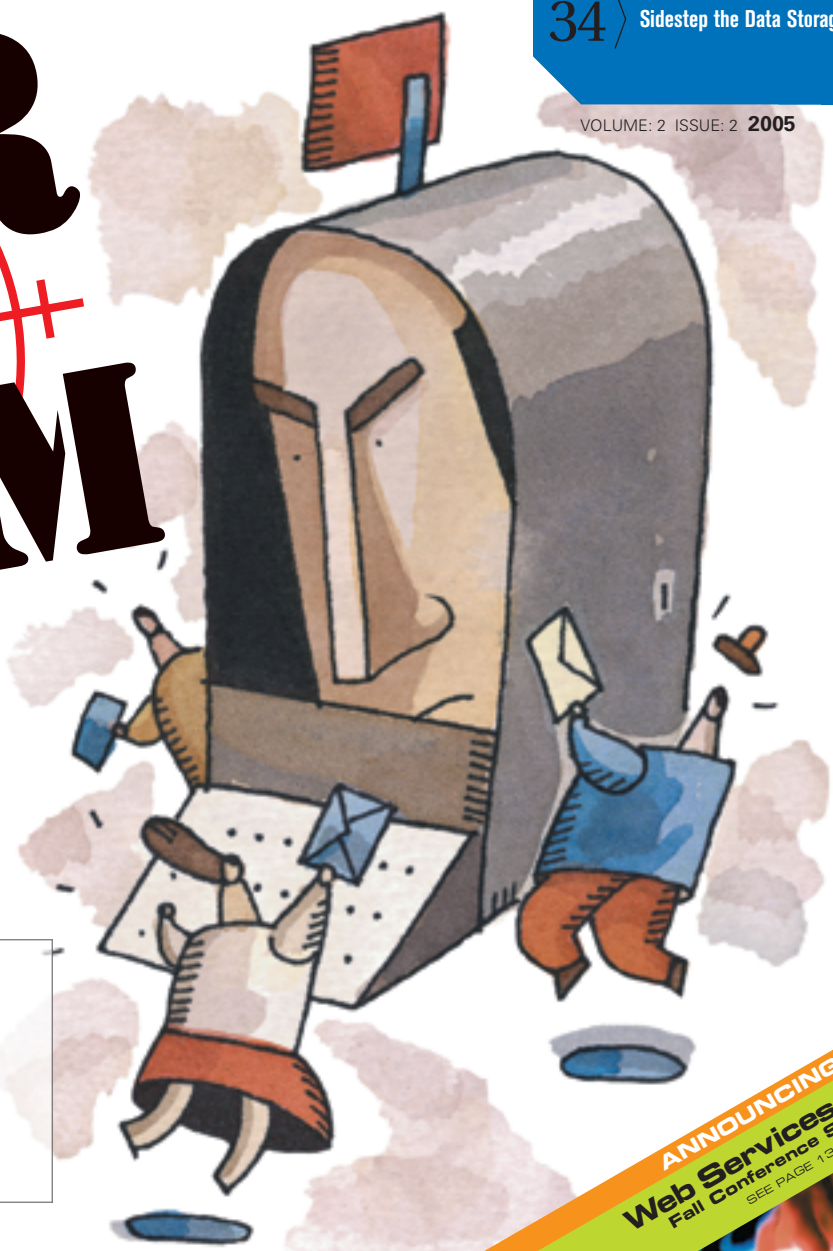
information
STORAGE+ SECURITY
journal
www.ISSJournal.com

In This Issue:

VOLUME: 2 ISSUE: 2 2005

THE WAR ON SPAM ⟨4

UPDATE — FROM THE — FRONT LINES



RECLAIM YOUR EMAIL

Visit us at NECC
Booth #358
June 27-30 Pittsburgh, PA

Spam and virus protection at an affordable price.

- No per user license fees
- Prices starting at $1399
- Powerful, enterprise-class solution

Barracuda Spam Firewall

BARRACUDA NETWORKS

Order a free evaluation unit at
www.barracudanetworks.com

POWERFUL EASY TO USE AFFORDABLE

Aggressive Reseller Program
Get more info by visiting **www.barracudanetworks.com/NECC**
or by calling **1-888-ANTI-SPAM or 408-342-5400**

information STORAGE+ SECURITY journal

Editorial Offices

SYS-CON Media,135 Chestnut Ridge Rd.
Montvale, NJ 07645
Telephone: 201 802-3000 Fax: 201 782-9638

Worldwide Newsstand Distribution
Curtis Circulation Company, New Milford, NJ

For List Rental Information:
Kevin Collopy: 845 731-2684
kevin.collopy@edithroman.com
Frank Cipolla: 845 731-3832
frank.cipolla@epostdirect.com

Newsstand Distribution Consultant
Brian J. Gregory/Gregory Associates/W.R.D.S.
732 607-9941, BJGAssociates@cs.com

SYS-CON MEDIA
The World's Leading i-Technology Publisher

From the Group Publisher

Storage + Security: Now a Major Technology "Sweet Spot"

BY JEREMY GEELAN

NOW THAT STORAGE professionals are obliged to know more about security and security professionals have to know more about storage, ***Information Storage + Security Journal*** is coming into its own just as we knew it would. All the articles in *ISSJ* are written by acknowledged experts in their field to get right to the heart of the storage-security problem matrix.

In a classic synergy of storage and security issues, for example, Winn Schwartau shows us how we can "Sidestep the Data Storage Blues" by ensuring that we sanitize our hard disks, rather than discard hard drives and tapes still containing valuable and private company data, as is still the case in a huge proportion of cases – amazingly enough, even in this age of GLB, SarbOx, HIPAA, and so on.

Symantec's L. D. Weller reminds us in "Best Practices for an Iron-Clad Backup and Recovery Plan" that the only way for businesses to ensure their data is adequately protected is to integrate security technology and policies with regular and effective backup of systems and important data. Weller warns that IT staffs "should take the time to implement and execute various security standards internally, incorporate disk and tape storage, partition hard drives, and plug other lurking holes in their system." If they don't change their priorities to make time, the outcome could be disaster.

Talking of changing priorities, in "Security's White Knight," SecureWave's Dennis Szerszen explains how "whitelisting" – setting a pre-defined list of applications and devices that can reside or function on corporate machines while blocking everything else by default – is replacing blacklisting. Whitelisting, his article shows, is simple, proactive, and lets administrators keep tabs on the scores of file types, devices, and applications that matter to a company's business – rather than the thousands that don't.

John Henze, director of marketing for what Cisco calls its "Caching Services" business unit, writes in detail about a financially effective way to manage an efficient storage infrastructure while ensuring user acceptance and satisfaction. Occasionally, Henze shows us, storage can be the enemy – in security terms – since a distributed storage infrastructure can often leave large amounts of data at risk, which undermines the effectiveness of the entire security infrastructure, so Cisco has worked on a way of letting IT managers give remote users high-speed access to the corporate data center, eliminating the risky local storage.

John Kelly, chair of the SNIA Storage Management Forum Requirements Committee, has written a fantastic breakdown of the benefits of ILM ("Six Steps to Building an ILM Foundation"), and Akbal Singh Karlcut writes about how organizations adopting Voice-, Video- and MultimediaOver-IP stand to reap huge benefits in productivity and cost savings, but are opening up holes in the network security fabric that can put the whole VoIP infrastructure at risk.

In our next issue, we will be bringing on board – as joint editors-in-chief – two noted business leaders. One is a CTO with many years' experience in addressing business challenges with blended IT solutions involving leading-edge database, Web, and hardware systems. The other has acted as chief architect, technologist, and director of several large-scale IT environments. Between them they are well known in the worlds of storage and security, and in their capable hands I haven't the slightest doubt that *Information Storage + Security Journal* will widen and deepen its coverage of what the industry agrees is a major technology "sweet spot," making it an even more compelling read each issue.

Have a productive and profitable 2005. And please continue letting us know your thoughts and suggestions as to which topics you'd like to see covered in *ISSJ*. Our e-mail, as ever, is issj@sys-con.com.

About the Author

Jeremy Geelan is group publisher of SYS-CON Media, and is responsible for the development of new titles and technology portals for the firm. He regularly represents SYS-CON at conferences and trade shows, speaking to technology audiences both in North America and overseas.

jeremy@sys-con.com

The War on Spam

UPDATE FROM THE FRONT LINES

BY YAKOV SHAFRANOVICH

THE INTERNET IS now indispensable to business at the cost of Internet abuse. Spam cascaded from an annoying trickle to a raging flood of ads, viruses, spyware, and phishing scams that pour into millions of inboxes everyday all over the world. With upwards of 80% of all e-mail traffic now spam, it's no wonder that organizations worldwide are looking for new ways to eradicate this blight. This article will discuss some of the newer developments in the "battle of the inbox."

E-Mail Authenication: SPF, Sender-ID, DomainKeys and IIM

Like traditional mail, e-mail is supposed to have a return address, often called a "bounce address," where undeliverable e-mail is returned. This address isn't always the same as the address of the author, which is called the "from address," just like in the real world where the return address on the envelope doesn't necessarily match the person who sent it.

Spammers fake these addresses since they don't want to pay for servers to handle all the undeliverable spam they send out, which can be in the millions of messages. Many times they make someone else's address the bounce address and the undeliverable e-mails are sent to some innocent bystander, a situation called a "joe job."

To add insult to injury, the bystander is usually blamed for sending the original spam. Spammers like to fake the "from" address, especially when phishing since it makes the message look more realistic to the unsuspecting.

In many ways, the e-mail system is modeled after the real world postal system. Unfortunately, just like real post offices, the return address isn't verified so spammers can fake the bounce and from addresses.

To combat the problem, several authentication schemes are being developed. The most popular ones are SPF or "Sender Permitted From," now renamed "Sender Policy Framework," (http://spf.pobox.com), Microsoft's Sender-ID (http://www.microsoft.com/senderid/) and Yahoo's DomainKeys (http://antispam.yahoo.com/domainkeys).

They all operate the same way: Domain name owners publish information about their domains in DNS and e-mail receivers check that information against the e-mail they get. For SPF and Sender-ID the information consists of a list of e-mail servers that are authorized to use the owner's domain in the bounce address (for SPF) or the author's address (for SID). For DomainKeys, domain owners digitally sign outgoing e-mail and publish the corresponding public keys in DNS.

The logic behind these techniques is that by having domain owners publish this information, it prevents anyone else from using their domain names without their permission. Receivers can check whether the information in incoming e-mail matches the data published by the domain owners, and can discard non-matching messages as fake.

Software support for e-mail authentication is sketchy. SPF enjoys the widest support, with implementations available

for most major MTAs (http://spf.pobox.com/downloads.html).

DomainKeys implementations are also available for most major e-mail software (http://domainkeys.sourceforge.net/) but unlike SPF, most DomainKeys implementations haven't been widely tested. Support for Sender-ID is very slim; even Microsoft won't support Sender-ID in Exchange until the second half of 2005 (http://blogs.msdn.com/exchange/archive/2004/12/22/330184.aspx).

There's no support for SID in any major MTA software except Sendmail (http://www.sendmail.net/sid-milter/) partly because of licensing issues with Microsoft patents and the collapse of IETF's standardization efforts.

However, the biggest sticking point is not necessarily lack of software support, but lack of participants. Few domain owners publish any records. Without the participation of domain owners, software support is basically useless.

Another issue is the lack of standardization. And there's no resolution in sight because IETF's MARID working group has disintegrated.

Nevertheless, SPF boasts a following of over 200,000 domains publishing SPF records (http://spftools.infinitepenguins.net/register.php).

While both Sender-ID and DomainKeys don't have many participants, they do have some big names such as AOL and Google publishing records and testing them. But compared to the number of domains and the volume of e-mail traffic on the Internet it's insignificant.

At the same time, significant technical and legal issues surround these schemes. SPF requires major changes in mailing lists and only protects the "bounce address," which is pretty thin. Questions have been raised about possible fatal flaws in Sender-ID's key algorithm. Sender-ID's license and patents have legal problems. Forwarding and mailing lists have to be reconfigured or changed. There's insufficient real-world testing and it's unclear these schemes can stand up outside the lab.

Nevertheless, the implementations that are dribbling out are letting organizations experiment. As field test data increases, it's hoped the flaws and technical problems can be resolved, though no technology is mature enough to be deployed in a production environment without pausing over the potential legal and technical issues. Since the protocols may be tweaked and changed many times before a final standard is set, managers must proceed with caution before spending valuable resources. While organizations may seek to protect themselves and rush to deploy these solutions, they must realize that e-mail authentication is only one step in a larger attack on spam.

Beyond Authentication: Reputation and Accreditation

Unfortunately a lot of media and analyst coverage has focused on e-mail authentication being the Holy Grail for killing spam. Nothing is farther from the truth since authentication only addresses a small subset of the problem. Actually almost no part of an e-mail message can be verified or authenticated. E-mail authentication proposes changes to the underlying e-mail architecture by verifying part of the e-mail message. Even if all e-mail were authenticated, it wouldn't end spam. This is where reputation and accreditation comes in.

In the United States one can't apply for a credit card or loan without a credit report. Sometimes when a person doesn't have a credit history, a third party has to vouch for his credit worthiness. Sometimes even that isn't sufficient, and the third-party guarantor has to promise to repay the loan in case of default. In essence, a person's credit history is his reputation or someone with a good reputation has to vouch for him in a process that can be called accreditation.

For either reputation or accreditation to work, there has to be a way to identify a person and verify his identity. In the U.S., social security numbers are unique identifiers. The same is true with e-mail – without authentication it's impossible to establish a unique identity or do either reputation or accreditation checking.

The current crop of e-mail authentication proposals provides a verified identity that can be used for reputation and accreditation. Different proposals provide different kinds of identities. SPF and Sender-ID merely provide a domain name. DomainKeys offer a cryptographic public key. But the end result is the same – a unique identity associated with an e-mail.

Because of the rising popularity of e-mail authentication, reputation and accreditation are being looked at. Various organizations and companies are discussing, developing, and testing different kinds of reputation and accreditation systems. The Insititute for Spam and Internet Policy's IADB system (http://www.isipp.com/iadb.php), Habeas's Accreditation program (http://www.habeas.com/senders/accredit.php), TRUSTE's Point of Collection program (http://www.truste.org/sealholders/point_of_collection_seal.php), IronPort's BondedSender (http://www.bondedsender.org/), and Cloudmark's Rating for Newsletters (http://rating.cloudmark.com/newsletter/) all provide accreditation services.

These programs vouch for a sender's identity and e-mail practices, and BondedSender provides a monetary bond in case the sender starts sending spam. Reputation systems such as CloudMark's Rating for Sender-ID (http://rating.cloudmark.com/senderid/) and Rele-mail (http://www.rele-mail.com/ seeks) provide reputation information about specific senders based on the collective opinion of many users. There's also talk from large SSL players such as Verisign about leveraging the accreditation information provided in SSL certificates for e-mail.

Some of these services are available now. Senders are usually required for pay a fee for accreditation and sign a written agreement with the accrediting organization, while reputation services don't require sender participation.

Access to accreditation and reputation databases is usually free to most receivers. Most of these services use a DNS-based protocol of information based on the protocol originally developed by MAPS for anti-spam blacklists (http://www.ietf.org/internet-drafts/draft-irtf-asrg-dnsbl-01.txt). The MAPS protocol is well known, making implementation and deployment easier.

However, these protocols vary significantly enough between accreditation and reputation services to require separate implementation for each provider. Even though technical information about how to access each system is available, support for them in e-mail software and spam filters is sketchy. The lack of participants, common standards and protocols and unproven track records are holding these services back from wider use.

