided source material for this document:

12. eWhisper.net Brad Geddes’s Thoughts on Advertising, PPC, and the Web
13. Gartner’s IT Security Summit in Sydney
14. Technorati
15. Computer World “Blogs Bubble into the Workplace”
17. Postini’s Annual Communications Intelligence report
18. Lumension Security, “Prevent Data Leakage and Costly Security Breaches” (Data Leak Vulnerabilities section of this document):
   a – “Boeing Employee Charged with Stealing 320,000 Sensitive Files”, Information Week, July 11, 2007
   c – “Credit card breach exposes 40 million accounts”, CNET News.com, June 17, 2005
   e – Ponemon Institute’s 2006 Cost of Data Breach Study