Global Information Assurance Certification Paper

Designing a Layered Defense against Email-Borne Malware

By
Aaron Wayman

GSEC Practical version 1.4b
March 2004

Abstract

As network security has become more high-profile in the past few years, the focus of security professionals has been on building strong perimeter defenses, often through firewall implementations. However, as perimeter defenses have become more robust and harder to circumvent, attackers have adapted to the new landscape. Rather than trying to find or create obscure holes in a firewall, it is much simpler to use holes that are known to be open. This is the reason why a recent trend for propagation of malicious software has been through the use of Email, or SMTP traffic.

It is nearly impossible for a company to stay in business today without using email, and many people purchase home computers primarily to use email to keep in touch with friends and family. This means that if malicious content can be designed to successfully spread using email as a medium, it already has an open path to nearly every computer on the planet.

Although email-borne malware poses a significant threat to secure computing, this threat can be greatly diminished if proper measures are taken. Email-based threats are some of the best examples of why any security measures must be designed with an in-depth, layered approach. This paper will attempt to provide a roadmap to designing a layered solution that can greatly reduce the risks associated with the use of email. We will examine why simply allowing email traffic into a network is creating risk. We will look at five layers where we can defend against email-borne malware, and the components that must be included at each layer to provide an effective defense. We will look at the difference between the proactive and reactive defenses in these layers, and why some of the best defenses are often overlooked, because they are not technology-based at all. We will also look at the changing landscape of network security, and the reasons why the "Strong Perimeter Defense" model of thinking can be dangerous when designing security solutions. Although our focus is on designing defenses for an enterprise corporate or business network, many of the ideas presented here can be applied on a much smaller scale, even down to a single home user.

Why Email is Used for Malware Propagation

In the abstract, we touched upon two major reasons why it is currently popular for attackers to attempt to spread malware via email. One reason is that email provides an extremely simple way to penetrate to the core of a victim's network.

In the past, an attacker who was trying to get around or through a firewall would spend time figuring out which ports were open on the firewall, or try to figure out what type of firewall was in place and try to compromise the firewall in order to bypass it. These were network-level attacks, and they could be time-consuming and difficult to execute.

An alternative is to perform an application-level attack, which attempts to attack a network by finding an application that is *designed* to penetrate or bypass

perimeter network defenses, and then either try to deliver attacks using that application, or attack the application directly. Email is a wonderful attack tool, when you look at it from this perspective. A significant benefit of email for an attacker is the ability to attach any kind of file to an email message, and firewalls must be designed to pass these messages (with their potentially malicious payload) not only into the network, but all the way to end users. This is why network-based defenses such as firewalls or intrusion detection systems provide almost no proactive defenses against email threats.

Email is a popular method of spreading malware because of its widespread use. Millions of people and businesses use email every day, which gives an email-borne attack a statistically excellent chance of achieving its purpose. Unfortunately, the majority of people who use email have had little or no security training around this application, which makes it easier to fool them into launching an attack.

To understand the scope of what we are defending against, we need to define malware, and examine its purpose.

Defining Malware

Webopedia.com defines malware as "short for **mal**icious soft**ware**. Software designed to damage or disrupt a system, such as virus or a trojan horse." [WEBOPEDIA: MALWARE]. Malware is a broad category that can include several different kinds of programs, each with different purposes. The phrase "malware" came about because most examples of malicious software today don't fit into a single categorical definition, such as a virus, worm, trojan horse, or backdoor. Today's malware is often a combination of these types of threats. For example, the recent Nimda strain includes a routine to create an administrator account with a null password on Windows computers as part of its automated propagation routine. [ANTIVIRUS.COM: NIMDA] This exhibits behaviors of both a worm program (a program whose sole purpose is replication) and a backdoor program (a program that undermines security by providing an illegitimate way to access a system). Malware may be designed for a variety of purposes, including password or information harvesting, destruction of data or systems, security breaches, etc. Since the varieties and purposes of malware programs are limited only by the creativity of the people writing them, we must design comprehensive defenses that will work against a variety of attacks.

