Malicious Code and Spam in Wired and Wireless Environments: Problems and Solutions

Joel Sing and Ben Soh
Department of Computer Science and Computer Engineering
La Trobe University
Bundoora VIC 3083, Australia
Email: joel@ionix.com.au, ben@cs.latrobe.edu.au

Abstract

Over the last several years digital pests - namely viruses and spam - have reached epidemic proportions, severely impacting the usability of digital communication systems, primarily affecting email. These pests result in increased bandwidth usage, increased operating costs, potential security threats and above all decreased usability. They also have the potential to create Denial of Service attacks, crippling the networks that they infect or target. Whilst to date they have been mostly confined to email, it is becoming apparent that they have the potential to impact digital communications infrastructure of the future. For example, concerns have already been raised with regards to the potential impact of spam on Voice over IP. Spam and viruses have also been appearing within other technologies and communication media, including mobile phone networks and hand held devices. To combat the abovementioned security problems, we propose an email scanning gateway using open source tools, whilst ensuring that the gateway can be readily deployed in a commercial environment. In this paper we present details of the design and implementation, and discuss several possible applications of our proposed system.

1. Introduction

Digital communication systems have provided an extreme level of connectivity, especially when compared to their non-digital counterparts. In particular, email has provided a mechanism for people around the globe to communicate, effectively without cost. Unlike normal mail, email is delivered with minimal delay and unlike the phone system, it is not dependant on the recipient being currently available. These key elements have resulted in email becoming one of the most widely used communication systems, being the backbone of many businesses. However, the key aspects of digital communication systems, those that have made it as successful and as usable as it is, have also allowed for it to be heavily abused. Digital pests have used the same aspects to their advantage, much to the detriment of email users. Additionally, there is not a single entity that controls the Internet, therefore people who abuse the system cannot be easily located and disconnected, unlike a phone network.

There are two main forms of digital pests, namely spam and viruses, both of which occur in various digital communication systems including email, mobile phones, hand held devices, instant messaging and Voice over IP (VoIP). The requirements for the successful transmission and replication of viruses and worms is discussed in section 2. Spam is detailed in section 3, including the motivation for spammers, the primary properties of digital communication systems that allow spam to work and the impact it has on digital communications. A summary of currently known technologies for combating viruses and spam is presented in section 4. The design and implementation of an email scanning gateway that uses open source tools to deploy a select number of these technologies, is proposed in section 5. Application of this implementation to various digital communication networks is discussed in section 6. Finally, possible further research is discussed and a conclusion is presented[1].

2. Viruses and Worms

Viruses and worms require a method of invoking themselves on the device they are attempting to infect, typically exploiting a security hole that exists due to poor programming or tricking the user into invoking it. A form of transmission is also required in order for the virus or worm to infect new systems. Transmission methods available in-

[1] Due to space constraints, some content has been omitted. For a full copy of this paper please contact the authors via email.

clude floppy disks, bluetooth, email and digital networks such as the Internet. Viruses and worms currently infect computer systems, particularly those running a Microsoft Windows operating system, mobile phones and hand held devices such as PDAs. In the not too distant future we may see VoIP equipment being infected if security holes within phone firmware are found and exploited - an example of previously found security issues is detailed in [2].

The last few years have seen a huge increase in the number of viruses and worms that transfer themselves via email, either exploiting a security hole in the end user's mail client or tricking the user into opening and executing an attachment. The Melissa virus which appeared during 1999 is arguably one of the first viruses to transmit itself via email, with many since following suit. Email is a simple and effective means of transferring data between systems, in this case the data being the viral executable.

3. The Spam Problem

Over the last couple of years Unsolicited Commercial Email (UCE), more commonly known as spam, has become a major problem with email communication. Simply put, spam is email that you receive without ever having requested it, typically being of a commercial or advertising nature. Whilst to date it has been primarily limited to email, a number of people have raised concerns with regards to other forms of digital communication infrastructure, including Voice over IP (VoIP) [9], Instant Messaging (IM) [11] and the Short Message Service (SMS) [10] provided by many mobile phone networks.