While many players praise e-mail authentication, reputation and accreditation there's a dark side. Rudimentary e-mail reputation services have existed for a while

in the form of blacklists. Given that IP addresses are easily spoofed, blacklists use IP addresses as the identity against which reputation information is provided. Some newer blacklists such as SURBL (http://www.surbl.org/) use links to spam Web sites for reputation information. Blacklists such as SORBS (http://www.dnsbl.sorbs.net/), SpamHaus's SBL and XBL (http://www.spamhaus.org), and the infamous SPEWS (http://www.spews.org/) are operated by different organizations and individuals with inconsistent policies and appeals processes, all providing different kinds of reputation data.

While some efforts have tried to manage blacklists better (http://www.shaftek.org/drafts/draft-irtf-asrg-bcp-blacklists-00.txt), things haven't changed. There are still lots of cases of IP addresses and domains being incorrectly listed in blacklists with little or no chance of getting off. Sometimes a blacklisted IP space is reassigned to a new owner, who then has to spend time, money, and resources clearing up its reputation.

Given the mixed history of blacklists and how easy it is to set one up, why would "new" reputation services fare any better? And as with e-mail authentication, it's unclear how spammers would react to these solutions. Organizations choosing to test, deploy, and rely on these services should do considerable research to establish whether they're really effective.

"The good guys are **losing the war on spam**"

Attacking the Source: Greylisting, Tarpitting and Rate Limiting

The concept of delaying and stopping incoming spam traffic at the time of the e-mail transaction rather than after the e-mail has been accepted has been getting some attention lately.

One of the earliest implementations was called "tarpitting" (http://www.palomine.net/qmail/tarpit.html) and intentionally slowed down or delayed illegitimate connections to an e-mail server. Lately a related tool called "greylisting" has gained some prominence (http://projects.puremagic.com/greylisting/). Greylisting denies unknown senders e-mail connections altogether by issuing a "temporary failure" message in the SMTP transaction. It figures most spam sources don't behave the same as "normal" mail systems and usually don't try to resend e-mail if the connection fails. Legitimate systems, on the other hand, can try to resend e-mail for as long as a week.

A more ambitious method developed by the ePrivacy Group, which later spun off as a separate company called TurnTide that Symantec then acquired. Its idea was called "traffic shaping." A hardware device limits the bandwidth available to a specific server that looks suspicious. Unlike tarpitting, which slows connections down, TurnTide only limits the bandwidth on the network layer. It doesn't slow anything down. It figures legitimate servers won't be affected, but spammers sending huge amounts of e-mail will.

For system administrators, greylisting, tarpitting, and traffic shaping may be valuable tools. The cost of fighting spam is often the cost of the new servers needed to handle the inflated e-mail load. These traffic techniques attack spam at the network level before the mail servers accept the e-mail.

Since most major e-mail systems now support tarpitting and greylisting (http://greylisting.org/implementations/), including Microsoft Exchange (http://www.windowsitpro.com/Windows/Article/ArticleID/44772/44772.html), these tools are cheap alternatives to installing spam filters behind your mail servers.

Unlike e-mail authentication, reputation, and accreditation schemes, participation by senders or third parties isn't required; these tools operate on your mail servers. But they aren't foolproof and remain controversial.

Their effectiveness is temporary at best since they're based on how spammers operate now, which might not hold true in the future. Tarpitting relies on spammers sending e-mail to multiple recipients at a single domain over a short period of time. Greylisting assumes they won't resend e-mail from the same source. Both assumptions are true until spammers behave otherwise and if these approaches become more popular, they will.

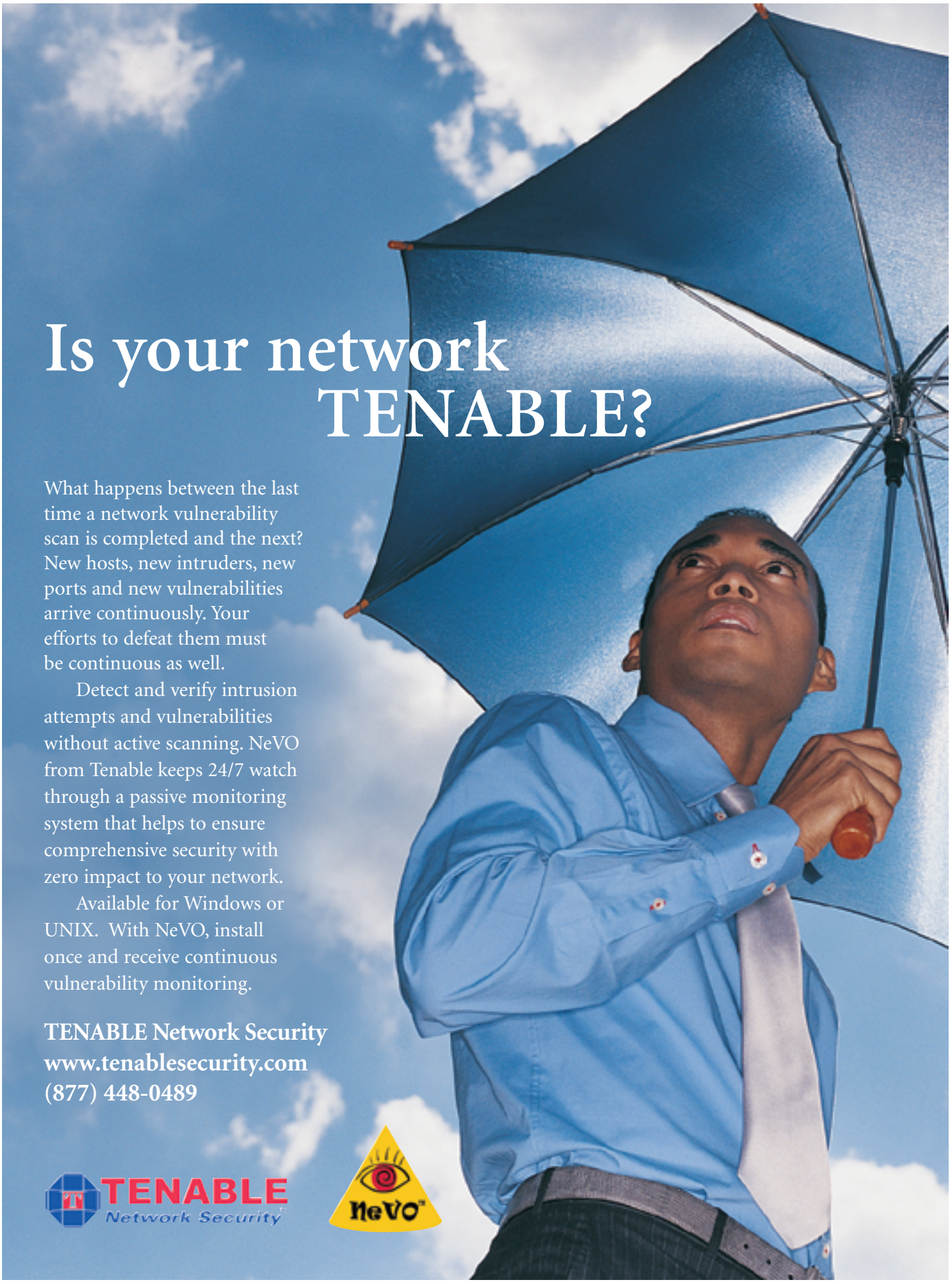
There's also the issue of collateral damage, which is true of most anti-spam technologies – legitimate systems can be taken for rogues. Some e-mail lists with lots of subscribers at the same domain can trigger tarpitting. Greylisting can incorrectly detect legitimate servers that don't try to resend e-mails, which means a whitelist has to be created (http://www.greylisting.org/whitelisting.shtml). Deploying inbound traffic control techniques on a legitimate mail transaction can tie up the resources of the incoming mail server and create all kinds of problems. Organizations deploying these schemes must tread carefully and test extensively.

Tarpitting and greylisting essentially limit the inbound rate of traffic since they control how much data comes in. A related concept is outbound rate limiting, which looks at the e-mail leaving an organization. This technique means setting up a monitoring program that clocks the e-mail traffic levels of all users that automatically takes action if those levels suddenly rise. The monitoring software can be configured to block or queue any further e-mail traffic until the issue is resolved. It's good at combating virus-controlled zombie PCs used to send spam. Some ISPs have begun testing and deploying it but software support is sketchy and it requires custom solutions. The notion of running spam filters on outgoing e-mail has been discussed but hasn't gained much traction. With limited budgets and low profit margins, ISPs tend to focus on the inbound traffic.

Follow the Money: E-Postage and HashCash

A recurring argument in anti-spam circles has centered on fees. Some hold that "If we charged for e-mail, spam wouldn't exist." This notion has given rise to e-postage schemes for e-mail such as "e-stamps" for e-mail, metering for ISPs, charging for the right to interrupt, charging only if the receiver thinks it's spam and monetary bonds.

E-postage would require a viable micro-payments system and there is none yet. There would also have to be a worldwide settlement scheme so ISPs could


Is your network
TENABLE?
What happens between the last time a network vulnerability scan is completed and the next? New hosts, new intruders, new ports and new vulnerabilities arrive continuously. Your efforts to defeat them must be continuous as well.
Detect and verify intrusion attempts and vulnerabilities without active scanning. NeVO from Tenable keeps 24/7 watch through a passive monitoring system that helps to ensure comprehensive security with zero impact to your network.
Available for Windows or UNIX. With NeVO, install once and receive continuous vulnerability monitoring.
TENABLE Network Security
www.tenablesecurity.com
(877) 448-0489
TENABLE
Network Security
NeVO

settle the charges incurred by their users among themselves. Since many ISPs don't communicate existing abuses, how can we expect them to trust each other and settle spam payments?

There is little data to suggest that e-postage would cut down on spam. Phones and faxes, which cost money, have their own form of spam. (Has a telemarketer called lately?) Spammers also hijack computers and could just tap into other people's e-postage accounts. E-postage advocates say that e-postage accounts could have preset limits that automatically stop any further spam once the limit was reached. However, if all that e-postage achieves is a preset limit, then outbound rate filtering would work just as well.

I see another problem with e-postage. It destroys the low cost of the Internet. E-mail, instant messaging, and VoIP are getting more popular than ordinary mail, printed catalogs, and phone calls because they're cheaper. Denied cheap e-mail, users will fall back on instant messaging and alternative e-mail systems so e-postage is unlikely to succeed.

Of course one could use CPU power for e-postage instead of money. Such systems, like hashcash, force the e-mail sender to solve a mathematical puzzle known to take a specific amount of CPU time. Supposedly legitimate senders wouldn't be hurt by this extra step since they don't send significant amounts of e-mail. Spammers, on the other hand, so the theory goes, would be forced to spend money on additional computers and CPUs to generate the answers, decreasing the incentive to spam. Of course, legitimate bulk mailers would be impacted too and spammers could probably afford to develop custom systems for computing hashcash stamps.

System administrators who would like to try e-postage should remember that no common protocol or standards exist for interoperability among systems, so they would have to fend for themselves. The too a significant percentage of the Internet would also need to use e-postage for it to be effective, but there are implementations are available for those who want to experiment (http://www.hashcash.org).

Other Kinds of Spam: Spim, Spit/Spyke and Comment Spam

Aside from e-mail, unsolicited ads are also finding their way into instant messaging, VOIP, and blogs.

Instant messaging has supplanting e-mail in some organizations. IM spam is called "spim." Obviously, this is a problem only when public networks are used such as AOL's AIM and ICQ, Yahoo's Messenger, Microsoft's MSN Messenger and the various Jabber networks.

Private networks aren't open to spim unless they're connected to a public IM network. Still the problem is becoming annoying enough to be noticed. Fortunately, public IM networks have a feature that e-mail doesn't have – a notion of identity. Unlike e-mail, the usernames used in public IM networks can't be spoofed easily, so people can block them using the free tools available in most IM clients and networks. On the other hand, virtually all public IM networks are free and don't verify identity in any way, allowing an unlimited number of new identities to be created easily, even cheaper than registering new domain names for e-mail spamming. Nevertheless, the availability of an IM identity combined with the fact that IM networks are centralized and more tightly controlled has been keeping spim to a minimum. It remains to be seen whether a decentralized network such as Jabber will fare differently. Commercial tools for fighting spim are available, but haven't been needed much.

Unlike spam and spim, VoIP spam, often called "spit," is currently a theoretical problem. Few verified cases have been seen but several commercial vendors are preparing tools to combat it. On VoIP networks connected to regular public networks such as Vonage and Packet8, spit hasn't been a problem because these VoIP providers basically extend the regular phone network, keep authenticated identity in place, and tightly control the thing. On pure VoIP networks such as Skype and FreeWorldDialup the likelihood of spit is higher and as pure VoIP networks get connected to regular phone networks, VoIP spam will grow and possibly spill over to regular phone users.

The comment spam in blogs is a new phenomenon that promises to become a major problem. Spammers post fake comments in blog posts that contain links promoting their sites, the same sites promoted by e-mail spam. Spammers want to get blog visitors to their site to raise the rank of their site in search engines, which rely on links as part of their ranking mechanisms (although recently search engines have begun to combat the problem http://www.google.com/googleblog/2005/01/preventing-comment-spam.html). Unlike spim and spit, comment spam has become a big enough issue for major blogs to start requiring that their readers register before posting a comment, and some have disabled comments altogether. In many ways comment spam mirrors e-mail and USENET spam in its lack of identity, reputation, or accreditation and decentralized control. Solutions borrow heavily from e-mail spam: blacklists, hashcash, filtering, and authentication.

Given the rising popularity of blogs as a way for companies to communicate with customers, the urgency of finding ways to prevent comment spam has risen. Most major blog programs include some kind of built-in comment spam countermeasure such as rate limiting and throttling. Plug-ins are also available such as MT-Blacklist, WP-Hashcash, and SixApart's TypeKey service.

Conclusion

As spam increases, the demand for solutions is swelling. New and innovative techniques such as e-mail authentication, outbound rate limiting, reputation, and accreditation services are being developed, implemented and tested widely. However, with e-mail the lifeblood of 21st century, organizations should be careful about deploying any new anti-spam "solution." ■

"Upwards of 80% of all e-mail traffic is now spam"

About the Author

Yakov Shafranovich is vice-president of a business software start-up and former co-chair of the Anti-Spam Research Group (ASRG).

comments@shaftek.org

INFITECH

X5 NAS

empower your data network

High Performance
Rack Mount Servers and Storage Solutions

> Simplify your network: X5 NAS will replace your file servers for Microsoft, UNIX and Apple clients. Manage a single network storage box vs. three legacy file servers. When more storage is required, simply plug another X5 NAS to an open network port.

> Remote, secured management: X5 NAS can be configured, maintained and monitored from anywhere in the world, as long as you have connection to the Internet. Use secured, HTTP(S) access for protection against unauthorized access.

> Faster access, more simultaneous clients: X5 NAS has proven to be faster and more responsive. Due to its optimized embedded OS, X5 NAS will outperform traditional file servers exponentially. Faster means more simultaneous users and getting jobs done quicker.

> Robust & highly available: Embedded OS, high quality hardware components, continuous on-going reliability test makes X5 NAS extremely reliable. Furthermore, its true server-to-server mirroring and real-time fail-over, makes X5 NAS the most highly available storage solution.