Lines of Defense

Concept: Defense in Depth

What exactly do we mean by "Defense in Depth?" A very simple definition is just that layers of defense are set up, so that if an attack compromises the first layer, there are more layers that an attack must penetrate to be successful. This is why the phrase "defense in depth" is often used interchangeably with "layered defense." However, this does **not** mean simply having more than one firewall, or

even using more than one kind of firewall software. To effectively design a layered defense in a network or computing environment, *different kinds* of defenses must be implemented at each different layer. As we implement each of the ideas suggested in this paper, we will see that there are ways that each layer can be bypassed. If every layer of defense is built upon the same concept or technology, then the ability to compromise the first layer creates a compromise in all successive layers.

Defense in depth is especially important in regard to email, because of the variety of malware threats that are present in this application. Since different types of malware have different targets and attack strategies, we must design defenses to prevent email-borne malware from: 1) reaching mail server(s), 2) reaching client workstations or end users, and 3) prevent the malware from executing. Then we must assume that all of these defenses will be compromised, and design systems to minimize the impact if malware does execute.

There are many attacks today that have known signatures, and we know specific ways to defend against these attacks. For the purpose of this paper, we will call these kinds of defenses *Proactive Defenses*. A proactive defense doesn't necessarily defend against one specific defense (for example, the Nimda virus), but can defend against a category of known attacks, such as sending .exe files via email. Specific types of these defenses will be explained in detail later. A *Reactive Defense* is a technology or mechanism that is not effective until an attack has already been successful. The purpose of a reactive defense is to minimize the impact of an attack after it has bypassed all of our proactive defenses.

Concept: Corporate Policy and Business Continuity

Before designing any network defense system, it is extremely important to review (or create!) corporate policies that define permission levels and acceptable use guidelines for the organization's IT staff, and for end users. Policy should clearly define when and how IT staff are permitted to review e-mail. Some of the questions that should be answered are:

- Should IT staff be allowed to review all email, or only email for certain users or departments?
- Should IT staff be able to open users' email accounts after email has been delivered to the server, or should access be restricted to only monitoring email as it comes into the network?
- If IT staff members have access to monitor and review email for malicious content, are they required to document (or even have) a reason for opening certain emails?

The policies and procedures regarding IT access to corporate email should be very clearly defined, and should also take into consideration that personal, confidential, or even illegal information might be discovered as email is reviewed. Email-specific incident response plans should be in place.

All users on a computer network must be informed that IT staff may have access to their email, and under what circumstances their email may be

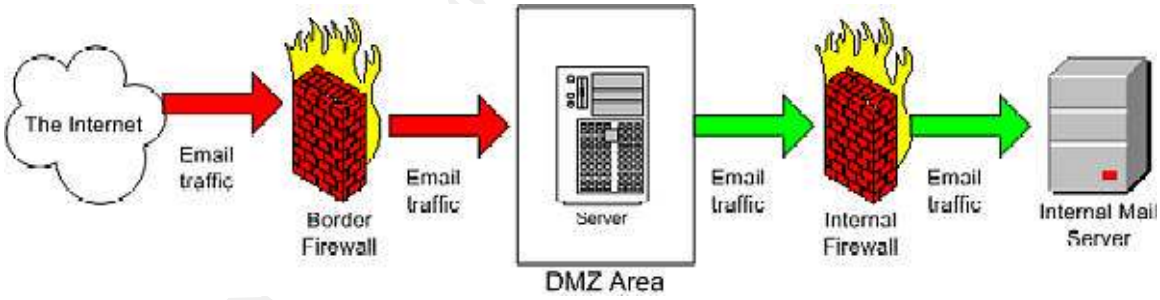
reviewed. Users should have a clear understanding of what level of privacy they can expect in regard to email (or any other electronic documents or communication). Also, users must be informed of an acceptable use policy. If it is not clearly stated that it is unacceptable for a user to intentionally email a virus from his home account to his work account, then legal recourse for that type of action may not be an option. An acceptable use policy regarding email may even state that file attachments may not be opened if the sender of the email is unknown.

All users on a network (including IT staff) should sign a written document that defines the acceptable use, privacy expectations, and authorized access levels for their job position. This will provide clear legal and practical boundaries for the types of defense measures and proper incident response plans that can be implemented.