The main problem with spam stems from people wanting to allow unknown and unauthorised people to contact them for legitimate purposes, meaning that anyone, anywhere, can contact and communicate with them. Whilst existing communications infrastructure such as the standard phone system and postal system could be used to send spam, two issues exist for the sender. Firstly, it is prohibitively expensive. In Australia it will cost around $0.25AUD to make a local phone call and $0.50AUD to send a standard letter. Generating either in massive proportions, as required for spam to be successful, would cost the originator a large sum of money. Secondly, neither of these forms of communications can be completely automated - a large amount of time has to be spent to generate and post the letters or make the phone calls.

Digital communication systems have a number of properties that are the opposites of their non-digital counterparts. Firstly, communication is cheap, with email and VoIP phone calls effectively being free of charge. Secondly, these services are semi-anonymous, allowing for messages to be generated that are hard to block and difficult to track. Thirdly, they can be easily automated, allowing for minimal human interaction.

Spam has a huge potential to destroy these precious digital communication mechanisms, as it severely reduces usability and causes numerous problems for the end user, including wasted time and effort in deleting unwanted messages. Additionally, for users who incur per megabyte data charges, the cost of receiving massive quantities of unwanted email adds up. The bandwidth usage is also of concern for users of mobile and ad-hoc networks, where bandwidth is in very short supply and/or extremely expensive. Mobile users connecting to the Internet via a digital mobile network, such as GPRS or CDMA 1X are examples of this.

4. Existing Technologies

A number of solutions exist to reduce the amount of spam received by email users. This section briefly covers a number of these technologies and identifies their associated strengths and weaknesses.

4.1. Filtering/Tagging

A large number of spam filters exist, most of which implement some form of pattern or word matching in an attempt to distinguish spam from non-spam (commonly known as ham). Many of the more successful spam filters implement Bayesian classification engines [5, 6, 13], calculating a weighting based on each word that appears within the email message. Once a message is identified as being spam a number of different approaches can be used. These include tagging the message, either by appending new headers to the RFC2822 [12] email message or by changing the subject line; storing the message in a separate account or folder, for later retrieval and deletion; or deleting the message.

Whilst such filtering can prove extremely useful, assisting the user in the task of separating the spam from the ham, either at the mail server or on the end user's system, a number of issues exist. There is still a potential for false negatives and worse, false positives. If the user is simply filtering messages off to one side and checking through them manually, the issues associated with this are greatly reduced. However, if messages are automatically deleted when they are considered to be spam, a single false positive will result in legitimate email being deleted, without any notification to the user.

4.2. Blocklists

Due to the fact that a large amount of spam originates from a small number of professional spammers, blocklists based on the sender IP address and envelope from address can be effective, if well maintained.

4.2.1. IP Based Blocklists

When a remote system connects to an SMTP server the SMTP session may be accepted or rejected, based on the remote system's IP address. For spammers that own a network block or have been allocated static addresses by their upstream ISP, IP based blocklists can be used effectively, rejecting email that originates from these systems. Spammers can however work around this by delivering email through other mail relays, particularly those running on systems that have been compromised. Regularly changing mail relays will bypass this mechanism, making it very difficult to implement effective IP based blocking. This has become even more of an issue of late, with spammers using viruses such as Sobig to deliver an SMTP engine to a large number of "zombie" systems.

4.2.2. Domain Blocklists

Like IP based blocklists, domain based blocklists can be used to reject email which appears to originate from a particular domain. Whilst this does not appear to be as effective as IP based blocklists, it can still be useful against spammers using throwaway domains and/or small companies who do their own mass-marketing via email. Most Mail Transport Agents (MTAs) allow the system administrator to specify a list of email addresses and domains from which email should be rejected, typically returning a hard SMTP error code in response to the *MAIL* command, if the envelope from address matches a domain or address on the blacklist.