> Server to Server Fail-Over & Mirroring
> Snap Shot Data Recovery
> Embedded OS
> RAID 0,1,5,10, and JBOD
> SATA, PATA and SCSI HDD Support
> Hot Swap HDD and PSU
> SCSI/Fibre Channel Subsystem Support
> PDC/ADS/NIS/Host IP Blocking
> Dual Gigabit NIC with Fail-Over
> Up to 3TB in 3U
> 64bit, PCI-X for I/O

Powered by NetEngine

Visit Us www.infi-tech.com

or Call 1-800-560-6550

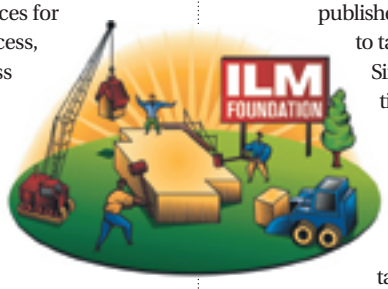
to Find Out More

Six Steps to Building an ILM Foundation

INVENTORY, CLASSIFY, ASSIGN, PROVISION, MONITOR, AND CHARGEBACK

BY JOHN KELLY

INFORMATION LIFECYCLE MANAGEMENT (ILM) is a complex, cross-functional, and interdepartmental strategy, a set of practices for managing the storing, access, and protection of business information in alignment with service-level and cost-of-ownership objectives. A successful ILM foundation enables IT organizations to align storage assets and costs with the applications that matter most to the business, while delivering a more cost-effective level of service to the applications and data that are not as critical. But trying to implement ILM without having a strong management foundation is risky, and can lead to mistakes, cost overruns, and project delays. ILM has grown in importance because of stringent new regulatory requirements in data retention, the increasing cost of retaining and managing data for extended periods of time, and the increasing cost of managing the complexity and capacity demands of unstructured data types (documents, presentations, media files).



The main goal of ILM is to ensure that strategic information and documents are stored on costlier storage with better response time, data retention, and recovery time characteristics. Conversely, ILM puts data that has been classified as less important (based on age, business relevance, etc.) on less expensive storage with weaker performance and availability qualities. By synchronizing the storage infrastructure with business requirements, enterprises can respond more cost-effectively to data reference patterns.

For example, a financial analyst's prospectus on a particular stock might have great immediate value to investors when it's first published, and so requires highly available storage. However, it's likely to be less valuable a year later when the market climate has changed and a new analysis has been published. At this point, archival to tape may be appropriate. Similarly, a sales presentation to a prospect has a high level of intrinsic value immediately before and after the presentation, but its value diminishes over time as new presentations are created and new prospects are targeted. Keeping the prospectus and sales presentation on high-performance storage where access and availability are optimized – even after the value of these documents has diminished – is a costly and potentially wasteful use of resources. Over the course of several weeks, these documents will be backed up at least once a week and sent off-site multiple times. Multiple copies of the same documents will have been e-mailed and copied to various servers in the same network. Lastly, recovering from a disaster may well mean recovering multiple copies of potentially obsolete information.

Collaboration applications such as e-mail are good candidates for ILM, since the intrinsic value of an e-mail drops dramatically after it's been read. At the same time, many industries have strict regulations for e-mail retention. With an ILM strategy in place, these messages could be put on lower-cost disk arrays or tape, archived in a warehouse, and/or eventually destroyed to provide the appropriate levels of service at more affordable price points.

Heterogeneous Storage Infrastructure: A Key Enabler of ILM

To implement an ILM strategy that offers the flexibility needed to satisfy varying service-level, cost, and data retention requirements, businesses must be free to create a heterogeneous storage infrastructure using a variety of vendors and/or technologies. Lock-in to a single storage vendor; using only high-end, mid-range, or low-end storage devices; or choosing only specific storage technologies – DAS, NAS, and SAN – or protocols – ATA, SCSI, Fibre Channel, iSCSI – are unlikely to satisfy all of an enterprise's present and future ILM needs.

Multi-Vendor SRM and SAN Management

To satisfy the core ILM requirements of storage resource flexibility, IT organizations must be free to select and implement the storage technologies best suited to their business and budgetary needs. The only way to cost-effectively manage a heterogeneous storage infrastructure optimized for ILM is to have a multi-vendor management platform in place that combines storage resource management (SRM), SAN management, and provisioning. With such a solution, IT managers will be free to mix and match storage technologies, without worrying about how to manage the overall capacity, performance, and health of the

ILM Benefits

IT managers planning an ILM initiative expect the following benefits:

- Reduce overall storage costs
- Align storage infrastructure with business requirements
- Optimize utilization of storage assets
- Reduce total cost of storage ownership (TCO)
- Improve service levels to business users and applications
- Improve the ability to comply with regulations and internal and external audits
- Defer unnecessary storage purchases

storage infrastructure. Of course, having a multi-vendor storage area management platform in place before you begin implementing your ILM initiative will help ensure the success of the overall project through better planning and preparation.

The Role of Industry Standards

The Common Information Model (CIM) standard and Storage Management Initiative Specification (SMI-S) can greatly simplify the management of ILM-optimized storage infrastructures. By selecting storage hardware that is compliant with CIM and SMI-S, or by choosing a management solution that monitors and manages storage infrastructure in accordance with these standards, businesses and IT organizations can both benefit from:

- ***Flexibility:*** IT can add new storage technologies in a plug-and-play fashion to meet changing business needs
- ***Scalability:*** IT can quickly increase capacity to meet ever-increasing data storage and data access requirements
- ***Manageability:*** IT efficiency and processes are maximized as a result of being able to manage the heterogeneous storage infrastructure through a single, consistent user interface
- ***Investment Protection:*** IT staff is trained on a single standards-based management platform, rather than a wide range of proprietary point tools
- ***Lower TCO:*** By leveraging the CIM object models native to Windows and some Unix systems, standards-built management solutions have lower memory and agent footprints.

Six-Step Methodology for Building an ILM Foundation

This section details a six-step methodology for building an ILM foundation that establishes a clear picture of where your storage infrastructure is today, and gives you a clear process to follow as you move forward into the future.

1. Take Inventory of Existing Storage Infrastructure

Before beginning an ILM implementation, it's critical to understand what is in your current storage resource inventory. By figuring out first what types of resources you have and how much capacity is being utilized by different business units and business applications, you'll be able to utilize existing assets for your ILM initiative better, avoid unnecessary investments in new raw capacity, and ensure that your new storage tiers are being properly set-up to satisfy service level requirements.

The following questions can help you gain a thorough understanding of what resources you have and how they are being utilized:

- ☑ How many and what kind of storage resources are there today, including DAS, NAS, and SAN?
- ☑ What are the total capacity and utilization levels?
- ☑ How are the resources logically and physically connected?
- ☑ What are the application, host and user dependencies on each resource?
- ☑ Which resources can be classified as Tier One, Tier Two, and Tier Three?
- ☑ Which resources have excess capacity that can be leveraged for ILM?
- ☑ What solutions are already in place for data protection and business continuity, such as types of RAID, mirroring, remote copies, and multi-path?

2. Classify Data Types and Map Data to Classification Model

"All data is not created equal." This is one of the fundamental premises of ILM, and requires that you understand the value of your data to determine where it should be stored, how it should be protected, and how much storage capacity will be needed for each ILM tier. While end users are best equipped to make value judgments about their data, surveying your entire end-user population to classify the value of every file is not practical. Fortunately, when it comes to classifying unstructured data, file-level attributes can be used to streamline this process. For example, every business application provides its own unique file extension – .jpg, .ora, .dat, .doc, .mpg, etc. – that can assist you in making judgments about data criticality and value. In addition, an analysis of when files were last modified and accessed can help you understand how important they really are.

Below are some questions you should answer to identify and classify your organization's data types based on its business value:

- ☑ What kinds of data are in the enterprise?
- ☑ How much of each type of data is there?
- ☑ What applications and users created this data?
- ☑ When was the data last accessed and last modified?
- ☑ What is the criticality and value of the data?
- ☑ What are the existing storage dependencies?

3. Assign Storage Resources to ILM Tiers

The third step in implementing ILM is understanding and defining service-level requirements for data access, recovery, and retention so that storage resources can be assigned to appropriate ILM tiers and data management disciplines can be incorporated that align business requirements with storage infrastructure.

Below are some guidelines for developing ILM service level agreements:

- ☑ What SLAs are in place today?
- ☑ Have SLA requirements changed?
- ☑ What are the performance characteristics of your applications and storage systems?
- ☑ What are the data retention and availability requirements of key applications?
- ☑ How fast must different applications be recovered in the event of a disaster or unplanned outage?

Once service-level objectives are understood and you have a thorough understanding of how data is being used, you're armed with the knowledge to assign classified data for differentiated treatment. The number of tiers defined will depend on the different classes of resources available in your environment as well as your ILM objectives (achieving compliance, reducing cost of ownership, controlling growth, etc.).

To define ILM tiers that work for you, answer the following questions first:

- ☑ Is the information business-critical? What is the business cost of not be able to recover this information?
- ☑ Does the information have data-retention requirements for regulatory compliance? If so, what is the time period required for retention and recovery?
- ☑ What are the expected access rates? Will this information be accessed:
  - ○ Once?

Common ILM Tier Definitions

Tier 1: Mission-Critical Data
Tier 2: Business-Critical Data
Tier 3: Business Operations Data
Tier 4: Historical Data

- ○ Several times initially then probably not again?
- ○ Frequently?

☑ What action should be taken once the data-retention requirements have expired?

- ○ Delete
- ○ Archive
- ○ Permanently destroy all records?
- ○ Alert the owner prior to action

☑ Does the information have disaster-recovery requirements such as multiple off-site copies?

☑ Is the information confidential and require secure access permissions?

☑ What access rates are acceptable?

☑ Is access time critical or is a slight delay when retrieving from near-line storage acceptable?

☑ If access rates are critical, does this information require high-performance disk arrays or will standard IDE- and ATA-based arrays suffice?

4. Provision Storage in Accordance with ILM Tiers

Once ILM tiers have been defined and populated with storage resources, there needs to be a fast, efficient, error-free process for provisioning additional storage capacity in accordance with your tier classifications. Common scenarios are that a new mission-critical application is about to be brought online or has exceeded the 80% utilization threshold, and a request for new storage has reached your IT organization. You need to be able to respond quickly by provisioning Tier 1 storage capacity, i.e., storage that is highly available with remote mirroring and multiple paths between the application and the storage.

Following are some guidelines for ensuring that you can quickly, easily, and cost-effectively allocate storage capacity in accordance with your ILM tiers:

☑ How will business applications get provisioned with the right tiered storage?

☑ How can a standard provisioning process be developed that supports the different kinds of storage products and different vendors required by the ILM strategy?

☑ How will provisioning-related SLAs be met and measured?

☑ How can provisioning be made easier so that more of the junior IT staff can offload provisioning operations from the handful of senior experts?

☑ How can provisioning jobs be scheduled during off-hours so that business operations aren't impacted?

5. Monitor ILM Infrastructure and Create Automated Policies

The fifth step in implementing ILM is to monitor your ILM-optimized storage infrastructure to ensure availability and create storage policies that automate IT's response to storage events and conditions. Proactively monitoring your storage infrastructure for events and potential outages and then adjusting your policies accordingly will help to optimize your environment and reach your ILM objectives. For example, policies can be created that send a notification when Tier 1 storage resources reach 80% utilization, create audit logs of SAN changes and provisioning requests, or move files that haven't been accessed in six months to Tier 3 storage.

When creating ILM policies, be careful that migration policies aren't too aggressive. Aggressive policies waste valuable time sending requests to the ILM policy engine and then retrieving data from near-line storage. At the same time, creating policies that are too conservative will result in inefficiencies and the purchase of unnecessary additional storage capacity. Some questions to consider when setting up policies include:

☑ How will events from the multi-vendor storage infrastructure be collected and correlated?

☑ How will capacity levels for each ILM tier be monitored and maintained?

☑ How will changes to the storage infrastructure be centrally controlled and audited?

☑ How will data be automatically migrated to the appropriate ILM storage tiers?

6. Chargeback to Business Units and Communicate Business Results

Now that you've implemented your ILM foundation, you'll want to make sure that business users and departments are held accountable for the tiered storage they're using. A storage chargeback solution is an excellent way to accomplish this. Storage chargeback centrally meters storage resources assigned to different business units for financial analysis, reporting, budgeting, and controlling storage-related capital expenditures. By assigning costs to various storage tiers and charging users for both the quantity and quality of storage utilized, you'll be reminding business users that what they're paying for storage capacity is directly proportional to the level of performance and protection they're getting. Chargebacks also serve the higher-level objectives of aligning storage infrastructure with business needs, and delivering storage in a utility model.

Before you implement a chargeback model, consider the following:

☑ How will the quality and quantity of tiered storage utilized by different business departments be tracked?

☑ How will the tiered storage provisioned in Step 3 be automatically incorporated into the chargeback model?

☑ How can new storage resources be incorporated into chargeback calculations?

☑ How will new business units be factored into the chargeback model?

How well business value is communicated often determines whether an IT project is perceived to be a success or failure. As a result, you should consider how you're going to measure and communicate the results of your ILM initiative to key C-level executives and business unit constituents.

☑ What kind of data should be in the reports to indicate the success of the ILM initiative?

☑ Which key individuals should get the reports?

☑ How often should reports be circulated?

☑ How should the reports be formatted to ensure they're read?

☑ How can the reporting process be automated to save time?

Summary

A successful ILM foundation enables IT managers to align storage assets and costs with the applications that matter most to the business, while delivering a more cost-effective level of service to the applications and data that are not as critical. Storage resource management (SRM) and SAN management solutions provide the essential building blocks required to implement ILM, and help ensure the success of ILM initiatives.

About the Author

John Kelly is director of product marketing at AppIQ, and chair of the SNIA Storage Management Forum Requirements Committee. John has been active in the storage management market for over 13 years as an industry analyst and as director of product marketing with HighGround Systems and Sun Microsystems. John has BS degrees in marketing and finance from Babson College and an MBA from Northeastern University.

john.kelly@appiq.com

www.sys-con.com/edge2005

web services EDGE conference&expo

FALL SERIES

[Engage and Explore]

the technologies, solutions and applications that are driving today's **Web services** initiatives and strategies...

CALL FOR PAPERS NOW OPEN!*

Coming to a City Near You ►

Web Services Edge
Fall Conference Series

3 Dynamic Conference Programs
Targeting Major Industry Markets

20+ seminars within 5 tracks will address the hottest topics & issues:

- ► Web Services: The Benefits and Challenges
- ► Web Services Security
- ► SOA (Service-Oriented Architecture) and ESB (Enterprise Service Bus) Strategies
- ► Interoperability, Incremental Integration, & Open Source
- ► The Management Process in Developing a Web Services Strategy

Program Features:

- ► Keynotes
- ► Tutorials
- ► Panel Discussions

Why Attend:

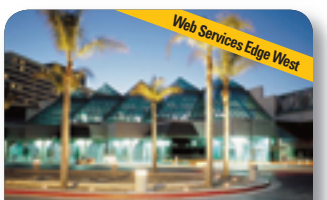
- ► Improve the return on your technology investment
- ► Develop & sharpen your strategy and identify key action steps
- ► Find new ways to reach and impress customers with Web services
- ► Maximize the power of your enterprise
- ► Protect your business from security threats
- ► Assess Web services as a viable option

Attention Exhibitors:

- ► An Exhibit-Forum will display leading Web services products, services, and solutions

The Westin Washington, D.C.
Washington, D.C.
September 7-8, 2005

1



The Westin Santa Clara Convention Center
Santa Clara, CA
October 24-25, 2005

2

Hyatt O'Hare Airport
Chicago, IL
Oct 31-Nov 1, 2005

3

Register Today! www.SYS-CON.com/Edge2005

Sponsored by

WebServices JOURNAL · XML JOURNAL · .NET JOURNAL · eclipse developer's journal · WebSphere JOURNAL · information STORAGE+SECURITY journal · wldj · JDJ · LINUXWORLD MAGAZINE · MX developer's journal · asp.netPRO · SD Times · CoDe · Software Test & Performance

*Call for Papers email: grisha@sys-con.com

For Exhibit and Sponsorship Information ► Call 201 802-3066

Produced by SYS-CON EVENTS

The Deep Inspection Firewall as VoIP Enabler

HOW SAFE ARE YOU?

