WHITE PAPER: SECURITY

The Hidden Menace of Embedded Links

Symantec.cloud

White Paper: Security

The Hidden Menace of Embedded Links

Contents

Introduction

As the arms race between malware writers and the IT security community continues to escalate, spammers and blackhats are using ever more sophisticated methods to attempt to bypass detection systems and achieve their objectives.

Many companies now filter out email attachment types thought to be harmful, and this has driven malware authors to find ways to bypass this strategy. As a result, a substantial amount of malware is spread simply by sending a link to the malware, together with some social engineering to try and induce the recipient into clicking on the link. Symantec.cloud has observed this trend since 2000 and recent months this trend has increased dramatically.

This trend has presented traditional anti-virus and anti-spam engines with a problem. Since the malware itself is not present in the email, conventional scanners are unable to detect even simple and well known threats. The inability to identify potential email threats has also exposed a weakness in corporate email security strategies. Quite simply, by switching the focus of malware detection from email scanning to HTTP, malware authors know they are attacking the weakest link in corporate security.

Symantec.cloud takes the view that both SMTP and HTTP scanning at the Internet level is now essential in order to ensure reliable security to internal networks. Indeed the two approaches are rapidly converging to create a comprehensive defense to protect customers' networks from viruses, malware and other unwanted content.

Problems of a Conventional Approach

In order to block malware, some email security vendors may decide to download and scan each link that is present in the email. But this is not an ideal solution for several reasons.

Unpredictible Processing Time

An email can contain a large number of links, many pointing to substantial quantities of data. Moreover, some links are at the end of very slow connections. This creates a long and unpredictable delay at the time of processing the email.

Accidential Confirmation

Some email links are accompanied by text such as 'Click here to unsubscribe', 'Click here to cancel your insurance' and perhaps even 'Click here to confirm your order of two new aircraft carriers.'

The very act of 'touching' (i.e. attempting to download from) these links causes the intended action to occur. Some links are also 'web bots' used by spammers to confirm that email addresses exist. Thus, touching these links guarantees the email address will be targeted for spam in the future.

IP Switching

Links to malware are often deliberately tampered with by blackhats, so that attempting to download the contents does not always lead to the malware, but to a harmless web site. There are several tactics in common use. For example, the malware can be set only to appear some time after the email has been sent, so that the malware code itself will not be present when the mail engine tries to resolve the link — but will be there when the recipient attempts to download.

Another option is to set the malware to appear only if the downloading IP address is either within a permitted range, or excluded from a disallowed range. For instance, phishing attacks on American banks have been crafted so that the phishing web site only appears to American IP addresses. Attempts to resolve the link from outside America result in a harmless site appearing. In a further twist, if the American IP address belongs to the bank under attack, then the bank also gets served with a harmless site.

No doubt many other schemes are possible. Scam artists certainly have the time and resources on their side to keep innovating. As a result it is impossible to guarantee that the object you scan at the time of processing an email is the same object that will appear when the email recipient clicks on the link.

Analysis of the Problem

Given the limitations set out above, Symantec.cloud has developed a number of unique solutions to stop threats posed by embedded links in email.

There are several proactive and reactive strategies that can be deployed to mitigate the chances of missing an email containing malicious links. And our philosophy is that the best defense will always be a combination of strategies. But first it is critically important to understand the real underlying issues.

Link Exploits

Several link exploits are known and emails that contain one or more of these exploits can be considered malicious, with a very low risk of false positives. Other exploits are less easily identified. Detection of one of these exploits is a good indication that the email is malicious — though not a definite sign.

The ZapfDingbat exploit

This uses links containing the hex characters 0x01 or 0x00. Unpatched versions of Internet Explorer display only the characters to the left of these characters, and clever URL construction can ensure that the subsequent displayed text is not related to the actual web site connected to. For example:

http://www.microsoft.com%00@www.badhacker.com

Thus a URL can be constructed which appears to point to a specific site, while in fact pointing to another.

Typosquatting

Using this exploit, a blackhat registers a domain name similar to the target domain. Combined with email, this type of attack is most often used for phishing, although typosquatting can also be used for non-email attacks, such as URL mistyping. For example:

http://www.barcalys.com

http://www.barclays-security.com

Defending against this type of attack is difficult. It is necessary to maintain a list of domains that could be attacked, and then to create algorithms that detect near misses. The quality of defense depends on the completeness of the list of attackable domains, plus the quality of the near-miss recognizer.