4.3. Sender Policy Framework (SPF)

Sender Policy Framework (SPF) aims to prevent the spoofing of email by stipulating which IP addresses email can originate from for a given domain. Within the DNS records for a domain an additional TXT resource record is added, specifying IP addresses of allowed senders. Email which appears to be from this domain and originates from other IP addresses will be considered suspect and may not be delivered, depending on the MTA configuration.

A number of problems currently exist with SPF. Firstly, spammers can quite easily register throwaway domains and configure SPF records for each domain, specifying their own IP address blocks as legitimate senders. This allows for such a mechanism to be easily bypassed. Secondly, whilst in many circumstances it will be easy for a company to have fixed IP addresses where email can originate from for their domain, in other cases it can be extremely difficult. Consider a member of staff working from home, connected via their ISP. Email sent from this machine will not be considered legitimate unless the IP address belonging to the ISP is added to the SPF record. Alternatively, the user would need to relay email via the corporate mail server. Thirdly, at least one implementation of SPF is patent encumbered, namely Sender-ID, severely impacting the potential implementation and use of this protocol.

4.4. DomainKeys

DomainKeys [3] aims to provide sender authentication, preventing the spoofing of email. Like SPF it uses DNS TXT resource records to provide authentication data, however public key cryptography is used as opposed to the IP addresses of legitimate senders; the process being similar to the ideas presented in [8]. To use DomainKeys the owner of the domain generates a key pair and publishes the public key via DNS. When sending an email the message is canonicalised and the private key is used to sign the message, with the signature being added as a new RFC2822 message header. When a system receives this message the signature can be verified by retrieving the public key from DNS and checking that the signature is authentic.

Whilst DomainKeys will allow recipients to verify that the sender is allowed to send email for the given domain, spammers will be also able to publish public keys for throwaway domains that they have registered. This means that DomainKeys, like SPF, will only be useful for preventing spammers from spoofing email to make it look like it originated from domains they do not control. Additionally, DomainKeys currently lacks any form of granularity, meaning that there is only one private key per domain. This requires that all email be signed at the corporate gateway, otherwise a copy of the private key would need to exist on each workstation, severely impacting the security of the deployment. This problem is only extended with road warrior users whereby the private key may be leaked to a third party if a laptop is misplaced. The canonicalised form of the message that was originally signed is also a potential source of problems, particularly for messages that have been modified by intermediate MTAs. Unless the exact message canonicalisation can be achieved, the signature will not verify - even if the message is legitimate and signed correctly.

4.5. Greylisting

Greylisting [7] is a technology that ensures an SMTP server behaves in a manner that complies with RFC2821 before accepting email from it. When a connection is received from an unknown SMTP server, a greylisting implementation will return a soft SMTP error code and record the tuple consisting of the envelope from address, envelope to address and remote IP address. Providing that the delivery of this message is reattempted with the same tuple after a given period of time (typically 30 minutes), the remote IP address will be added to a whitelist of known SMTP servers. Once whitelisted, an SMTP server can successfully deliver

mail without delay.

Many spammers use custom mass-mailing applications that will either give up if a soft SMTP error code is received or retry repeatedly for a short period of time. Additionally, the envelope to and from addresses are often changed at random, as is the mail relay in use. As a result, most spammers will never end up on the whitelist and their mail will never be accepted.

5. Implementation

An email scanning gateway was built for use in a commercial network, using only freely available open source tools. This section details the design and implementation of this system, before discussing the results achieved from its deployment.

5.1. Design

The design goals of the system are to prevent the delivery of email borne viruses and to reduce the amount of spam received by users. Primarily three of the previously mentioned technologies were deployed, along with a number of other systems to increase the overall effectiveness (see figure 1). The Mail Transfer Agent (MTA) used was Qmail, a modular, robust, secure and flexible system that is easily modified to alter its behaviour. Qmail-Scanner, a Perl based content analyser, ensures that a message is compliant with RFC2822, before scanning the message for viruses using Clam Anti-virus, an open source anti-virus toolkit. If the message is found to be a virus, Qmail-Scanner will quarantine the message and notify the postmaster. If the message does not contain a virus SpamAssassin is used to determine the likelihood of the message being spam. Finally, Qmail-Scanner adds two additional RFC2822 headers to the message, detailing the rankings returned from SpamAssassin. These headers can be used for filtering within the Mail User Agent (MUA).