BY AKBAL SINGH KARLCUT

COMPANIES IMPLEMENTING VOICE-OVER-IP (VoIP) technologies to cut communications costs shouldn't overlook the security risks associated with a converged voice and data network. Tempted by the thought of lower phone bills, centralized management and rapid deployment, VoIP security and network integrity are often neglected. There are numerous weak points to consider in a VoIP network – the call servers and their operating systems, the phones and their software, even phone calls themselves are vulnerable.

This article examines the issues and complexities of deploying a secure Voice- and Video-over-IP network, and how a VoIP-capable firewall can address these concerns.

Evolution of the Firewall in a VoIP Network

The traditional role of a firewall in a VoIP network is undergoing a radical evolution.

In the past, its primary job was simply to "behave well." VoIP relies on the predictable static availability of IP-based resources across the Internet, while the firewall's strong desire to keep ports closed as well as its network address translation (NAT) functionality inherently breaks the VoIP network. Through pinholing and other techniques, security vendors have found ways to interoperate with VoIP infrastructures.

With network-based threats getting ever more sophisticated, however, the firewall has evolved from behaving nicely to enabling and protecting the complete infrastructure.

From end-user devices such as IP-based phones, soft-phones and wireless communications devices to infrastructure equipment such as H.323 gatekeepers and SIP proxy servers, there's a lot of exposure in an organization-wide VoIP deployment. From simple denial of service (DDoS) attacks aimed at limiting the availability of the IP-based voice infrastructure to full-blown application-layer attacks targeting the VoIP protocols themselves, the threats are very real...and growing.

Elements of a Secure VoIP Infrastructure

For any successful VoIP implementation, three key factors have to be considered: VoIP security, VoIP network interoperability/protocol support, and VoIP vendor interoperability.

The big security factors that have be considered in any deployment are access, availability and implementation.

Access

VoIP calls are vulnerable to session hijacking and so-called man-in-the-middle attacks. Without proper safeguards, an attacker can intercept a VoIP call and modify its parameters/addresses. This opens up the call to spoofing, identity theft, call redirection, and other attacks.

Even without modifying VoIP packets, attackers can eavesdrop on conversations carried over a VoIP network. If VoIP packets are traveling unprotected over the Internet, attackers can access the information they carry.

With a standard public switched telephone network (PSTN) connection, intercepting conversations requires physical access to phone lines or access to the private branch exchange (PBX). Voice/data networks, on the other hand, which typically use the public Internet and the TCP/IP protocol stack, don't provide the physical wire security of phone lines. By gaining access and monitoring network traffic at certain points on a network infrastructure (such as to/from a VoIP gateway), an attacker can capture and reassemble VoIP packets. Publicly available tools such as Vomit (http://vomit.xtdnet.nl/) can convert these packets into a .wav file so an attacker can eavesdrop or even record and replay conversations.

Availability

The availability of a VoIP network is also a big concern. PSTN availability has reached 99.999% – attackers need physical access to telephone exchanges or have to cut the phone lines to have any impact. A simple DDoS attack aimed at key points of an unprotected VoIP network can disrupt, or worse cripple, voice and data communications.

VoIP networks are especially susceptible to DDoS attacks such as:

- ***The Malformed Request DDoS:*** Carefully crafted protocol requests can exploit a known vulnerability resulting in partial or complete loss of service. Attackers can not only crash the target but gain control over it.
- ***DDoS on media:*** VoIP media is carried in Real-Time Protocol (RTP) packets, and is vulnerable to any attack that congests the network or slows the ability of an end device (a phone or gateway) to process the packets in real-time. An attacker who has access to the part of the network where media is present simply needs to inject a large number of media packets or high

Quality of Service (QoS) packets to contend with legitimate media packets.

> ***Load-based DDoS:*** A DDoS attack doesn't necessarily need to use malformed packets to achieve its goal. Flooding a target with legitimate requests can easily overwhelm a poorly designed system. Even without an actual VoIP request, a DDoS attack such as TCP SYN Flood can prevent a device from accepting calls for long periods of time.

Implementation Problems

VoIP encompasses a large number of standards – such as the Session Initiation Protocol (SIP), H.323, the Media Gateway Control Protocol (MGCP) and H.248. These are complex standards that leave the door open to bugs in the software implementation. With PSTN, phones are just dumb terminals – all the logic and intelligence resides in the PBX. There's not a lot an attacker can do to disrupt access to a PSTN network.

With VoIP, the same bugs and exploits that hamper every operating system and application available today can also hit VoIP equipment. Remember, many of today's VoIP call servers and gateway devices are built on vulnerable Windows and Linux operating systems. One only has to look at the CERT advisories that have been issued for H.323 (CERT-H.323) or SIP (CERT-SIP) to see the number of vulnerabilities that have been found and the dozens of vendors affected by them.

VoIP Network Interoperability and Protocol Support

The second critical element of a secure VoIP infrastructure is network interoperability and protocol support. For a firewall, VoIP is more complicated than a standard TCP/UDP-based application. Because of the complexities of VoIP signaling and protocols, as well as the inconsistencies introduced when a firewall modifies source addresses and source port information with NAT, it's difficult for VoIP to traverse a firewall.

Here are a few of the reasons why:

> VoIP operates using two sets of protocols – signaling (between the client and VoIP Server) and media (between the clients). The port/IP address pairs used by the media protocols (RTP/RTCP) for each session are negotiated dynamically by the signaling protocols. Firewalls need to track and maintain this information dynamically, opening selected ports for the sessions securely and closing them at the appropriate time.

> Multiple media ports are dynamically negotiated through the signaling session; negotiations of the media ports are contained in the payload of the signaling protocols (IP address and port information). Firewalls need to deep-inspect each packet to get the information and dynamically maintain the sessions. That demands extra firewall processing.

> Source and destination IP addresses are embedded in the VoIP signaling packets. A firewall supporting NAT translates IP addresses and ports at the IP header level for packets. Worse still, fully symmetric NAT-firewalls frequently adjust their NAT bindings, and can arbitrarily close the pinholes that let inbound packets pass into the network they protect, eliminating the service provider's ability to send inbound calls to the customer.

> To support VoIP effectively, a NAT firewall has to do deep packet inspection, transform embedded IP addresses and port information as the packets traverse the firewall.

> Firewalls need to process the signaling protocol suites that consist of the different message formats used by different VoIP systems. Just because two vendors use the same protocol suite doesn't mean they interoperate.

VoIP Vendor Interoperability

The last element in a secure VoIP infrastructure is ensuring that the firewall will interoperate with all of the VoIP devices used in the infrastructure. Some VoIP vendors have slightly different implementations of the standard VoIP protocols based on RFCs not all of which are compatible. Furthermore, some vendors implement so-called standard-compatible proprietary VoIP protocols. Because of this, it's important for the firewall to interoperate with as wide a range of VoIP end devices as possible. A partial list of devices includes IP phones, videophones, videoconferencing gear, SIP proxies and H.323 gatekeepers. It's largely up to the security appliance vendors to ensure they interoperate with VoIP infrastructure devices.

Current VoIP Security Alternatives

As VoIP adoption grows, the number of VoIP security options is increasing. A few of these alternatives are described below.

"The traditional role of a firewall in a VoIP network **is undergoing a radical evolution"**

Simple NAT Traversal Solutions

Devices (such as IETF-STUN) that bypass the firewall to translate the IP addresses embedded in VoIP signaling packets. The only real advantage in these devices is that they co-exist with existing firewalls. There are many disadvantages of this approach such as:

> No/limited network security on "open" ports – VoIP devices are still exposed

> Won't work through symmetric NAT [IETF-TURN]

> Work only for UDP – won't support H.323 or SIP over TCP, and RTCP may not work since the port number it uses is tightly coupled to the one used for RTP

Session Border Controllers (SBCs)

SBCs were popular before firewalls (and other networking devices) had the "deep packet inspection" ability to deal with VoIP signaling and security. The SBC sits on the Internet side of a firewall and tries to control the border of a VoIP

network by terminating and re-originating all VoIP media and signaling traffic. In essence, SBCs are proxies for VoIP traffic. While SBCs deal effectively with the NAT translation issue, there are many drawbacks of this approach:

> SBCs add an additional hop in the VoIP session network path, resulting in unavoidable additional latency in peer communications that can affect the quality of a session.
> SBC devices can't protect the VoIP infrastructure at the application layer, or even against DDoS attacks targeting the infrastructure.
> SBCs may require that client software be installed on each network, introducing a bottleneck and jitter since VoIP endpoints have to direct all traffic (signaling and media) through it.
> SBCs QoS controls tend to be limited. For a successful VoIP implementation, both inbound and outbound bandwidth management is needed to provide high levels of traffic control.
> No encryption. If an organization wants to connect remote sites through a VPN tunnel, it requires separate hardware.
> SBCs are primarily designed for service providers. They're expensive and feature-overkill for most business-class applications.
> SBCs produce additional administrative overhead – another "box" to manage.

The VoIP-Enabled Firewall

Firewalls with VoIP capabilities are currently one of the most attractive VoIP security solutions. Most networks already have a firewall protecting the LAN as well as connecting remote sites and users through secure VPN technology. Because of their omnipresence, firewalls are a natural place to add facilities for VoIP security.

However, there are reasons more firewalls aren't VoIP-compliant. First, the firewall must understand the VoIP protocols it wants to protect. This requires elements of or even complete protocol stacks such as SIP, Megaco, and H.323. Second, the firewall must be able to read and operate on packet content at every layer in the stack. Then the firewall needs to track each VoIP call from the first signaling packet requesting a call setup to the point where the call ends.

It also needs to monitor for traffic legitimacy and for the syntax validation of all VoIP signaling packets. This puts incredible pressure on a device that was designed just to inspect TCP and IP packet headers.

A few firewall vendors have risen to the occasion with varying degrees of VoIP proficiency.

One advantage that VoIP-enabled firewalls have over SBCs and other VoIP security solutions is that they provide intrinsic protection against both Layer 3 and 4 attacks (such as DDoS and man-in-the-middle attacks), which is a part of the original job function of a firewall. A few firewall devices even have application-layer awareness and protect the VoIP protocols themselves. A smaller group of vendors provides virus scanning, intrusion prevention and other security services on VoIP traffic.

A major potential drawback of the VoIP-enabled, deep packet inspection firewall is that if the packet transform algorithms aren't implemented efficiently, or if the device doesn't have enough power to deal with the rigors of real-time application-layer processing, the VoIP traffic can suffer latency, reducing call or video quality. Vendors should be evaluated thoroughly in real-world environments before a selection is made.

Summary

While organizations adopting voice-, video- and general multimedia-over-IP stand to reap huge benefits in terms of productivity and cost savings, they are opening up new holes in the network security fabric that can put the whole VoIP infrastructure at risk. While solutions are available to help alleviate these problems, the VoIP-enabled firewall is gaining popularity among IT managers because of its effectiveness, simplicity, and low cost when compared to solutions such as SBCs. Care should be taken when selecting a solution for VoIP security, however, as no one solution does everything...no matter what the data sheet says. ■

"Organizations adopting voice-, video- and multimedia-over-IP stand to reap huge benefits in productivity and cost savings, **they are opening up holes in the network security fabric that can put the whole VoIP infrastructure at risk"**

About the Author

Akbal Singh Karlcut is a principal software engineer at SonicWALL responsible for a number of areas, such as VoIP and other stateful protocols. He has worked on several papers for the IEEE as well as patents in data communications. Akbal got his BSc in computer science from Kingston University in the UK and his MSc in computer engineering from San Jose State University in California.

21st Annual Gartner PlanetStorage Summit 2005

The Changing Face of **Recovery**

June 13 – 16, 2005 • Orlando, Florida • Hyatt Regency Grand Cypress

To register, or for additional information go to gartner.com/us/storage or call 1 800 778 1997.

Huge Changes Are Racing Through the Storage Industry. Are You Keeping Up?

Outside the walls of the cool, climate controlled world of your data center, huge changes swiftly transforming the storage technology landscape.

Data recovery is accelerating, regulatory requirements are tightening, and storage solutions are rapidly evolving to enable greater business agility.

You've got important tactical decisions to make today. And new technology budgets to plan for tomorrow. So go straight to Gartner PlanetStorage for actionable advice on your best storage moves for 2005.

Keynote Guests

John Thompson
Chairman of the Board and CEO, Symantec Corporation

Kenneth Thibodeau
Director, Electronic Archives Records Program, National Archives & Records Administration (NARA)

Conference Chairs

John Monroe
Research Vice President, Gartner

Carolyn DiCenzo
Research Vice President, Gartner

Robert Passmore
Research Vice President, Gartner

Priority code: STRISSJ

Gartner PlanetStorage Summit

Threat Assessment and Its Input to Risk Assessment

RISK ASSESSMENT AS A BUSINESS PROCESS

BY ANDY JONES AND DEBI ASHENDEN

About this Article

This article is an excerpt from *Risk Management for Computer Security: Protecting Your Network & Information Assets.* Printed with permission from Butterworth-Heinemann, a division of Elsevier. Copyright 2005. For more information about this book and other similar titles, please visit www.books.elsevier.com.

IN THIS EXCERPT we examine the role of threat assessment and its importance in the accurate and effective assessment of risk.

Threat

It seems appropriate to start this chapter by explaining what is meant by a threat assessment. In information security, this is probably one of the most abused and misunderstood terms and is often used interchangeably with the term "vulnerability." In this book, the word "threat" is used to describe those "things" that may pose a danger to the information systems, and for clarity, the term "threat agents" is used. What we are actually referring to is those agents, either intentional or accidental, that have the opportunity and that may exploit a vulnerability in the security of information systems.

The Internet Request For Comments (RFC) Glossary of terms describes threat in the following ways to cover differing environments:

> ***Internet usage:*** A potential for violation of security, which exists when there is a circumstance, capability, action, or event that could breach security and cause harm. That is, a threat is a possible danger that might exploit a vulnerability. A threat can be either "intentional" (i.e., intelligent; e.g., an individual cracker or a criminal) or "accidental" (e.g., the possibility of a computer malfunctioning, or the possibility of an "act of God" such as an earthquake, a fire, or a tornado).

In some contexts, such as the following, the term is used narrowly to refer only to intelligent threats:

> ***U.S. government usage:*** The technical and operational capability of a hostile entity to detect, exploit, or subvert friendly information systems and the demonstrated, presumed, or inferred intent of that entity to conduct such activity.

British Standard (BS) 7799, which has been developed into International Standard (ISO/IEC) 17799:2000 – Code of Practice for Information Security Management, is one of the most relevant documents and standards in this area and defines threats, risks, vulnerabilities, and assets as follows:

> ***Threats*** are anything that could cause harm to your assets, and vulnerabilities are weaknesses in your security arrangements that make it easy for these threats to occur. For example, if you have no backup of your data you are vulnerable and make the threat "loss of data" likely to occur.
> ***Risks*** describe the probability that a damaging incident is happening (when a threat occurs because of a vulnerability), as well as the possible damage if this incident takes place
> ***Assets*** are something that has value to your company and how it is carrying out its business operations.

The BS 7799 definition of information security also defines those aspects that it is safeguarding, as follows:

> ***Confidentiality of information:*** Ensuring that it is accessible only to those authorized to have access.
> ***Integrity of information:*** Safeguarding its accuracy and completeness.
> ***Availability of information:*** Ensuring that authorized users have access to it when required.