IDN Attacks

The IDN attack is similar to typosquatting, except that an extended alphabet of international domain name (IDN) characters is used, including accented characters. So it is possible to construct a domain name that looks similar to a real domain name. For example:

http://päypal.com

As with typosquatting, this exploit is typically used in a phishing attack, luring the victim to a counterfeit site where sensitive information can be extracted by clever social engineering. Defending against this type of attack is similar to protecting against typosquatting, but because the number of changes is more limited, the chances of successful detection are greater.

Url Containing Usernames and Passwords

The URL specification allows for the addition of a username and password which, if present, are inserted before the domain name. This enables the attacker to create a very misleading URL. The username is crafted to look like a domain name, and the real domain name is hidden thereafter. For example:

http://barclays.com: security.cgi @121.27.11.201

This type of URL is not common in legitimate usage — there is not much point in sending a username and password in clear text over the Internet. Therefore strongly weighted heuristics can be used for detection.

Suspicious Constructs in Links

IP Addresses

The blackhats need to find a web site to host their malware, and this can be achieved in various ways. Using their own machines is a non-starter, because this greatly increases the possibilities of arrest. Usual tactics include compromising an unsecured legitimate domain, hosting malware on a free hosting site and using a PC which has been compromised by an earlier virus or Trojan.

If they choose the latter path, then by and large the compromised PC will not have an associated domain name. They can register a new domain and use DNS to forward it to the new IP, but this involves significant work and does not happen often. It is likely therefore that the blackhat will reference the compromised PC by IP address, rather than by domain name.

Using IP addresses in URLs is comparatively rare, and indicates a high probability that the link points to a compromised PC. However it is not a definite sign. Indeed, Symantec.cloud processes several thousand legitimate emails each day that use IP addresses instead of domain names.

The presence of an IP address instead of a domain name therefore remains only a good clue, indicating that further investigation is needed, rather than a sure sign of malware. Other suspicious features include a bank name in the path, which is usually a good indicator of a phish, and a link pointing to an executable file, which strongly suggests a link to a trojan.

Ports of Entry

The standard port for hosting HTTP traffic is port 80, so any link pointing to a different port is suspect. However, there are plenty of exceptions to the rule, and Symantec.cloud processes thousands of emails daily with links pointing to ports other than port 80.

False positives can be reduced by maintaining lists of commonly used legitimate IP+port or IP+domain combinations. There are also some applications that generate legitimate emails with links formatted in this way, so recognizing these emails can also help reduce false positives.

Free Hosting & User Accounts

Malware authors commonly serve out their malware from free hosts like AOL or Netscape. It is fairly unusual for business email to contain links to executables located on free hosting sites, so detection of an executable hosted at a free web site is usually a reliable indication of a trojan being present.

The Brazilian bank-trojan gangs often spread their password stealers in this manner, sending out an email designed to look like a postcard:

http://hometown.aol.co.uk/caixa1012cart/Cartao.exe

Similarly, executables hosted at a user account directory are also suspect. User account directories can be identified by a directory name starting with a twiddle, such as:

http://8.10.120.10/~samsungf/musical.scr

http://24.195.242.82/~hamid/images/saudades_0512.scr

Link Hiding

Because malware authors are hosting malware on a PC that does not belong to them, they need to keep their presence hidden from the legitimate owner. They may therefore use unusual directory paths to the malware that would not occur in legitimate email. Examples of this might be:

http://210.205.6.34/.../T9904295V52L12J009385291029WT053

http://66.237.222.11/.%20/cib.ibanking-services.com/

The '...' sequence and the %20 character are both highly suspect.

URL Encoding

Legitimate emails never use obscure URL encoding. However, spammers and malware authors often use this technique in a bid to evade filters. They also use it as a tactic to confuse and misdirect the email recipient:

http://%36%32%2E%31%39%37%2E%37%32%2

Detection of obscure URL encoding is a sure sign of an unwanted email, although classifying as spam or malware is often a problem.

Redirection

Some URLs redirect through intended or accidental redirection services such as Google or MSN. For the most part this is an obfuscation technique designed to bypass filters or confuse the email recipient. However, there can be occasional legitimate uses of these redirection services.