At the firewall, *spamd*, a spam deferral daemon included with the OpenBSD operating system, is used to provide SMTP greylisting (see figure 2). The Spamhaus SBL is also used in conjunction with *spamd*. SMTP connections from all blacklisted IP addresses are tarpitted, wasting the spammer's time and resources. Additionally, local IP and domain blacklists are maintained, rejecting the receipt of email from various unwanted sources. It is also worth noting that the entire system was designed and implemented in a manner that allows for easy replication.

5.2. Results

The results observed after deploying the above mentioned system are far from scientific. However, anecdotal

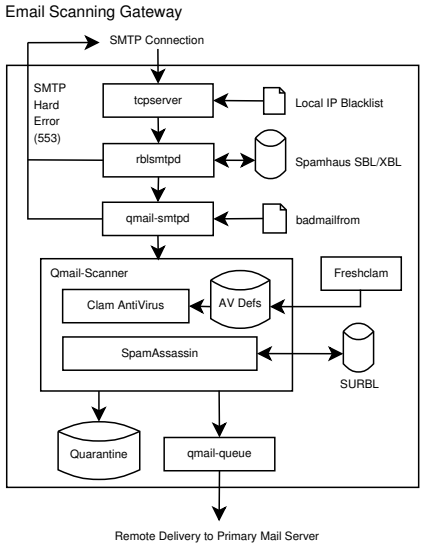

Figure 1. Email Scanning Gateway Schematic

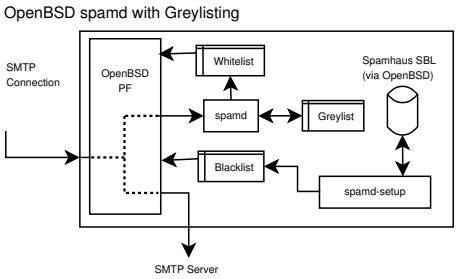

Figure 2. Operation of spamd

evidence suggests that the use of an IP based block list, such as the Spamhaus SBL, can be highly effective when combined with SMTP greylisting. A number of users were receiving in excess of 100 spam messages per day prior to the deployment of the filtering system. After the introduction of the email scanning gateway, without greylisting support, the same users were only receiving between 5-10 spam per day. Greylisting was implemented approximately a month later and resulted in the the same users receiving 1 or 2 spam per day, a significant decrease from the 100+ per day previously received. Additionally, in the first month of deployment around 2000 messages were quarantined due to containing viral executables. During the following month, after the addition of greylisting, only 447 viruses were quarantined. This seems to be a pleasant side-effect of greylisting, as it would appear that the internal SMTP engines used by many viruses do not adhere to retry intervals, hence delivery is not reattempted at a later point in time.

6. Applications

This section discusses the application of anti-virus and anti-spam technologies, such as the ones used to implement the previously detailed email scanning gateway, within various digital communication networks and infrastructure. In all cases the aim is to prevent viruses and spam from being received by our systems, preferably also preventing the transmission of digital pests to other network hosts.

6.1. Geostationary Satellite

If a geostationary satellite link is used as the last hop to a corporate network that hosts a local mail server, there may be benefits in deploying scanning and filtering systems on both sides of the link, effectively being located on each of the terrestrial gateways. This would allow for viruses and spam, both outbound and inbound, to be combated prior to reaching the geostationary satellite link, reducing the amount of traffic being routed via satellite. In this configuration, SMTP traffic that is destined for the corporate mail server would be scanned for viruses, blocked via IP address blacklists and/or greylisted, prior to the SMTP connection occurring over the satellite link. Likewise, outbound SMTP traffic would also be scanned prior to leaving the network, preventing the transmission of unwanted traffic over the link.