Another important consideration when designing network defense systems is the idea of business continuity. It is very easy to implement automated defense systems that are too restrictive and will interfere with critical business processes. All defensive measures, especially those working with mission-critical communications such as email, must be analyzed and implemented in a way that creates effective defenses without disrupting business.

Defense Layer 1: Implement an Email Gateway

An email gateway is a logical single point of entry for all email to enter a network. We will discuss this topology as a single gateway/single mail server environment, but the logic can be applied to a load-balanced or clustered system with multiple front-end gateways and multiple mail servers.



Gateway Topology Diagram

The email gateway has two primary purposes. The first is to identify and eliminate attacks at the network entry point before they reach any internal parts of the network. This will be accomplished by using our second-layer server-side defenses, but the gateway's placement in the network topology makes it our first layer of defense against email-based threats.

The second purpose of the email gateway is to act as a decoy or sacrifice for attacks that are able to defeat our server-side software defenses. As shown in the diagram, the gateway should be in a traffic-restricted DMZ area of the network, so any type of attack that compromises the integrity of the box will not give an intruder access to the internal network. If any type of Denial of Service

(DoS) attack is successful against the target mail server, then only the gateway will fail, leaving the internal mail server operational. An email gateway server should be relatively simple to repair or replace in the event of an outage, whereas an internal server crash could be much more resource-intensive to repair.

The email gateway is based on the "perimeter defense" model of thinking, which is ineffective by itself (as we will see later). However, we should still use some of the methodologies and ideas presented by a perimeter defense as part of the design of our overall defensive system.

In the next section, we will examine the components that run on our email gateway and our internal mail server.

Defense Layer 2: Server-side Software

At each layer, there are certain components that must be included to make that layer a potential stopping point for an attack. The two main tools to implement at our second layer are antivirus programs and content filtering programs.

Antivirus Software

Antivirus software simply has to be on any networked computer to prevent downtime and data loss. High-quality antivirus software running on the email gateway and internal mail server will stop the majority of email-based attacks that an organization will see. However, problems with antivirus software arise when a very new or unknown attack is present, or if the antivirus signatures are not current. Antivirus software works by maintaining a database of specific signatures for known attacks and matching file content against those signatures to identify an attack. There are two scenarios that can make this defense ineffective. First, a rapidly-spreading attack can infect hundreds or thousands of systems before antivirus vendors are able to identify the attack, and then create and publish a software update that contains a new attack signature. Even after the new signature is published, an attack could still have a window of several hours (or even days) to spread before our antivirus software performs its update routine to include new virus signatures. Antivirus software is currently a reactive defense, because it must rely primarily upon knowing the signature of an existing attack in order to be effective.

The second problem with a signature-based defense system comes from polymorphic viruses. Symantec defines a polymorphic virus as, "a virus that can change its byte pattern when it replicates; thereby, avoiding detection by simple string-scanning techniques." [SYMANTEC] Today's sophisticated antivirus products have been designed with polymorphic viruses in mind, so they can often detect when a known virus has changed its signature in an attempt to avoid detection. Unfortunately, virus writers are aware of this, and will attempt to create a polymorphic virus that changes so radically that it will completely avoid detection from one variant to the next.

While antivirus software has its limitations, we can reduce the risk of new attacks bypassing our antivirus products by using more than one scanning engine. "If you have only one antivirus engine, you depend on that vendor to

continually update its virus definitions, and to be able to immediately identify virus-like or worm-like code when it arrives in an email." [ZDNET] A safer alternative is to use antivirus products from multiple vendors, or use a single product such as GFI's MailSecurity ®, which allows the use of multiple antivirus engines in a single email-aware antivirus product. Increasing the number of antivirus engines that scan incoming email will increase the chances of catching new attacks; the idea is that if one scanning engine is weak in a certain area, a second or third engine will be able to successfully identify and eliminate the threat. [GFI]