Multiple redirections, where a link passes through two or more redirectors, are always malicious. Examples might be:

http://go.msn.com/HML/6/5.asp?target=http://www.g	oogle.sc/url%3fq%3d

http://%09%73c2%62j%648%%%.%%09d%%%A%%252E%%%%%2552%%%09%2575/

Environment Analysis

Just considering the link in isolation can, as we have seen, provide many clues. However considering the link within the wider context of the containing email can also divulge many fruitful hints

Aliasing

It is fairly easy to use HTML to disguise the true destination of a link. For instance, the following link will appear to point to www.microsoft.com, a potentially trusted site. However it actually points to malware.com, a much more dubious destination.

<A url='malware.com'>www.microsoft.com</A>

This technique is called aliasing. The problem is that many legitimate emails also contain aliasing; for instance, this is a common technique used by bulk email senders. When these businesses send out email campaigns on behalf of their clients, they use their client as the destination to which the link apparently points.

However the link actually points to their own servers. When the email recipient clicks the link, the connection is first made to the bulk email sender's website, where it is recorded for tracking and statistical purposes. The request is then forwarded to the client server. Another common legitimate situation is where one link is a subdomain of the other. Therefore, it is not possible to say that every aliased link is indicative of malware and the email must be investigated for further clues.

Lists of commonly used bulk email servers and alias pairs can be used to help reduce false positives. Other clues can also help in making the decision. For instance, if the real link is actually an IP address, this is more suspicious than if it is a domain.

A mismatch in file types is also suspect — for example, if the real link points to an .exe file, and the fake link is a .html file, this is very questionable.

Text

Many blackhats operate their scams for months, so it is difficult for them to come up with different social engineering emails each time. As a result, there tends to be a relatively small number of phrases that are used time after time to induce the victim to follow the link.

Analysis of the email text can be used as a powerful clue. From time to time malware writers introduce new phrases, but rarely change the whole email. Thus, as long as the phrase list is kept up-to-date, detection is extremely likely.

A typical example:

> Dear eBay user,
>
> During our regular update and verification of the accounts we could not verify your current information. Either your information has changed or it is incomplete. As a result your access to your eBay account will be restricted.
>
> According to our site policy you will have to confirm that you are the real owner of the eBay account by completing the following form within 24 hours or else your account will be suspended without the right to register again with eBay.
>
> In order to complete this verification just click the link below.
>
> eBay Customer Support

A typical second example:

> Dear eBay user,
>
> Either your information has changed or it is incomplete. As a result your access to your eBay account will be restricted. During our regular update and verification of the accounts, we could not verify your current information.
>
> Our site policy requires you to confirm that you are the real owner of the eBay account by completing the following form within 24 hours or else your account will be suspended without the right to register again with eBay.
>
> In order to complete this verification just click the link below.
>
> eBay Customer Support

Email Headers

Blackhats tend to use the same bulk email engines each time to spam out links to trojans. Similarly, new viruses sending out links also often use email engines exploited by previous viruses. This means that the email headers are very similar from run to run, and often look quite different to legitimate emails.

Thus, in cases of doubt, the email headers can often be used as a good differentiator. For example:

From: “Citi” user-support3@citibank.com
X-Mailer: The Bat! (v2.00.6) Business
Reply-To: “Citi” user-support3@citibank.com
X-Priority: 3 (Normal)
Message-ID: 1666540507427598254366158550048338@yahoo.com
To: bill.hook@xx.xxx.org
Subject: official Notice for all users of Citibank!
MIME-Version: 1.0
Content-Type: multipart/mixed;
boundary=”----------BC4D7C15E946920”

HTML Constructs

The HTML used in the email is often very different from HTML generated by standard email clients or mailing engines. Sometimes this is because there is an attempt to defeat filters; sometimes it is just laziness or ignorance on the part of the blackhat generating the email.

An example of laziness would be encoding the <HTML> tag in lower case, as <html>.

An example of filter evasion would be to use text reversing elements to disguise the text used in the email, or to include the text in a picture, so that a human could easily read it, but a filter looking for specific elements could not.

Forms, Maps and Objects

In HTML email, most links use <IFRAME> or <A> tags. However, in an effort to avoid filters and make detection harder, blackhats occasionally put links in other, more obscure places such as <MAP> <FORM> or <OBJECT>. These constructs occur relatively infrequently in email, and so can be treated as highly suspect.

They do legitimately occur sometimes, but it is relatively easy to construct a set of rules picking out the good from the bad.