In developing a common vocabulary of terms, it is important that we recognize other standard definitions such as the ISO/IEC Guide 73 Vocabulary for Risk Management – Guidelines for Use in Standards. In this document, risk is defined as "the combination of the probability of an event and its consequence." Risk assessment is defined as "overall process of risk analysis and risk evaluation."

Threat Assessment

A threat assessment is an integral and essential element of the risk assessment and risk management processes. If an organization wants to undertake an effective risk assessment for its information systems to enable rational and considered decisions to be taken, then it is essential that an accurate picture of the threats to the organization are understood. It must be clearly understood that risk assessment is a business process. The need to carry out these assessments of the risks to information assets or to other assets of an organization has been brought about as a result of the proliferation in the use of information and communications technologies and the convergence of these technologies over the last three decades. This massive increase in the use of these systems and the subsequent dependence on them has resulted in significant changes in the level and type of threat to the information environment that we have, whether knowingly or in ignorance, come to rely on.

The way in which we assess the threat that is posed to an information environment has not developed at a pace that has

matched the rate of change and adoption of the technologies, with the result that we are still using tools and techniques from a previous environment. It is also a reality that the way in which we assess threat has not yet transitioned from art to science. As a result of using tools and techniques that were developed for non-technology-based systems, there is currently no way in which the threats, as opposed to the vulnerabilities, to information systems can be either modeled or quantified in any meaningful or repeatable manner that will allow the decision makers to take informed decisions.

In this heavily dependent and rapidly changing environment, where technology is offering new opportunities and the matching problems, all types of organizations, from governments to commerce to academia, are increasingly needing to produce meaningful risk assessments on which they can make decisions on the appropriate level of investment required to establish and ensure that they maintain the appropriate levels of confidentiality, integrity, and availability to their information. This is not possible without assessing threats as well as vulnerabilities.

Security standards such as BS 7799 and the related ISO 17799 start from the assumption that organizations and governments understand the threats they face to their information systems. For them to achieve a better quantification of the risk to an information environment, it is increasingly important that the information on which decisions are based is as up-to-date and accurate as possible and is expressed in terms that have a common meaning and basis. If a term is used in the assessment of threats to one information system, it should be understandable to those involved in the preparation of a threat assessment for another system, not least because interdependence between systems is a fact of life in the networked world. Unfortunately, in the high-technology environment reality is far from this. Even common terms such as threat and vulnerability are used almost interchangeably. If the input on the level of threat that is used in the risk assessment process is to be improved, then an accurate representation of the threat to information systems must be achieved.

The threat agent is not the only factor that must be considered when determining the level of threat to an information system. Other issues that must be addressed include the probability of an attacker carrying out a successful attack and the impact that a successful attack would have on the business. After all, no matter how capable and motivated a potential threat agent might be, if the countermeasures in place at the target are already at a level higher than the attacker can overcome, then there is no prospect of success. Also, if the information asset that is being targeted is of little or no significance to the business, then the potential impact to the business is low or nonexistent. (It may be appropriate in the last case to question why the system is being used if it has no value or impact to the business; the very existence of a system is a cost to the organization in terms of hardware, software, management, and maintenance.)

For the owners, the custodians, or the insurers of information systems to understand the risks that come into effect as a result of using a particular high-technology device in a particular set of circumstances, it is necessary to carry out a risk assessment of the relevant information environment. The assessment of the risk is essential in the modern environment because it will provide guidance on the system with regard to the likelihood of an event occurring after all the identified threats and vulnerabilities have been taken into account and the selected countermeasures have been implemented. From this, the relevant parties will have a better understanding of the residual risk they will be accepting if they choose to operate the system in the manner that has been defined and can explore the relative benefits of options to reduce this residual risk even further that are available to them and the relative costs and benefits of those options. The following factors must be considered when conducting such a risk assessment:

- The Agent that is causing a Threat to the system.
- The exploitable Vulnerability within the system (Note: The significant word here is exploitable; if it cannot be exploited, then it does not require investment to protect it).
- The Impact of a successful attack.
- Mitigating factors (countermeasures).

Some History

The assessment of threats in the political and physical environments has been undertaken since time immemorial at the national and international governmental level. More recently, large organizations in the commercial sector have also started to undertake threat assessments to meet legal and regulatory requirements and to ensure that the protection they implement is cost effective. At the government level, assessments have been undertaken by experienced and skilled analysts who have carried them out over an extended period of time. The assessments produced by these analysts have then been applied to potential threats to the nation states' physical assets. The analysts who have, historically, carried out the analysis of the threat have worked in an environment where the time scales were relatively long and the assets they were analyzing had a physical basis. Even in this environment, we have seen how difficult it is to produce an effective analysis of something like the threat from a nation state. A recent example is Iraq, where despite U.N. weapons inspection teams trying to detect weapons of mass destruction over a number of years, there was still considerable disagreement between a number of nation states on the capabilities of the country. With the benefit of good old 20/20 hindsight and a unique scrutiny from the United States, the United Kingdom, and other countries of the strength and accuracy of the intelligence that was used by the coalition governments to justify their actions, the picture became even more confused. Given that this is an assessment of the threat in which physical assets were being analyzed, you may begin to understand the problems that exist when we move into the new and more complex arena of information and information technologies.

Typically, threat analysts have looked at the threat that is posed by other nation states and terrorist groups. Every country will look at the threat that is posed to its interests, both at home and abroad, and will have skilled analysts who spend their careers specializing in, in all probability, a small section of the threat spectrum. They may concentrate on the threat from one geographical area or country and, over time, gain an in-depth knowledge of the threat capability of that entity. They will isolate key indicators of intent and capability, such as the movement of ships or aircraft, the movement of troops, or perhaps even the movement of key individuals. They will look for indicators of intent in the diplomatic arena (remember

that countries normally use the military as a last resort). The point is that where physical action is contemplated, the time scales are normally protracted, with a period of diplomatic activity that is then followed by a period of preparation, where the logistics and armaments are moved to locations that will enable the country to undertake operations. In this period, there is time for the analysis of likely actions and outcomes. If the group of interest is a terrorist group, although there may be no diplomatic phase, the group will still need to acquire the knowledge and equipment needed to carry out the attack and to deploy its resources to the location where they are going to carry out the attack. Although the same is fundamentally true of an attack on an information system, the level of resources required, the preparation time, and the number of observable indicators are all significantly different. In an attack on an information system, the attacker is likely to have available the resources to carry out the attack, and they are not detectable as other types of weapons would be. The preparation time is shortened because there is no requirement to move resources to a location from where they can reach the target, and the threshold for initiating the attack may be at a far lower level.

The threat posed to such an intangible and volatile environment as an information system has never, to date, been successfully assessed. That is, it has not been carried out in a provable and replicable manner. In the past, the threat agents considered have been, primarily, either other nation states or terrorist organizations. An example of how difficult the problem is was highlighted in February 1998, when the U.S. Department of Defense computer systems came under what was, at the time, described as a systematic attack. The attack pattern was highly indicative of preparations for a coordinated attack on U.S. Defense Information Infrastructure at a time when the U.S. Air Force was being readied for a deployment against the Iraqi Regime. The attacks all appeared to be targeted against Department of Defense network domain name servers and were exploiting a well-known vulnerability in the Solaris Operating System. (Incidentally, the patch for the vulnerability had been available for quite some time.) The attack profile consisted of a probe to determine whether the vulnerability existed in the server, which was then followed by the exploitation of the vulnerability to enter the computer. Once into the system, the attacker would insert a program that gathered data and, at a later date, return to retrieve the data collected.

The attacks were widespread and appeared to be well coordinated, and a large number of the attacks followed the same profile. The attacks seemed to target key elements of the defense networks, and over the period the attackers collected a large number of network passwords. The attacks could not be characterized or attributed to a specific source, but there was an obvious potential connection with the deployment for impending operations in Gulf.

After a considerable period of investigation (reported as up to 17 days), involving a wide range of the resources available to the U.S. Government, it was discovered that the attackers were what became known as the Cloverdale Kids, two youths aged 15 and 16 from Cloverdale, California, who operated under the nicknames of Too Short and Makaveli. They had also been given assistance and "mentoring" by a third person, identified as an 18-year-old Israeli youth, Ehud Tenenbaum, who used the nickname of Analyzer.

The reason for giving this example is that an attack, which was considered to have been initiated by a foreign nation and which involved a wide range of the resources available to the United States, was eventually attributed to three youths with the resources and equipment that can be found in the average home. This gives some insight into the level of difficulty we currently face. It is also worth pointing out that if this incident had occurred in almost any other country in the world, the time taken to isolate the perpetrator would probably have been considerably greater. If the attacker had indeed been a foreign power, how much damage could they have caused?

The threat posed both by, and to, commercial organizations or non-terrorist non-government organizations has not, in the past, been considered. In the current environment, however, it has been demonstrated that the potential impact from, and to, these other groups could be much more significant than was previously thought. The threats posed to non-government organizations and the elements of the Critical National Infrastructure have not been considered until recently, because in the past there was no single individual or group that was concerned with or had a sufficient understanding of the problem.

What Is a Threat Agent?

To understand what threat is, it is necessary to identify the separate elements that make up a threat. The elements identified below are not an exhaustive set but have been selected to demonstrate a good cross-section. The characteristics of these elements are as follows:

1. ***Natural threats and accidents:*** This group consists of non-intentional threat agents and includes those natural incidents such as earthquakes, typhoons, naturally occurring fires and floods, and the unintentional actions of humans. They are described separately as natural and accidental.
2. ***Malicious threats:*** This group consists of those threat agents that result from the intentional actions of individuals and groups and have the following characteristics that affect them:
   > Capability
   > Motivation
   > Catalysts
   > Access
   > Inhibitors
   > Amplifiers

Natural and Accidental Threats

These are two relatively well-known and understood groups of threats, and some knowledge of them can be gained from the insurance industry and the actuarial history they retain regarding the effects of earthquakes, fires, wind, water, and lightning. For the second group, accidental damage, there is, again, a wealth of information available within the insurance industry with regard to the likelihood of an accident occurring in the physical domain (i.e., someone dropping a piece of equipment). What cannot be avoided is our inability to accurately predict the incidence of such incidents. Unfortunately, in the electronic environment, with the exception of the cases recorded by Peter G. Neumann in his book, Computer Related Risks, there is little or no documented information that is publicly available for incidents that have occurred in the electronic environment; as a result, there is little that can be gained from any past experiences in this domain.

NETWORLD+INTEROP

LAS VEGAS • MAY 1–6, 2005

Network Infrastructure and Services
Wireless
Security
Performance
VoIP and Collaboration
Data Management and Compliance

See it All in One Place

ALL SYSTEMS GO

350+ Top Exhibitors on the Exhibit Floor with 8 Targeted Technology Zones and Pavilions

100+ Educational Sessions, Including 6 Comprehensive Conferences Revolving Around 6 Key Themes, 3 Special Interest Days and 36 Tutorials and Workshops

6 Visionary Keynotes by Leading Industry Executives

NetworkWorld Survivor Las Vegas

Visionary Keynotes

John Chambers
President and Chief Executive Officer, Cisco Systems

Hossein Eslambolchi
President—AT&T Global Networking Technology Services, Chief Technology Officer and Chief Information Officer, AT&T

Scott Kriens
Chairman and Chief Executive Officer, Juniper Networks

Sean Maloney
Executive Vice President General Manager, Mobility Group, Intel

Andy Mattes
President and Chief Executive Officer, Siemens Communication Networks

Your Source for Building a Better IT Infrastructure

MediaLive INTERNATIONAL

Register Today at www.interop.com

Use priority code MLAHNV42 and receive $100 off any educational product.

For this group of natural and accidental threat agents, each type is reviewed in isolation because they have only tenuous links to each other, and the main area of commonality is that they are not planned or directed.

Earthquake

The possibility of damage as a result of an earthquake is largely geographically dependent, but again there is considerable experience and documented case histories in the insurance industry of underwriting this type of event.

Fire

The likelihood of a direct effect on an information system from fire can easily be calculated, and there is considerable experience and a large number of documented case histories in the insurance industry of underwriting this type of event.

Wind

The possibility of damage from wind, normally most often thought of as a result of tornados or hurricanes (typhoons), is largely geographically dependent, because some locations are far more prone to wind damage than others. Again, there is considerable experience and documented case histories in the insurance industry of underwriting this type of event.

Water

The likelihood of a direct effect on a system from water, either from tidal wave, flood, rain, or damaged pipes, is again easily calculable, and there is considerable experience and documented case histories in the insurance industry of underwriting this type of event.

Lightning

Again, the likelihood of a direct effect on a system from the effects of lightning is easily calculable, and there is considerable experience and documented case histories in the insurance industry of underwriting this type of event.

Accidents

The threat to an information system from accidental misuse or damage is very different from the categories in the other groups above, because it can and will be affected over time by the attitude, disposition, and training of the staff, in addition to the environment. What separates this group from the malicious threats discussed later is the absence of malice or motivation. Again, this type of threat is generically well understood, and the probability of an event occurring as the result of an accident can be reasonably predicted from the actuarial data held by the insurance industry.

It is possible that more than one of these natural threats will affect an information system at the same time or shortly after each other. An example of this might be an earthquake that is followed by a fire as a result of the disruption to the gas or electrical services that the initial event caused. It may then, in turn, be affected by water used by the emergency services to douse the fire.

Malicious Threat Agents

For a malicious threat to exist, there must be an "agent" (an individual or a group of individuals) that will implement the threat. That agent must have sufficient motivation to carry it out, the capability and the opportunity to do so, and something to cause them to carry it out at that specific time (a catalyst). The threat agent will also be affected by other factors that will either enhance or reduce the likelihood of it being initiated by an attacker or an attack being successful (amplifiers and inhibitors).

Motivation

The motivation of an attacker to carry out a malicious attack on a system could arise from any number of drivers, which may affect the attacker either individually or in combination. There are a number of commonly accepted motivational drivers:

- Political
- Terrorism
- Secular
- Personal gain (including recognition)
- Religious
- Revenge
- Power
- Curiosity

Capability

The capability of an individual or a group formed into some type of organization to mount an attack and to sustain it at an effective level will vary with the complexity, resources, and sophistication of both the attacking force and the target. It may be sufficient for an attacker (threat agent) to mount an attack at any level to achieve their objective, but it may also require a high level of resources over a long period to have the desired effect on the target.

Opportunity

For an attacker to initiate an attack on a system, the attacker must have the opportunity to carry out the attack. This may be the result of a number of circumstances coming together, but for the purposes of this book, we constrain opportunity to mean either physical access or direct or indirect electronic access to the target. For a threat agent to carry out an attack on an information system, it must gain either physical access to the system (the threat agent gaining direct access to the place where elements of the system are located) or through either direct electronic access (through a connection from other networks) or indirect electronic access (eavesdropping). Without this, there is no opportunity for an attack to be initiated.

Catalyst

A catalyst is required to cause a threat agent to select the target and the time at which the attack will be initiated. The catalyst may be something that has an effect on either the target or the threat agent. An example of a catalyst might be the one that was considered earlier, when the U.S. Air Force were deploying to the Gulf. This could have been the catalyst for Iraq or its sympathizers to carry out an attack on the U.S. military in an attempt to prevent or delay the deployment.

Inhibitors

A number of factors (affecters) inhibit a threat agent from mounting an attack either on a specific target or at a specific time. As mentioned before, these may affect either the target or the threat agent. An example of this may be the perception by the attacker that the target system is well protected and that any attempt to attack it will be quickly detected. Another inhibitor might be the fear by the attacker of being caught as a result of publicity of successes by relevant law enforcement agencies.