Content Filtering Software

Another tool we can implement on our email servers to protect against antivirus software failures is email content filtering. An effective content filtering product for email will have the ability to scan all components of each incoming email item and validate those components based on certain rules. There are certain criteria that will be common to nearly all email based attacks, so we can implement content filtering as a proactive defense against future attacks. For example, malware that is attached to (or embedded in) an email message must execute on an end user's system in order to infect the computer, or execute on the server for a server-targeted attack. Therefore, by removing (at the server level) any executable or scripting components that are embedded or included as file attachments (extensions such as .exe, .pif, .vbs, etc.) we eliminate the possibility of launching malware at the client. Most new email clients will block many of these file types; a list of file types that are blocked by default in Outlook 2003 is listed at http://www.microsoft.com/office/ork/2003/three/ch12/OutG07.htm This list provides a good reference of extensions that should be blocked at the server, in case email clients are used that do not provide adequate file blocking.

The ability to customize filtering settings in content filtering software can provide ad-hoc defenses against certain types of attacks. For example, HTML-encoded emails can contain links to malware that is stored on the Internet, without actually carrying a malicious payload in the email itself. Antivirus software will not be able to identify this type of attack, but content filtering software can look for customizable keywords or characteristics and block delivery, or simply remove all HTML encoding in incoming email.

As with antivirus software, there may be weaknesses in content filtering products. Perhaps a high volume of email will cause the software to "fail open" and quit scanning mail content. Many products will only scan a few layers into compressed files; perhaps simply zipping an executable attachment several times will bypass the restrictions that are in place.

Server-side defenses could still be viewed as perimeter defenses; we are attempting to clean or block traffic before it is passed to more internal parts of the network. Unfortunately, the weaknesses in signature based scanning and the potential for failure in content filtering software require us to implement more defenses, in case our perimeter security fails.

Defense Layer 3: Client-based Defenses

It is unlikely that a malicious email can produce a successful attack by simply passing through our network. In order for an attack to take place, some action must be taken, either by a person or a program, to open or view the email. Since the email client is where most email-based attacks will actually occur, we must be especially diligent in protecting this layer. Here are four steps we can take to harden our client systems against these attacks:

1. Disable execution of scripting and executable file attachments within e-mail client software. As we mentioned previously, newer email clients are designed to block execution of many of these items by default, but watch for options to adjust or work around items that are blocked. If a user can execute a program or script through their e-mail client, the probability of accidentally launching an attack is very high.
2. Patch client operating systems and applications. Viruses often take advantage of well-publicized vulnerabilities in either the client's operating system or an email-related application. For example, the Bugbear virus and it variants send HTML-encoded email that take advantage of a flaw in certain versions of Internet explorer to run their malicious code. This design is a result of Internet Explorer components being used in Microsoft's Outlook and Outlook Express email clients. [MICROSOFT: BUGBEAR, ANTIVIRUS.COM: BUGBEAR] If all current updates and patches are applied, these types of attacks will usually fail. Unfortunately, slow patch releases from software vendors can weaken this defense.
3. Limit user rights on client computers. Perhaps a threat will bypass all of our defenses to this point by taking advantage of a new or unpublished vulnerability. If the user who is logged on to the client computer doesn't have permission to install programs or write to system files, there is a good chance the attack will fail, even if the code is executed. Users should be unable to disable antivirus software, and when possible, be unable to change security settings within applications. Operating systems such as Windows 2000 offer very granular control when restricting user rights. However, there will always be some users who will need full administrative access to their computers.
4. Run *current* antivirus software on all client systems. The restrictions of a signature-based defense that we saw at the server level are still present on our clients, but antivirus on every client computer is crucial nonetheless. As we discussed previously, using a different scanning engine on clients than is used on servers will provide another level of protection.
5. Consider disabling HTML-encoded email. The ability to embed links to the internet in an e-mail opens a new window for email-based threats. Content filtering software and client may block delivery or execution of malicious attachments within an email, but clicking a link in an email may launch a separate web browser window with less restrictive

settings. If business practices allow it, disabling HTML in email will force a user to copy and paste any links, and hopefully bring attention to the true nature of any false links.

Even after all of this is done, it is very possible for threats to present themselves on client systems without ever encountering **any** of our first and second layer defenses, and can circumvent some of our third layer defenses. In the next section, we will see how this can happen.

Think we're secure? Not yet!