<object data=”ms-its:mhtml:file://C:\\MAIN.MHT!

http://privatemailboxrentals.com/main1.chm::/main1.html”

type=”text/x-scriptlet”></object>

Ironically, the more the blackhats use special tricks in an attempt to obscure or confuse, the more their emails differ from legitimate mail, and so the easier it is to detect them.

Combined Approach

Putting together a successful attack is hard work, involving many elements. For instance, to spam out a link to malware may require getting hold of a set of compromised PCs to send out the email, creating the email, finding a compromised PC to host the malware, creating the malware and, of course, arranging in some way to benefit from the malware.

This is a lot of effort and these attacks are occurring on a constant basis. The malware writers vary their attacks slightly each time to avoid detection, but in general they do not create a completely new attack each time.

Therefore, for instance, even if they discover a completely new link exploit undetected by scanning systems and scoring 0, there is still an extremely high chance of detection from other factors, such as the text of the email, the engine used to send the email, and so on.

This is a similar concept to the virus DNA that Symantec.cloud uses successfully against executable malware. Symantec.cloud carries out continual monitoring of threats propagated by email links, and adapts detection strategies to meet new behavior patterns as they arise.

Symantec.cloud Services 0ffer Protection

Link Signaturing

We can identify undesirable links via several routes, backed up by a fast link signaturing capability. This enables us to put defenses in place very quickly whenever we learn of a possible new link. Symantec.cloud signaturing allows for both exact matches, partial matches and pattern matches, to cover the range of different incidents which may occur.

Recursive Link Following

For those links we are unsure about, actually downloading and examining the content can provide a useful second opinion. Of course it is always possible for the attacker to feed us a non-malicious sample, and the email recipient a different, malicious sample. However, currently this is still a rare phenomenon, and so this is a good reactive method. The faster this can be done, the quicker protection can be put in place and customers protected.

Symantec.cloud also gathers intelligence to prevent downloading of 'action' links that may accidentally trigger some side effect.

For performance reasons, link following is not done in real time; rather we use hardware separate from the main email processing engines. Suspect links are fed into a scheduling system that processes the links as soon as possible. The scheduler has a level of intelligence, so the links we think should be processed first are given priority.

The link is downloaded and scored by several AV engines, including Skeptic and the commercial scanners we currently use. From time to time we sometimes use other AV engines or security products as well.

Occasionally, a link's contents may refer to other links. In this case, these are also scheduled for link following up to a maximum depth, so the system does not infinitely loop. Otherwise a malicious attacker could DoS the service through a clever choice of links.

If the link contents are found to be undesirable, then various actions take place. These include signaturing the link so that future references to it are blocked. Internal alerts are generated so that customers to whom the link got through can be contacted. Interested parties may also be contacted, and the sample(s) stored for possible further investigation.

Traffic Analysis

Even if the link contents cannot be downloaded, or are manipulated by the attacker to present harmless content occasionally, it can still be possible to use traffic analysis to determine whether a series of emails are malicious or harmless. This will not catch a one-off targeted attack, but will catch a mass mailed link to a trojan or virus.

Traffic analysis works by building up patterns of known good and bad emails. When a suspicious new link is detected, the email pattern is compared to the known good and bad patterns and a decision taken as to which category it belongs.

As an example, a new link appearing in emails originating from many different IP addresses is very suspicious. This looks like someone spamming out a link to a trojan from a number of compromised PCs. A new link appearing in emails originating from one IP address is not so suspicious — looking like some type of legitimate mail shot. Symantec.cloud has two patents for traffic analysis that are applicable to this area.

Sometimes the traffic analyzer is unable to make a decision, but has seen enough copies of a link to make it advisable that a decision is made either way. In this case the decision can be escalated to a human operator who can then make the necessary decision.

HTTP Scanning

The best defense is to provide both SMTP and HTTP scanning. If a malicious link passes through email scanning, it will be caught by the HTTP scanner. It does not matter if the attacker sometimes presents harmless objects at link resolution time; if a harmless object is presented it gets through, but if a harmful object is presented it is stopped. In neither case does any harmful object get through to the customer.

There are some other protocols that are common in links, which can also be covered by this method. One popular example is FTP. Links to FTP traffic can be treated in much the same way as HTTP.