Amplifiers

A number of factors (affecters) may encourage a threat agent to carry out an attack at a particular time against a particular target or group of targets. Again, these may affect either the target or the

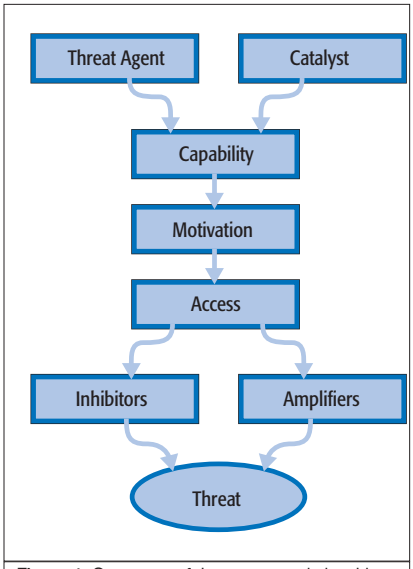


Figure 1: Sequence of threat agent relationships

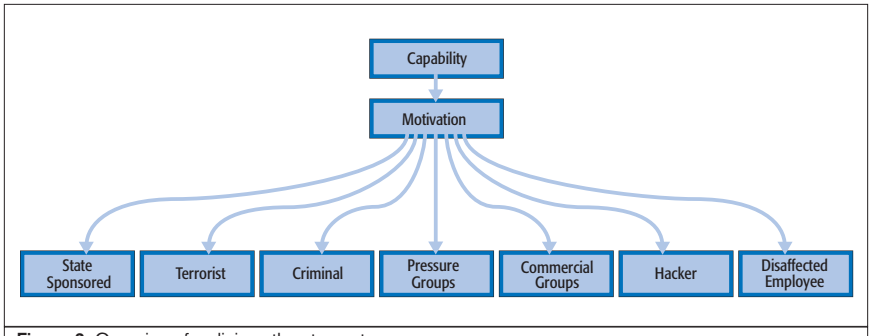


Figure 2: Grouping of malicious threat agents

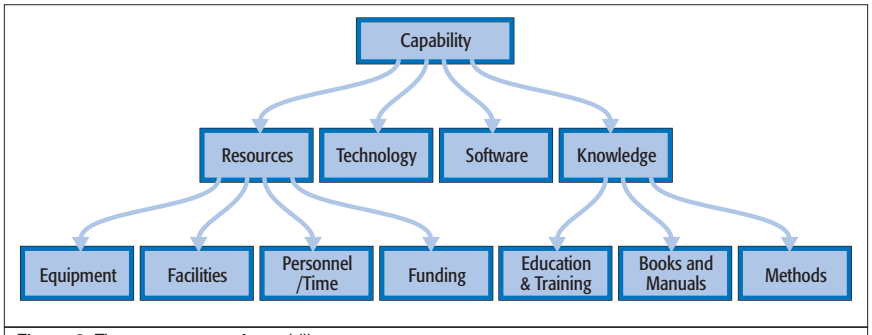


Figure 3: The components of capability

threat agent. Examples of this may be the perception that the target system is not well protected during a certain time period and that an attempt to attack it will not be detected or, even if detected, that no follow-up action will be taken.

System

For a threat agent to carry out a successful attack on a system, there are at least two system-related factors that must be present. The first is that there must be an exploitable vulnerability in the system the threat agent can use. For a vulnerability to be exploitable, it must be known, or there must be an expectation that it will be known to the attacker, and he must have sufficient access to the system to carry out the attack. The vulnerability may exist in the hardware, the operating system software, the applications software, or the physical environment in which the system is contained. The second factor is that the target system must be important enough to the organization that the loss of it or a degradation in its confidentiality, integrity, or availability would have a large enough impact on the business process of the organization to be considered a success by the attacker and/or the organization. Alternatively, it may be the only target available to the threat agent that will satisfy their requirements.

Sequence of Factors Involved in a Threat

As described above, for a threat agent to pose an actual threat to an information system, a number of factors have an influence. In reality, for it to pose a real threat to an information system, the threat agent must possess a capability and be able to gain either physical or electronic access. The level of access and its capability will influence the potential impact that such a threat agent will have. The likelihood of the threat agent being able to mount a successful attack will be reduced by factors that inhibit its ability to and will be enhanced by other factors. In addition, some type of catalyst will cause the threat agent to act when it does, depending on the motivation of the threat agent. The components of a malicious threat and their interrelationships are detailed in Figure 1.

Malicious Threat Agent

Malicious threat agents can be categorized into one of a number of groups. The groups detailed below are neither exclusive (the threat agent may belong to one or more of them) nor exhaustive. The main groups are shown in Figure 2. A malicious threat agent can be generated from any one of the groups or group combinations identified in Figure 2. This is not an exhaustive list of potential sources or groupings of malicious threat agents, because these change over time as high technology, education, national and international politics, culture, and a host of other factors have an effect.

Capability

For a malicious threat agent to be effective, it must have the perceived or actual capability to carry out and, if necessary, to sustain an attack and perhaps totally destroy the target and any subsequent replacement. The main constituent elements of the capability of a threat are detailed in Figure 3. For malicious threat agents to be able to carry out an attack, they must have the means in terms of personnel and equipment and the necessary skills and methods to be successful. They must also, in some cases, have a sustainable depth of capability to achieve their aims.

Inhibitors

A range of influences and factors can either inhibit or assist a malicious threat agent in carrying out a successful attack. These have been labeled as inhibitors and amplifiers. It is possible that the same influence, with a different value or in different circumstances, can act to inhibit or amplify either the likelihood of an attack or the potential for success of an attack. An example of this might be the security measures in place to defend a system. If they are weak, this will encourage an attacker

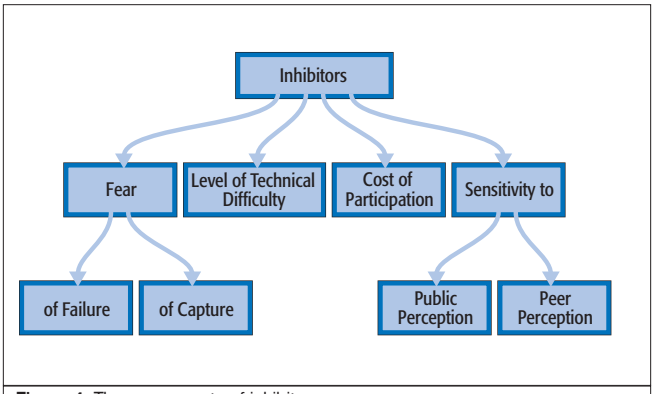


Figure 4: The components of inhibitors

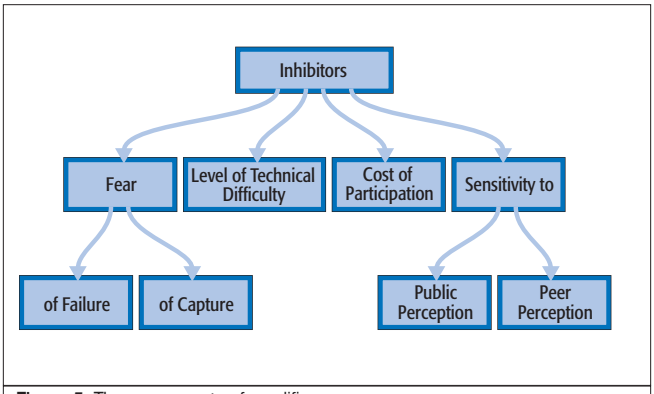


Figure 5: The components of amplifiers

to mount the attack; if they are strong, it may deter an attacker or prevent them from succeeding.

The influences that might act as inhibitors are detailed in Figure 4. An inhibitor can work in a number of ways. First, it can reduce the inclination of a threat agent to initiate an attack. Second, it can prevent a threat agent from initiating or carrying out a successful attack. Third, it can minimize the impact a successful attack will have. The fear of being captured as a result of conducting an attack may well act as a sufficient deterrent to the threat agent and cause it to decide not to carry out the attack. If the threat agent perceives its peers or indeed the public will hold him or her in contempt for attempting the attack (for example, if the target was a hospital or a charity), this may be sufficient to inhibit the attack. Also, if the level of technical difficulty that the threat agent encounters is sufficiently high, the threat agent may decide it is not worth the investment of effort required to attempt or continue the attack either on the initial target or at the current time. The factors that come together to inhibit an attack are, or may be, used as part of the protection and defense of the system and can assist in the reduction of the risk to the system.

Amplifiers

As mentioned above, the influences that may be an inhibitor in one environment can be an amplifier in another. The influences that might act as amplifiers to an attack taking place or being successful are detailed in Figure 5. The types of influences that amplify or increase the possibility of an attack occurring or being successful are varied and are dependent on the type of threat agent but include factors such as peer pressure or the level of skill or education. In the first of these amplifiers, there is the desire of the threat agent to be well regarded by his or her peers. His or her desire is to gain the recognition and respect of peers through the demonstration of skills, and this will strengthen his or her resolve to carry out the attack. The level of education and skill an agent possesses, or can gain access to, improves the confidence of the threat agent and also increases the likelihood of a successful attack. Another factor can be the ability to gain access to the information the agent needs to mount an attack, in terms of information on the target, other relevant information systems, organizations, or in terms of programming scripts and tools that can be run to conduct an attack; these may also increase the possibility of a successful attack.

Catalysts

The causal factor in a threat agent deciding whether and when to carry out an attack on an information system may be the result of an event, such as a publicity event for an organization with which the threat agent has a dispute or a dislike, or perhaps the start of an armed conflict between the threat agent's country or one for which they have sympathy, and an opponent. Another factor may be the circumstances of the threat agent, and any change (perhaps in location, social grouping, or employment status) may affect their ability or desire to carry out an attack or to be successful.

An attack can also be triggered by the advent of a new technology that makes what was previously not achievable a possibility. Finally, the commercial imperative to gain advantage against a competitor may cause a threat agent to conduct an attack. Figure 6 details some of the main groups of factors that can act as catalysts for an attack being initiated.

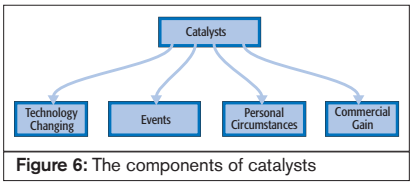


Figure 6: The components of catalysts

Motivation

The motivation of the threat agent is, by definition, a subjective area, and the threat agent may be influenced by a wide range of factors. Influential factors depend on the grouping or combination of groupings from which the threat agent originates. In some cases a number of these will act together to influence the threat agent.

In this excerpt the term threat, as it is used in this book, has been defined and explained, and the elements that need to be present for a threat agent to cause a problem have been examined. In summary, for a malicious threat agent to be effective, it must have the capability to carry out its attack and also the motivation and the opportunity.

About the Authors

Andy Jones is a research group leader at the Security Research Centre for British Telecommunications where he is doing research into the security of information and communication systems.
andrew.28.jones@bt.com

Debi Ashenden is a senior research fellow in information assurance at the Royal Military College of Science, Cranfield University, U.K.
d.m.ashenden@cranfield.ac.uk

FREE Gartner Research Note! Predicts 2005: Security Focuses on Attack Prevention
Go to gartnerinfo.com/securityresearch and enter priority code SSJ1

Gartner IT Security Summit 2005

Crack the Combination to Better IT Security

June 6-8, 2005 • Marriott Wardman Park Hotel • Washington, D.C.

Keynote Guests

Dr. Edward Amoroso
Chief Information Security Officer, AT&T

Roger Cressey
Former Chief of Staff to the President's Critical Infrastructure Protection Board

Phillip Q. Maier
VP Information Security, Emerging Technology, Visa/Inovant

Howard Schmidt
Former White House Cybersecurity Advisor and Chief Security Strategist of the U.S. CERT

Bruce Schneier
Renowned Security Technologist and Best-Selling Author

Bob Woodward
Pulitzer Prize-Winning Investigative Journalist

Amit Yoran
Former Director, National Cyber-security Division, Department of Homeland Security

1st Annual IT Security Innovation Awards

gartner.com/us/securityinnovation

Three high-impact days deliver hundreds of high-impact solutions.

See how to:

- Deliver appropriate levels of security and privacy protection with minimal risk, cost and complexity
- Reduce your organization's exposure to attacks and theft
- Discover the strategies and best practices that will improve your enterprise's resilience in the event of a disaster
- Determine optimal organizational structure, staffing and budget for your IT security operation
- Meet regulatory compliance in the most cost-efficient manner
- Be prepared for the new threat environment in the event of a disaster

For complete details go to gartner.com/us/itsecurity or call 1 800 778 1997 for a conference brochure.

Information Storage and Security Journal is a Proud Media Partner of Gartner IT Security Summit

information STORAGE+ SECURITY journal

Securing Remote Office Data with Wide Area File Services

HAVING YOUR CAKE AND EATING IT TOO

BY JOHN HENZE

FOR IT MANAGERS, consolidating all the corporate data in a single storage infrastructure at the data center is the easiest, most cost-effective way to manage and protect the data. To branchoffice users, WANs delay access to the centralized data and make a consolidated infrastructure unworkable. As a result, more than half of all corporate data is stored on largely unprotected branch office file servers and computers. But now, Wide Area File Services (WAFS) technology is letting IT managers give remote users high-speed access to the corporate data center, eliminating the risky local storage.

The Challenge of Remote Access to Centralized Data

Some businesses have consolidated their storage, security, and data protection assets in the corporate data center because centralized assets are easier to manage and less expensive to maintain than a distributed infrastructure. In most companies, however, critical data is stored outside the data center because the WAN connection between the branch office and the data center can't provide the instant data access that users demand. Network users tend to expect a lot of performance and tolerate little delay. A user opening a Microsoft Word document or PowerPoint presentation wants that file to open at the click of the mouse – not seconds or, worse, minutes later. That kind of immediate gratification is possible when the data is LAN-accessible. Accessing the same data over a WAN, however, causes a noticeable delay because of the architecture of the file access protocols.

The two primary file access protocols used today – the Common Internet File System (CIFS) for Windows environments and the Network File System (NFS) for Unix - send hundreds of administrative messages at each file open, save, and close command. These messages authenticate and authorize users, determine how an application will open a file, and define how data is presented. Most of these messages are short, synchronous communications that can't be compressed. They are the necessary overhead of the file protocols, and run quickly and transparently at LAN speeds. At WAN speeds, however, the quantity of sequential messages creates an unacceptable delay. Even high-speed WANs can't overcome this dilly-dallying, because the additional bandwidth does nothing to alleviate the lag in hundreds of consecutive round trips.

To give branch office users LAN-speed data access , most IT managers install a separate storage infrastructure in each branch office. This infrastructure includes file-and-print servers, a tape backup or replication system, and associated software, all attached to the branch office LAN. Administering this kind of storage infrastructure takes specialized expertise to create and implement regular backup processes and manage disaster recovery. However, many branch offices have little or no IT expertise on-site, and a limited budget for IT personnel.

How well data is protected at each branch office varies widely at enterprises without adequate on-site IT staff. Some have excellent data protection, while others may have none at all. And few enterprises have the kind of uniform data protection across all branches that's critical to meeting current regulatory compliance objectives.

In general, a distributed storage infrastructure leaves large amounts of data at risk, which undermines the effectiveness of the entire security infrastructure. Duplicating storage infrastructures at multiple branch offices also creates high total cost of ownership (TCO) for enterprise storage resources.