So, let's recap: We built a perimeter defense by establishing an email gateway and running antivirus and content filtering on the gateway and our internal mail server. Since we know new or unpublished attacks may bypass this perimeter, we hardened all the client software on computers attached to our network. These steps will identify and eliminate the majority of attacks that come through a corporate email system. The problem is, there are gaping holes in this defensive system, because it is extremely difficult to maintain a static perimeter that has any real integrity in today's corporate world. This section will present several ways that a perimeter defense can be defeated, not only in regards to email, but also in other areas of security.

1. Laptops & VPN connections are dangerous. Often, someone who is using a laptop or a home PC to connect to their business network, they will have full administrative rights on that computer. *Our* corporate email systems may be secure, but what about personal email accounts? What if unsafe email client software that doesn't perform proper attachment blocking has been installed? What if a spouse is using the same computer to connect to another corporate network that is not as well-protected as ours? These systems can be compromised while disconnected from our network and our defenses, then re-attach and infect us. We must be absolutely sure that all of our third layer, client-based defenses are used on **all** client computers that connect to our network.
2. POP mail opens a new avenue of infection. Network users may have personal e-mail accounts that they are able to check via Post Office Protocol (POP). If a user can configure his email client software to check a POP mail account at www.maliciousemail.com (for example), then he can download e-mail that will bypass our first and second layer defenses. If possible, we should consider blocking access to POP mail, or restricting access to only certain users.
3. Web mail can circumvent defense layers one, two, and three. Many personal email accounts are web-based, so restricting access to them is difficult, and filtering email content stored in these accounts is impossible. Also, by using a web browser instead of email client software, execution of scripting and dangerous file attachments becomes possible. Depending on the network design and business policies, it may be possible to restrict access to sites that provide web-

based email. The only other practical defense against this threat is in the all-important fifth layer.

4. Encryption can hide malware from defenses. If users are able to use Public Key Infrastructure (PKI) technologies to encrypt email, then our antivirus and content filtering products may be unable to examine the emails for malicious payloads. If business practices require that some (or all) email be encrypted, we must be aware of this, and be especially diligent in hardening client systems. There should be a clear policy regarding encrypting email, and everyone should know what that policy is.
5. Our end users are creative. We will create defenses for every avenue of attack that we can think of, and someone with a little creativity will think of one more. Perhaps someone thinks they need to open a certain executable file attachment, and change the settings that block this action. They may figure out a way to disable the client antivirus software in order to open a file that is being blocked. Even without malicious intent, user actions can be very detrimental to the health of our networks.
6. Dialup ability breaks the perimeter. If someone is able to create a dialup connection to the Internet, then they are able to check POP mail or web mail, regardless of restrictions we have placed on our network traffic. When possible, computers on our networks should not have the ability to dial out and connect to other networks that may not have the same level of defenses as ours.

Defense Layer 4: Network-based Defenses

A very important concept to remember when designing any type of network defense is what to do if, despite best efforts, there is a compromise. There are steps we can take to minimize the impact of, or recover from, a successful attack.

The first step to take at the network layer is to use a Network Intrusion Detection System (NIDS). A properly implemented NIDS can send an alert if a computer on the network is infected by an email-based attack, and ideally provide for removal of that client from the network before the attack gets out of control.

We should also establish a baseline for email traffic on the network. Since the malware we are looking at propagates via email, a widespread network infection can significantly increase the amount of email traffic on that network. Knowing the volume of email that a server normally sends will allow anyone monitoring this statistic to recognize a spike in emails leaving the network, indicating a possible infection. Since many viruses spread via their own mailing routine, the network should be monitored for attempts to send email outside the network from computers other than designated mail servers.

Defense Layer 5: Non-Technological and Preventive Methods

Too often, we focus all of our attention on implementing technology-based defenses against technology-based attacks. However, some of our best defenses are non-technological solutions.