HTTP scanning defends against protocol-related attacks. Because the attacker cannot connect directly to the target, but has to go through the HTTP proxy, these attacks can be eliminated as long as the proxy is aware of them. Even if not aware, a well-coded proxy may eliminate most attacks anyway as a by-product of the way it works.

Link Altering

This technique would be useful for customers who have SMTP scanning but not HTTP scanning. There is little point in offering this for customers who plan to use Symantec.cloud HTTP scanning services, although it would be a potentially useful add-on for customers who use other providers' HTTP scanning services. Currently, this technique is being evaluated, and is not in live usage.

If implemented the email service would change all links to point to HTTP servers under our control, and record a mapping between our link and the original link. When the email recipient reads the emails, the link is resolved and our HTTP server contacted. Our server contacts the original server and downloads the object. The object is then scanned and, if found to be safe, passed on to the client. This technique can be bypassed by emails that do not contain the link directly, but contain scripts which alter the email to add the links.

More Information

AMERICAS

UNITED STATES
512 Seventh Avenue
6th Floor
New York, NY 10018
USA
Toll-free +1 866 460 0000

CANADA
170 University Avenue
Toronto, ON M5H 3B3
Canada
Toll-free :1 866 460 0000

EUROPE

HEADQUARTERS
1270 Lansdowne Court
Gloucester Business Park
Gloucester, GL3 4AB
United Kingdom
Tel +44 (0) 1452 627 627
Fax +44 (0) 1452 627 628
Freephone 0800 917 7733

LONDON
3rd Floor
40 Whitfield Street
London, W1T 2RH
United Kingdom
Tel +44 (0) 203 009 6500
Fax +44 (0) 203 009 6552
Support +44 (0) 1452 627 766

NETHERLANDS
WTC Amsterdam
Zuidplein 36/H-Tower
NL-1077 XV
Amsterdam
Netherlands
Tel +31 (0) 20 799 7929
Fax +31 (0) 20 799 7801

BELGIUM/LUXEMBOURG
Symantec Belgium
Astrid Business Center
Is. Meyskensstraat 224
1780 Wemmel,
Belgium
Tel: +32 2 531 11 40
Fax: +32 531 11 41

DACH
Humboldtstrasse 6
Gewerbegebiet Dornach
85609 Aschheim
Deutschland
Tel +49 (0) 89 94320 120
Support :+44 (0)870 850 3014

NORDICS
St. Kongensgade 128
1264 Copenhagen K
Danmark
Tel +45 33 32 37 18
Fax +45 33 32 37 06
Support +44 (0)870 850 3014

ASIA PACIFIC

HONG KONG
Room 3006, Central Plaza
18 Harbour Road
Tower II
Wanchai
Hong Kong
Main: +852 2528 6206
Fax: +852 2526 2646
Support: + 852 6902 1130

AUSTRALIA
Level 13
207 Kent Street,
Sydney NSW 2000
Main: +61 2 8220 7000
Fax: +61 2 8220 7075
Support: 1 800 088 099

SINGAPORE
6 Temasek Boulevard
#11-01 Suntec Tower 4
Singapore 038986
Main: +65 6333 6366
Fax: +65 6235 8885
Support: 800 120 4415

JAPAN
Akasaka Intercity
1-11-44 Akasaka
Minato-ku, Tokyo 107-0052
Main: + 81 3 5114 4540
Fax: + 81 3 5114 4020
Support: + 852 6902 1130

About Symantec.cloud

Symantec.cloud uses the power of cloud computing to secure and manage information stored on endpoints and delivered via email, Web, and instant messaging. Building on the foundation of MessageLabs market leading software-as-a-service (SaaS) offerings and proven Symantec technologies, Symantec.cloud provides essential protection while virtually eliminating the need to manage hardware and software on site.

More than ten million end users at more than 31,000 organizations ranging from small businesses to the Fortune 500 use Symantec.cloud to secure and manage information stored on endpoints and delivered via email, Web, and instant messaging.

Symantec.cloud helps IT executives to protect information more completely, manage technology more effectively, and rapidly respond to the needs of their business.

For specific country offices and contact numbers, please visit our website.

Symantec.cloud North America
512 7th Ave.
6th Floor
New York, NY 10018 USA
1 (646) 519 8100
1 (866) 460 0000
www.MessageLabs.com

Symantec helps organizations secure and manage their information-driven world with **managed services**, **exchange spam filter**, **managed security services**, and **email antivirus**.

1/2011 21167355