A number of gains may be experienced by deploying an email scanning gateway prior to the geostationary satellite link. Firstly, many carriers within Australia charge per megabyte for data, thus preventing the transmission of unwanted email will result in a cost reduction for the user. Secondly, the connections to the scanning gateway do not have to cross over the satellite link, which typically exhibit large Round Trip Times (RTT) and high Bit Error Rates (BER). Once again, this will reduce the amount of traffic on the satellite link, reducing costs and decreasing latency for other traffic traversing the network.

6.2. Wireless Networks

With wireless networks, a single mobile user may change between any number of base stations during the course of using the network. Anti-spam and anti-virus technologies could be deployed within wireless base stations, preventing unwanted email from leaving the network. A well configured firewall, along with the proxying of protocols such as HTTP [4] and SMTP, would allow for scanning of outbound email traffic. If SPF became ubiquitous, one option would be to implement the appropriate checks within the base station or local SMTP proxy. For example, when the envelope from address is specified via the *MAIL* command during the SMTP session, the SPF record can be retrieved for the domain and if the sender's IP address is not listed as a valid originating IP address for the domain, the SMTP session could be terminated with a hard error code. The same technique could be applied with DomainKeys, however this would not be able to prevent spammers from transmitting spam having a from address of a throwaway domain for which an appropriate public key has been published via DNS.

6.3. Voice over IP (VoIP)

Concerns have been raised with regards to the potential for spam to impact the usability of VoIP. Like email, VoIP increases the level of connectivity available by providing voice based communication that is effectively free of cost. Additionally, many VoIP systems provide voicemail. Due to the low level of cost involved, it is possible for a spammer to make hundreds or thousands of automated calls, delivering a pre-recorded message to VoIP users or to their voicemail service. In comparison, the standard phone system uses a fixed identifier that is costly and awkward to change. Calls are typically expensive and are difficult to automate.

Unlike email, VoIP is not a store and forward system, instead it is a real-time process that allows for normal conversation. As such, technologies such as greylisting become useless, as do filtering and tagging systems. At this stage caller identification appears to be the only weapon against abuse of VoIP. For example, you could configure your VoIP phone to only accept calls from specific people. This however, would largely defeat the purpose of this communication system, as people would not be able to contact you without prior arrangement. An alternative system would be one analogous to those used by many instant messaging protocols, whereby when a new contact attempts to communicate with you, an option to accept or reject the contact is provided. This however, will not prevent new spammers from annoying you at 4am with an incoming pre-recorded phone call. Callback or CAPTCHA [1, 14] based techniques may provide a solution, requiring human interaction to validate the VoIP call before the receiver's phone actually rings.

6.4. Mobile Networks

Many users of mobile phone networks receive SMS based spam from time-to-time, something which appears to be more of a problem in the United States where the recipient pays, as opposed to the sender pays model used in Australia. It may be possible to identify and block SMS based spam by deploying a filtering system within the mobile network SMS infrastructure. Unlike email, SMS spam is typically delivered to the mobile network in the same manner as standard SMS messages, making technologies like

greylisting impractical. However, content analysis could be highly effective, particularly if coupled with a system for users to report the receipt of SMS spam. This would allow for the creation of a message blocklist, whereby after receiving multiple reports of a spam SMS, a hash algorithm such as SHA1 could be used to identify identical messages and terminate delivery. As with email, this may be avoided by sending the same message with minor changes. Rate limiting would be another technique that may be effective, restricting the number of messages that can be sent from a single user within a given period of time.