Speeding Data Access

The demands of regulatory compliance and business continuity are forcing large enterprises to protect and manage the data generated by branch-offices better. Businesses also are looking for ways to cut operating costs. So, IT managers are implementing solutions that give remote users higher-performance access to the corporate data center. This access lets IT managers consolidate branch office data in the data center, eliminating the expense, complex management, and questionable security of the distributed storage infrastructure.

Solutions for improving remote –data access over a WAN have traditionally focused on either network-optimization technology such as compression, or storage optimization such as replication. Compression-based products include PeriSphere from Peribit and ExpandView from Expand Networks. Replication-based products include Double-Take from NSI

Software, OnCourse from EMC/Legato, and Replication Exec from Veritas.

Both compression and replication technologies offer limited performance improvement for remote storage access. Compression can't deal with the latency issues inherent in file access protocols. Replication-based products tend to translate into higher server –management costs and may hobble data restoration. An emerging technology called Wide Area File Services (WAFS), which combines networking and storage management, can give branch offices high-performance WAN access to a storage infrastructure at the corporate data center. Latency and bandwidth-optimization techniques provide near-LAN access performance between the branch and the data center, while sophisticated storage caching with data integrity features enhances performance and network security. WAFS is compatible with CIFS and NFS, so users can access remote data using familiar business applications.

When a user opens a file, the local appliance determines whether it has a copy of it in its cache. Because branch office users generally work on their own files or ones created by a counterpart in the same office, most of the requested files exist in the local cache. The branch office appliance checks with the core appliance in the data center to ensure that the local cache file is the latest copy of the file.

If the cache copy is current, the user simply works from that while WAFS locks down the master copy. File locking helps ensure data integrity by guaranteeing that only one user can access the file at a time. If the cache copy isn't current because another user has accessed the file and changed it, just the changes are streamed over the network to update the cache copy. The interaction between the local and core appliances and the storage media is transparent to the user.

When the user saves the file, WAFS sends the changes made since the last file revision to the data center and updates the master file copy. WAFS also automatically flushes the appliance buffer at each close command to help ensure that the changes are actually propagated to the master copy before the session is released.

"WAFS offers the performance of a WAN **without the latency hassles"**

WAFS solutions include the File Engine from Cisco Systems, the IShared Remote Appliance from Tacit Networks, and the FilePort and FileController from DiskSites. WAFS is implemented by replacing branch office file servers with a network appliance and adding a similar appliance at the data center. These appliances are easy to install and require little management.

The appliance at the branch office attaches directly to the LAN and operates like a file server to local applications. It maintains cache copies of any file that a user opens, while the master copy remains stored in the data center on a central file server, storage area network (SAN), or network attached storage (NAS) device. The data center appliance connects directly to one or more NAS gateways or file servers, and does WAN-optimized file requests on behalf of the remote appliances.

Better Storage for Rapid ROI

Branch office storage infrastructures distribute a significant amount of valuable company data across multiple offices in complex and expensive infrastructures with limited resources to manage and protect them properly. WAFS combines the benefits of centralized storage with the user experience of local file-and-print services, helping enterprises cut the cost and complexity of managing critical data resources. A consolidated storage environment is simpler to manage, and high-speed access for remote offices gives users all the benefits of local storage. The WAFS solution is transparent and non-disruptive to branch users, providing the same ease-of-use as the local storage infrastructure. Users can easily access files from different sites, increasing productivity, and promoting distributed collaboration.

Consolidating the storage infrastructure improves data security, because the master copies of all the files generated at the branch reside in the data center. There, IT managers can protect data more effectively and simplify disaster recovery plans (DRP).

Finally, consolidating branch office storage at the data center dramatically cuts storage costs. For big enterprises with hundreds of branches, the cost of a distributed storage infrastructure is hefty. Cutting the amount of capital equipment needed for storage at branch offices eliminates much of this cost, resulting in significant savings. With a centralized model, the storage costs for each branch office are limited to the remote storage appliance and the actual storage media required at the data center. This storage media has a lower TCO than decentralized media, because of the higher utilization and more efficient management and security possible in a centralized environment.

With these combined savings, typical WAFS implementations have a fast return on investment (ROI). Cisco estimates that the typical ROI for a Cisco WAFS implementation is less than six months. With rapid ROI and a low TCO, WAFS is a financially effective way to manage an efficient storage infrastructure while ensuring user acceptance and satisfaction.

About the Author

John Henze is the director of marketing for Cisco's Caching Services Business Unit.

johenze@cisco.com

Wireless Security: Is Your Company Protected?

THE FIRST THING TO DO IS EVALUATE

BY TONY REDMOND

AS WIRELESS USE increases, companies that deploy corporate Wireless Local Area Networks (WLANs) open new dimensions of security vulnerability. Clearly, these companies need to address wireless security management as part of their overall security policies and architecture.

What'se surprising – and potentially more dangerous – are the security issues for companies that decide not to implement WLANs. The irony is that some companies choose not to implement a WLAN because they perceive that wireless is too risky, and they discover that this can be the most damaging security mistake of all.

The problem usually begins without malicious intent. Eager for increased mobility, employees can buy readily available wireless access points at any computer store. Someone who scans for unprotected networks can use these devices to attempt to access corporate environments. Also, when these wireless access points are attached to corporate LANs, the entire corporate network can be exposed. Often, "information trespassers" don't even need to try to crack any encryption. By default, most commercially available access points are completely open.

It's important to understand that wireless isn't a fad; it's an unavoidable new part of the communications landscape. Wireless connectivity has the potential to dramatically improve the productivity of mobile professionals, sales personnel, and field service technicians. By removing the procurement costs and effort of laying cable, it can substantially reduce network deployment expenses. Yet the benefits can easily be overshadowed by a single breach in security introduced, officially or unofficially, with the new technology.

Regardless of whether a company intends to implement a wireless LAN, it's imperative that wireless connectivity be a component of their security policies and procedures. They must include constant monitoring (either manually or with automated tools) of the airwaves for unauthorized traffic.

A Range of Risks and Vulnerabilities

While wireless has unique security aspects and industry standards, a WLAN isn't an island, it's part of an overall corporate information system. This relationship is also the source of much of the risk.

A rogue access point – or an unsecured access point on a corporate WLAN – is a hole that gives potential access to anyone who discovers it. It's not just the data being transmitted via wireless that's at risk. The trespasser can tap into the entire corporate network and could possibly read or tamper with any data. Thus, the security risk covers all corporate data, including confidential financial materials, e-mails, sales information, unfiled patents, and cached passwords.

Unlike fixed desktops, which are typically only used on the corporate network, mobile devices are frequently used on publicly accessible networks. Users will connect their laptops and PDAs on mobile networks or use them to download their mail or surf the Web. Outside the perimeter of the IT infrastructure, these devices are exposed to a greater range of viruses and worms and, once infected, can introduce them into the private LAN.

Besides entry to a corporate network via a rogue access point, improper configuration, or by cracking the WEP [Wired Equivalent Privacy] encryption on a WLAN, access can be provided through a wireless device that gets into the wrong hands through loss or theft. If it has an active connection, or cached user credentials, the new owner can easily access private applications and data.

The first step to take to avoid these and other risks and vulnerabilities is to do a thorough evaluation of corporate security, including wireless.

The Essential Security Evaluation

For an existing WLAN, or one in the planning stages, a number of key factors must be evaluated before deciding the security approaches that are needed. These factors include:

- Network topology and infrastructure
- Types of users and requirements
- Applications to be supported
- Value of the data (and financial impact if compromised)
- Existing security management solutions and policies across the organization
- Existing standards support
- Building structure and other devices in use or transmissions occurring in the vicinity (for potential of interference and to determine required bandwidth)

Cost analysis is a key element. The value of the data, and the financial impact if compromised, must be balanced against the price of combinations of security measures.

User convenience and speed of access must also be evaluated. Clearly, a major goal in creating a WLAN is the freedom and flexibility of mobile access to enhance business productivity. Some very stringent

security measures could be self-defeating if users fail to cooperate because they are complex or time-consuming.

Basic Recommendations

To ensure that a corporate WLAN remains secure, an automated assessment or security management solution is a fundamental requirement. For wireless-free environments, a regular assessment based on company security policies is the minimum requirement, and an automated assessment solution can assist in maintaining a wireless-free environment.

Until fairly recently, WEP (Wired Equivalent Privacy) was insufficient to encryp WLANs,. Knowing this some businesses avoided wireless, not understanding that there are many ways to deploy wireless security and maintain a secure WLAN.

Today enterprises are told to implement one or more of the following solutions: WPA (Wi-Fi Protected Access, WEP's replacement), 802.1x on both wireless and wired networks (which includes RADIUS authentication), and VPN solutions that a company may have already deployed for remote access. Role-based access control and password or smartcard-based authentication can also be mixed in.

Device security is also essential, and will be discussed in more detail in the next section.

Augmenting Device Security

Wireless-enabled mobile devices create vulnerabilities because they are highly portable and are used outside a secure corporate facility. As with wireless security in general, effective device security must be an integral part of overall company security. The standard procedure of reporting stolen or lost devices immediately can be a vital first line of defense.

It can be challenging to enforce security and impose strict authentication on devices such as PDAs. For example, long passwords can be difficult on devices without standard keyboards. Some products help encrypt the stored data, so that if the device is lost or stolen, attempts to read directly from it are thwarted. Other products are designed to erase all data if a wrong password is entered a set number of times, or if the device doesn't access the network in a designated time period.

Summary and Conclusion

As wireless use increases, so do the challenges for achieving and maintaining adequate enterprise security. Choosing to avoid the issue by not implementing a WLAN may be the most dangerous choice of all, because employees might turn to alternatives that open the door directly into the corporate network.

The first step to achieve effective wireless security is evaluating a wide range of factors from network topology and the value of the data to the types of users and applications. Most important of all, wireless security must be analyzed and implemented as part of a company's overall security planning. Effective wireless security is an integral part of the whole enterprise security strategy.

About the Author

Tony Redmond is the vice-president and chief technology officer of HP Services and the HP Security Program Office. He is responsible for setting the strategy and direction of HP's security initiatives that span its business units. These initiatives include trusted platforms, identity management, and proactive security management.

tony.redmond@hp.com

Best Practices for an Iron-Clad Backup and Recovery Plan

PREPARATION IS KEY

BY L.D. WELLER

TODAY'S SECURITY THREATS have become increasingly sophisticated and often combine several types of technology to maximize their impact on organizations. Even though businesses can't always prevent hardware damage from disasters like fires and hurricanes, they can protect data and information from disaster, manmade or natural.

The only way for businesses to ensure that their data is adequately protected is to integrate security technology and policies with regular and effective backup of systems and important data. To combat attacks, prevention tactics should include a multi-tiered approach that covers antivirus, firewall, content filtering, vulnerability management, and intrusion detection. In the event of a successful attack or crisis, businesses should set up a comprehensive backup and disaster recovery plan to reinforce the protection of their information and data from all sides.

The business costs associated with network downtime and data loss following a virus make secure backup and recovery an economic necessity. A recent survey of IT managers conducted by Insight Express, a research firm, said that 30% of the respondents figured that their companies lose at least $10,000 in revenue and productivity after a server failure. Data recovery can take anywhere from a few minutes to hours. For 85% of the respondents, recovering from a server failure takes two or more hours. Not surprisingly, however, 55% of them need five hours at the very least– with 13% needing more than 10 hours.

Disaster recovery should play an integral part of every organization's strategy for guaranteeing the safety and availability of mission-critical information. The Insight Express survey reveals that not all organizations implement backup and recovery measures; 35%of those surveyed don't back up on a regular basis, or only back up once a month. Even more startling, more than half (55%) don't back up their entire system on a daily basis.

These numbers are alarming, particularly since 72% of the respondents said their organizations suffer at least one server failure a year. By using a backup and disaster recovery solution, IT can recover quickly from a server failure caused by a virus or worm and get back to its "original state."

The risks associated with failing to implement an adequate disaster and recovery plan have proven to impact the success of a business significantly. Gartner notes that two out of five enterprises that experience a disaster go out of business in five years. Even more, Insight Express found that 54% of its respondents don't see the need to back up the entire system more often. Almost 29% said backups are too time consuming and there aren't enough resources to back up the entire system more frequently.

Despite the size and type of organization, it's imperative that all companies employ every available defense to protect mission-critical data and operations and prevent the loss of time, money, and customers. To successfully implement a backup plan, they should take a few best practices into consideration.

Implement a Comprehensive Recovery System

A good backup solution will create compressed images of a server's volumes, including its operating system, server settings and preferences, which will enable complete restoration of system and data volumes, or individual files and folders, in a matter of minutes. A system with 5GB to 8GB of data can be recovered in 15 minutes.

Disaster recovery, however, doesn't stop at the server. With a variety of technologies to choose from and use inside or outside the office, desktops, laptops, and handheld devices have become vulnerable targets. Anyone who's lost information on a laptop or had it die on them understands the headaches of retrieving the information stored on the hard drive. Insight Express notes that 72% of those in its survey don't include laptops in their backup routines, 50% don't include desktops, and 81% don't include PDAs. As these technology devices saturate the workplace, and an increasing amount of critical data is stored on them, it's imperative that IT departments consider desktops, laptops, and handheld devices a priority, not an afterthought.

Verify Backups

Being able to recover data is imperative to securing it. Organizations often find that the problem isn't in creating backups, but in verifying their recoverability. "False backups" have proven detrimental when organizations realize that the backups failed and they lost data after a virus attack. Regularly scheduling test recoveries ensures that backup procedures work properly when they're needed.

Partition the Hard Disk

Partitioning a hard disk can help organizations reduce the amount of data needed for backup stores. By creating separate partitions for data and applications, IT can quickly back up mission-critical data after a virus attack without utilizing valuable storage space on applications.

Partitioning can also improve organization and simplify the backup and recovery process. By assigning a set of files represented by its own drive letter, IT can keep track of the partitions that has to be backed up in accordance with the disaster recovery method selected.

Disk-based vs. Tape-based

To back up data and information businesses have the flexibility of using both tape-based and disk-based solutions. Many organizations, however, leverage the strengths of both these solutions to create one comprehensive solution that uses tape as a direct backup and disk as a day-to-day backup. Disk backups provide flexible and immediate access for everyday use, without having to shut down servers and take a company off-line. Once saved to disk, it's then wise to convert the disk backups to tape where companies can store them for a long period of time.

Backup Policy and Procedure

Having a backup plan that actually recovers information is crucial. Insight Express found that at least 36% of its survey respondents didn't have a backup strategy delivers complete recovery 100% of the time.

Implementing specific procedures for creating backups and creating an action plan for recovery are essential to any modern business. Disaster and backup plans, however, don't come in a "one size fits all" package. Businesses should tailor recovery plans to meet their specific needs. For example, financial system should be backed up as often as possible, while backing up word processing documents once a day might be sufficient, depending on the organization.

The first step in planning for recovery from a virus attack is an environmental assessment. When considering what to include in the plan, companies should keep the following in mind:

- The value, monetary or otherwise, of the most important network resources.
- The possible threats these resources face and the likelihood of the threats being realized.
- The impact of those threats on business, employees, or customers, if the threats are realized.
- The resources that need to come back online.
- The amount of time each resource can stay down.

After making a complete assessment, companies should set an allowable downtime for each resource and create a decontamination process for viruses and worms. This simple assessment can equip IT departments with the necessary information to bring systems back online quickly and successfully and retrieve data.

Protection at Every Level

While backup isn't meant to replace regular security measures, organizations that deploy an effective storage and recovery strategy are well on their way to protecting mission-critical data. To successfully protect from today's threats, organizations must take both security prevention and recovery tactics into consideration.

As threats continue to evolve and increase in complexity, IT should take the time to implement and execute various security standards internally, incorporate disk and tape storage, partition hard drives, and plug other lurking holes in their system.