Training

Training end users can be our most effective weapon in fighting email-based attacks. Since these are the people that open emails and attached files, we need to educate them on proper email security. Regular training sessions should be held that are focused on how to safely use email. Topics should include:

- Never open an attachment from someone you don't know
- Never attempt to install anything that claims to be a program or software update that is attached to an e-mail, regardless of the sender
- Be extremely cautious when clicking links in emails
- Understand social engineering as it applies to email. Senders can be spoofed, official-looking pages can be faked, etc.
- Understand email client software. Many people falsely believe that by using a "preview pane" (such as the one in Microsoft Outlook), they are not actually opening an email, and are therefore safe from any malicious content. People should understand *why* certain files are blocked in client email software, and why they should not attempt to circumvent these restrictions.
- General understanding of what threats can be present when using email, and how to avoid them.

IT staff should be trained to never distribute software updates or patches via email. Doing so would negate any training concepts that people have learned.

Another defense that IT staff can implement is simply researching new attacks and attack methodologies. For example, the number of file extensions that should be blocked in email applications has grown significantly in the past few years. Anyone in charge of email security needs to understand what attacks are possible and how to defend against them.

Testing

After implementing all of the defenses discussed thus far, there is one more tremendously important step in the plan around email security: testing. The Eicar test string is available to test antivirus software functionality; it is completely benign, but including it in an email should trigger an alert with any antivirus software running on an email gateway or server, or on client workstations. Test file attachment blocking by sending non-malicious forbidden file types via email. Check for the ability to trick content-filtering software by using double file extensions (.exe.doc) or adding characters to forbidden file extensions (.exe_). Try zipping a file that is normally blocked and mail it in. Does our defensive software scan inside compressed files? What if the file is compressed several times, or is compressed using different archive programs?

The most obvious purpose of this testing is to make sure that the defenses we have implemented are actually providing the security we expect. A second benefit of testing is that it forces the tester to examine his or her own defenses

from an attacker's perspective. Sometimes we will find that our defenses are not adequate against a certain type of attack, and it is much better for us to identify and patch any holes than for a new strain of malware to take advantage of them.

Review: Perimeter vs. Layered Defenses

The ideas presented here regarding the perimeter defense model apply to any area of network security, but especially to email systems. It is very difficult to set up a secure perimeter that defends against malicious email, because the application is *designed* to be delivered through the perimeter and into the core of our networks. Also, POP mail and web-based personal email services create avenues for malicious email to enter our network without encountering our "perimeter" defenses.

Despite its weaknesses, it is still a good idea to use a best-effort approach to creating a secure perimeter. A properly-configured email gateway and internal server with effective antivirus and content filtering will probably eliminate 95% or more of the email-based threats that enter a network. However, if there are no internal defenses, an attack that does manage to break the perimeter will have free reign to replicate and have the potential to destroy data and create excessive downtime.

Implementation of some of the layers suggested here will take some time, research, and money. Other layers are very simple to implement. The important thing to remember is that no single layer is going to provide a complete solution to email-based threats. However, by combining several layers into a comprehensive defensive system, we will be able to minimize the time we spend recovering from attacks, and spend more time developing new technologies that will lead to increasingly secure networks.

List of References

[ANTIVIRUS.COM: BUGBEAR]
URL:http://www.trendmicro.com/vinfo/virusencyclo/default5.asp?VName=PE_BUGBEAR.B

[ANTIVIRUS.COM: NIMDA]
URL:http://www.trendmicro.com/vinfo/virusencyclo/default5.asp?Vname=PE_NIMDA.A&VSect=T

[GFI]
URL: http://www.gfisoftware.com/mailsecurity/wpmultiplevirusengines.htm
"One Virus Engine Is Not Enough"

[MICROSOFT.COM: BUGBEAR]
URL:http://www.microsoft.com/technet/treeview/default.asp?url=/technet/security/bulletin/MS01-020.asp

Microsoft: Attachments blocked in Outlook 2k3
URL: http://www.microsoft.com/office/ork/2003/three/ch12/OutG07.htm

[SYMANTEC]
URL: http://securityresponse.symantec.com/avcenter/refa.html#Polymorphic_virus
[Glossary entry: Polymorphic Virus]

[WEBOPEDIA: MALWARE]
URL: http://www.webopedia.com/TERM/m/malware.html

[ZDNET]
URL: http://techupdate.zdnet.com/techupdate/stories/main/0,14179,2876822,00.html
Rash, Wayne. "One is not enough." August 7, 2002