7. Further Research

Whilst many of the technologies detailed within this paper are effective, it is becoming evident that the intelligence of the network may need to be improved in order to increase the level of control within the Internet; currently a loosely controlled system of interconnected networks. In particular it may be highly advantageous to push the blocking of traffic out into the network, preventing the need for traffic to be blocked at the last hop, prior to the corporate network. This would need to be controlled on a per user or per network basis, allowing for the current level of fine grained control afforded to firewall and network administrators. Such a system would allow for the blocking of SMTP traffic from particular hosts, for example those on the Spamhaus SBL, at their upstream ISP, reducing the amount of traffic traversing the Internet.

Another interesting approach would see the creation of an SMTP IP address whitelist, providing the IP addresses for SMTP servers that are known to be legitimate and well administered. One possible source of this information is the whitelists generated by the greylisting process, at the very minimum it would provide a list of IP addresses known to be running appropriately behaving SMTP servers. This could be used in conjunction with blacklisting and greylisting, preventing the need to delay delivery from all unknown servers.

VoIP would appear to have a very large potential for communication in the future, however it also has some of the biggest issues as far as abuse is concerned. Identifying ways to avoid the misuse and abuse of VoIP will be critical if it is going to be widely and effectively used. Without such mechanisms its usability will be severely impacted, possibly in ways worse than that seen to date with email.

8. Conclusion

Digital communication systems have resulted in increased communication around the globe and provide a highly cost effective means of communicating with anyone, anywhere. However, the same principles that have made it highly effective and usable have also allowed for it to be abused by spammers. Viruses have also taken their toll, using communication systems such as email to transmit themselves around the Internet.

Many of the existing technologies discussed in this paper can be implemented successfully, as outlined in the email scanning gateway used for a commercial deployment. It is worth noting that there does not appear to be a single solution to digital pests, rather a multi-layered approach is required in order to achieve maximum effectiveness. Spam appears to be an ongoing problem and numerous approaches will probably be required in order to keep it under control. Whilst we may not be able to eliminate spam in its entirety, a significant reduction is far better than none, increasing the usability of what are arguably the world's most important communication systems.

References

[1] The CAPTCHA project. http://www.captcha.net/.

[2] CERT/CC. CERT Advisory CA-2003-06 Multiple vulnerabilities in implementations of the Session Initiation Protocol (SIP). http://www.cert.org/advisories/CA-2003-06.html, February 2003.

[3] M. Delany. Domain-based email authentication using public-keys advertised in the DNS (DomainKeys). Internet Draft, Yahoo! Inc, 2004.

[4] R. Fielding, J. Gettys, J. Mogul, H. Frystyk, L. Masinter, P. Leach, and T. Berners-Lee. HyperText Transfer Protocol - HTTP/1.1. Standards Track, RFC 2616, Internet Engineering Task Force, 1999.

[5] P. Graham. A plan for spam. http://www.paulgraham.com/spam.html, August 2002.

[6] P. Graham. Better bayesian filtering. http://www.paulgraham.com/better.html, January 2003.

[7] E. Harris. The Next Step in the Spam Control War: Greylisting. http://projects.puremagic.com/greylisting/whitepaper.html, 2003.

[8] M. Jakobsson, J. Linn, and J. Algesheimer. How to protect against a militant spammer. *Cryptology*, 16, no. 2(2), March 2003.

[9] J. Leyden. Spam gets vocal with VoIP. http://www.theregister.co.uk/2005/02/17/spam_gets_vocal_with_voip/, February 2005.

[10] J. Leyden. Users choke on mobile spam. http://www.theregister.co.uk/2005/02/10/mobile_spam/, February 2005.

[11] L. D. Paulson. Spam hits instant messaging. *IEEE Computer*, 37, no. 4(4):18, April 2004.

[12] P. Resnick. Internet message format. RFC 2822, Internet Engineering Task Force, 2001.

[13] S. J. Vaughan-Nichols. Saving private e-mail. *IEEE Spectrum*, 40, no. 8(8):40–44, August 2003.

[14] L. von Ahn, M. Blum, and J. Langford. Telling humans and computers apart automatically. *Communications of the ACM*, 47, no. 2, February 2004.