When choosing backup and recovery products, it's important to pick ones that make it possible to automatically adjust backup routines to occur prior to a new application installation, user logon/logout, or storage upgrade. To avoid disrupting the network, pick products that will let you throttle back the speed of creating a backup during working hours.

To safeguard information from a catastrophic event, such as a natural disaster, organizations should implement physical security measures by keeping duplicate server backups in a different locations from the actual servers.

Technology should always be accompanied by internal policies and procedures that emphasize caution. A multi-faceted approach to enterprise computing, incorporating security prevention tactics and disaster recovery plans, will ensure the best possible defense against attacks. There's no doubt that cyber-attackers will continue to use new technologies to design more powerful viruses, creating more problems for IT staffs everywhere. Preparation is key to avoiding a complete wipeout of your data.

About the Author

L.D. Weller is senior product manager for Symantec's enterprise administration business unit.

Security's White Knight

UNIFY CORPORATE ENDPOINT SECURITY WITH A 'WHITELIST'

BY DENNIS SZERSZEN

FOR THE BETTER part of a decade now, companies have been buying defensive security technologies to secure their IT networks by identifying, defining, and then blocking the threats. By constantly updating a "blacklist" of things that should be barricaded outside of the network, security administrators figured that they could keep their PCs and servers from being infected by malicious code.

In the current environment, however, blacklisting has become a Herculean task of decreasing effectiveness.

Zero-day attacks are now common. That's when there's no blacklist signature for the malicious code until after the damage is done.

New worms, viruses and vulnerabilities are discovered daily, and a new generation of blended threats – attacks that combine some of the most harmful and pernicious characteristics of the latest worms and Trojans – are taking their toll on corporate systems and networks.

Organizations have become so reactionary in defense of their systems – and so narrow in focus – that they're spending a lot of their resources on the ad hoc defense of single exploits.

Every time a big enterprise mobilizes to test and apply a patch, it can strain both time and the budget – emergency patches often cost hundreds of thousands of dollars. And a zero-day attack would render the updating useless.

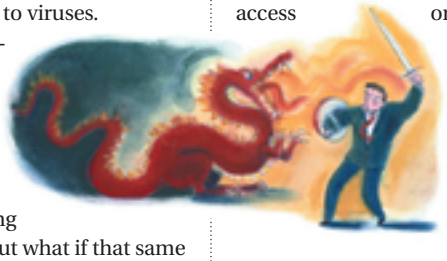
How big is the problem?

According to Carnegie Mellon University's CERT center (www.cert.org), 3,784 vulnerabilities were reported in 2003. In the first half of 2004 alone, there were 2,683. Keeping up with these vulnerabilities is nearly impossible as shown by the Slammer worm outbreak last year, which exploited a vulnerability in Microsoft SQL Server 2000, a vulnerability that was identified at least six months before Slammer hit.

Enterprises need to be more proactive. The best solutions emphasize a number of things, including technology better suited to the distributed nature of today's enterprise, as well as the better configuration of existing systems. However, perhaps the best solution is to take what organizations already know and flip it on its head.

Enter "Whitelisting"

"Whitelisting" is the opposite of blacklisting. It means setting a pre-defined list of applications and devices that can reside or function on corporate machines while blocking everything else by default. Whitelisting shelters administrators from spinning their wheels maintaining blacklists of devices, testing and deploying every software patch, updating threat signatures and frantically reacting to viruses.



For example, a security manager can clearly define what applications types are allowed on an employee's PC such as Microsoft Excel for doing spreadsheet analysis. But what if that same employee tried to download an executable file from eDonkey, Kazaa or another unapproved file-sharing Web site? The file would be blocked by default, along with anything – such as viruses, worms, Trojans and the like—associated with that download that could potentially devastate the company's IT environment.

Whitelisting also lets enterprises implement a security solution that's unified across departments, an issue that's been facing organizations for years.

Unification begins with managing credentials, but it needs to move further towards the comprehensive use management of an enforceable security policy such as whitelisting. Ever since systems were "decentralized" in the 1980s, enterprise IT organizations have struggled to "recentralize," beginning with their server base.

Just because companies have rigid change management processes in place doesn't necessarily mean they can control which applications are running on their servers. The problem is one of maintaining the performance, availability and security levels of networked storage. When these systems are impacted so are the users who rely on them. One place where change control can't be enforced is at a storage farm. The enterprise may own the farm, but not the individual storage allotments. Unauthorized, sometimes rogue applications, can make their way into network storage with no way to stop them.

The ideal that unified security aims for is managing user access privileges. The overall security principle is called "least privilege" and means granting users access only to the resources they need to do a specific job. The concept is tantalizing – being able to group users by job function, location or some other similar attribute.

The thing that lets enterprises assign rights – and the corresponding credentials – is fairly robust since every major enterprise has bought into this concept and has implemented a solution from a major vendor. Unification of access rights solves several major problems, including controlling who can get into the network and what information they can access. It also promises to solve the administrative nightmare that any centralization solution encounters. Unification is just the first step, however, toward policy-based control over a company's information resources.

What early solutions promised – but couldn't deliver on their own – was centralized control over the applications (and consequently the data) that made up an enterprise's overall information assets.

Another problem that access control can't solve by itself is the performance and availability of the corresponding systems and applications, along with the interaction of unauthorized applications with authorized ones.

Change control strives to add stability to the IT environment, but how can it do that if it can't control applications? Besides, access control is limited to just enabling or disabling access. It can't control how or when applications are used because it doesn't provide enforcement.

Something must reside with the end users and servers to be effective.

Some solutions would have you believe that adding access control to a network is enough. For example, making sure you have the right credentials, the right software installed and the right updates in place prior to permitting network entry. But this is naïve because a Trojan or other pieces of malicious code could be sitting, watching, and waiting to attack while users go through their daily routines.

Another drawback of access control is its inability to keep track of what applications are being used and by whom. Most enterprises can't track software use down to its lowest common denominator – the end user – because asset and inventory controls simply count software occurrences, not whether applications are actually being used or by whom. Without that information, how can an enterprise declare effective control over its installed base?

If all these missing pieces could be tied together with a policy-driven access management system, enterprises could establish a unified strategy that would solve most of the complex issues facing network security.

Whitelisting addresses each of these issues by empowering the IT administrator. Only applications that are known to be safe and policy-approved can work on corporate endpoints, adding an enforcement component to the policy. For example, if corporate policy prohibits instant messaging but can't stop such software from being downloaded, employees could defy policy and install IM channels on their PCs. By leaving all IM applications off the whitelist, companies can be sure the policy is enforced.

Security solutions based a whitelist also lets enterprises group end users by department. So, if the accounting department needs software that's of no value to the marketing department, the company could add the software to the accounting department's whitelist and outlaw it on the marketing department's machines. That way companies can rest easy knowing that the programs on their machines are safe and serve a specific business purpose, right down to individual departments.

Not to oversimplify, but which is it easier, to list the people you would let in your house, or to track the billions of people you wouldn't?

Whitelisting is simple, proactive and lets administrators keep tabs on the scores of file types, devices and applications that matter to a company's business rather than the thousands that don't. ■

About the Author

Dennis Szerszen is vice president of business development at SecureWave (www.securewave.com), a Luxembourg-based maker of endpoint security software with offices in Washington, DC, and Research Triangle Park, NC.

dszerszen@securewave.com.

ISSJ | Advertiser Index

Sidestep the Data Storage Blues

AN ARGUMENT FOR SANITIZING HARD DISKS

BY WINN SCHWARTAU

ANTONIO MARCELLI KILLED people for a living. At least a few he admitted to. The feds caught him, he turned state's evidence, testified in open court against the capos and subsequently entered the witness protection program. He was safe until his new name and location hit the Internet.

A computer junkie from Kentucky had bought a heap of old hard drives that the Justice Department had discarded. Lo and behold, names and addresses of people in the witness protection program popped up in a perfectly readable format.

Embarrassing? Yes. Deadly? Potentially. What went wrong? The DOJ forgot a simple fact: the value of data doesn't die when a hard disk (or tape, etc.) is tossed in the garbage.

In 2004, techno-journalist Simpson Garfinkel investigated the lack of care he suspected permeated every organization. He bought a slew of used drives from various places and discovered that 90% of them still contained readable data.

Was that data valuable? Maybe not to Simpson, but to *someone* certainly.

The fact is your discarded hard drives and tapes still contain your company's valuable and private data. You probably just upgraded, changed suppliers or formats. Maybe some of the data is old and useless – maybe. But how can you be sure without a complete, time-consuming, and expensive contents analysis? You can't.

That's where hard disk sanitization comes in.

To protect confidential company data and comply with all the myriad regulations (GLB, SarBox, HIPAA, etc.), we can't be complacent about the resilience of magnetic storage. For example:

> When a file is erased in some systems, only the file name is discarded. The data remains.
> If a file is overwritten with random '1s' and '0s,' forensics experts can still retrieve the original data by means of fringe track analysis.
> Even if the data is overwritten several times, there are advanced magnetic analysis techniques that recover several layers of data deep.
> Slack files and other hidden temporary data system repositories are often forgotten by the OS and can be easily recovered by COTS forensics tools.

We all spend a great deal of time protecting the security and privacy of data when it resides in our data centers, wired to the world, but we sorely lack in protecting that same data when it comes off-line and is relegated to the dust heap of the local computer fairs.

The first thing a company needs is a data-destruction policy. In the physical world this means policies and procedures for shredding papers, burning waste and mangling CDs and floppies. In the logical world, a destruction policy has to be quantum in nature, making sure that the magnetic orientations of the ferrous oxide particles produces no valuable information. Here, as in so many other areas, policy – from the top down – is critical.

There are three fundamental ways to destroy data on magnetic media.

The most extreme involves totally disassembling the drive, scratching off the magnetic surfaces of the disks and then melting the constituent components into post-data sludge. For the truly paranoid – think the government and military – this is the only rational approach.

For most of us, this is overkill. Keep in mind that security is never perfect. All we can do is raise the bar by increasing the impediments and obstacles to put any potential data recovery well beyond any cost-benefit analysis.

The second way is to overwrite the entire disk with random data. Not bad, but if someone really, really wants to target your firm, and is willing to go through some extra effort, a certain amount of the original data is still recoverable.

Third, and the one I prefer is degaussing. A degausser generates an intense magnetic field, strong enough to scramble the magnetic bits and pieces so there's really no reliable or cost-effective way to glean any of your secrets.

The degaussing method is the easiest, most effective method, requires the least manpower and no CPU time. Merely cycle the drives through, say the shipping department, and procedeuralize the degaussing. It's that simple.

When developing a sanitization or data-destruction policy, look at your needs from several different angles:

> Time is a key determinant in how to sanitize data. How long is the data going to be valuable? One day? One week? A year? Forever? The longer the data has real value the higher the quality of the data sanitization needed.
> How many compliance standards do you have to meet? Keep in mind that your compliance requirements live on, even if the hard disks are long gone.
> Ask your lawyers about the company's liability if proprietary, employee, or customer data reaches the public domain – read the Internet.
> What is your firm's downstream liability if the data disclosure affects other organizations and people not directly under your control?
> Is creating a policy and assigning one person to carry out the sanitization process reasonable insurance against any future litigation?

Think about the information your systems contain that you *don't* want on the front page of the New York Times or posted at www.thesmokinggun.com. Think about how a corporate adversary could use your discarded drives to profit from them and harm you. Think about your personal responsibility in data protection.

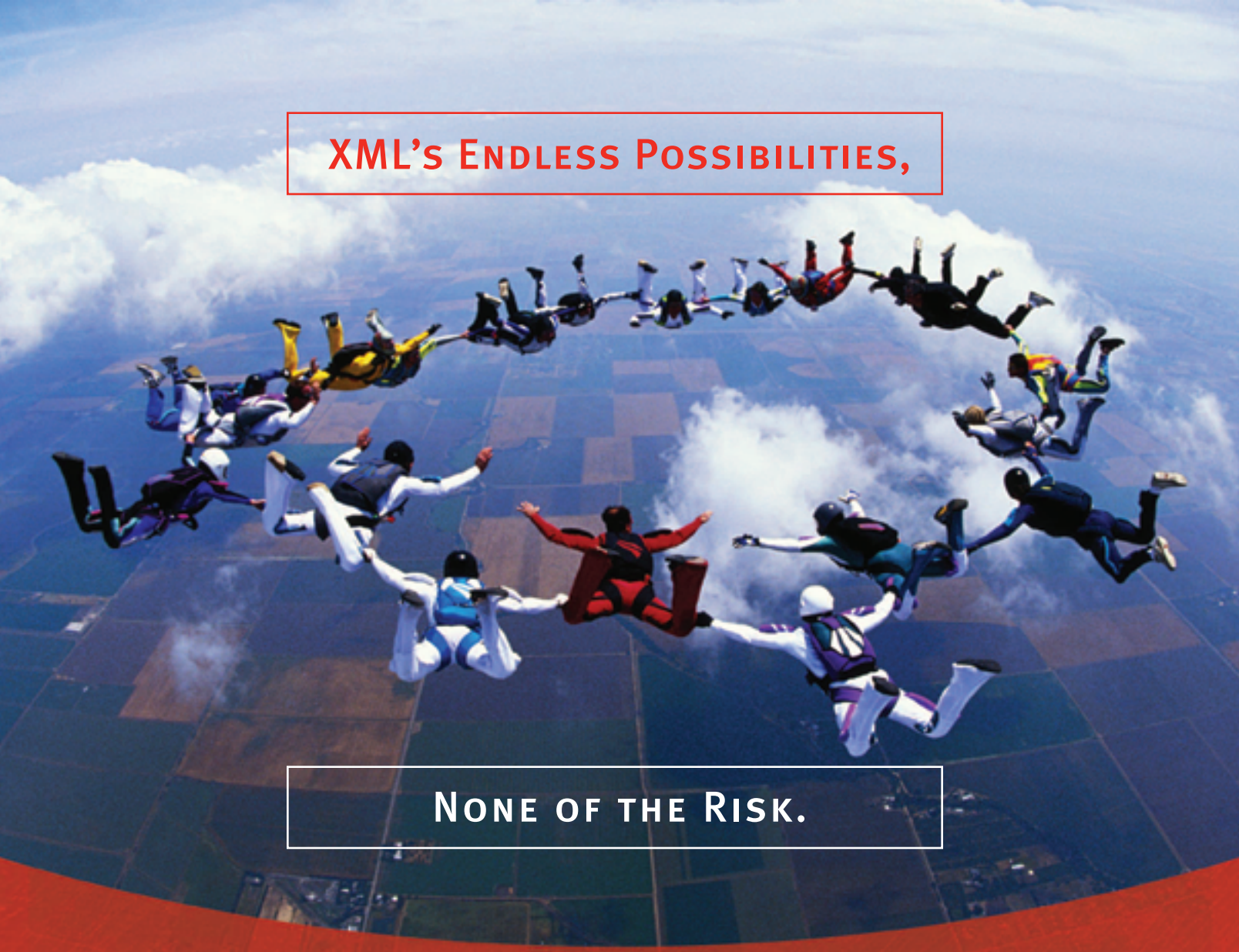
Just think about it. And then make sure you aren't singing the Post Storage Blues.

Resources:

> www.datadev.com/hdhardisdriv.html
> www.datalinksales.com/degaussers/hd1.htm
> www.cyberscrub.com
> www.diskwiper.com

About the Author

Winn Schwartau is CEO of www.TheSecurityAwarenessCompany.Com and Trusted Learning, Inc. www.TrustedLearning.Com. He's a popular author and speaker with thousands of credits to his name.
winn@thesecurityawarenesscompany.com

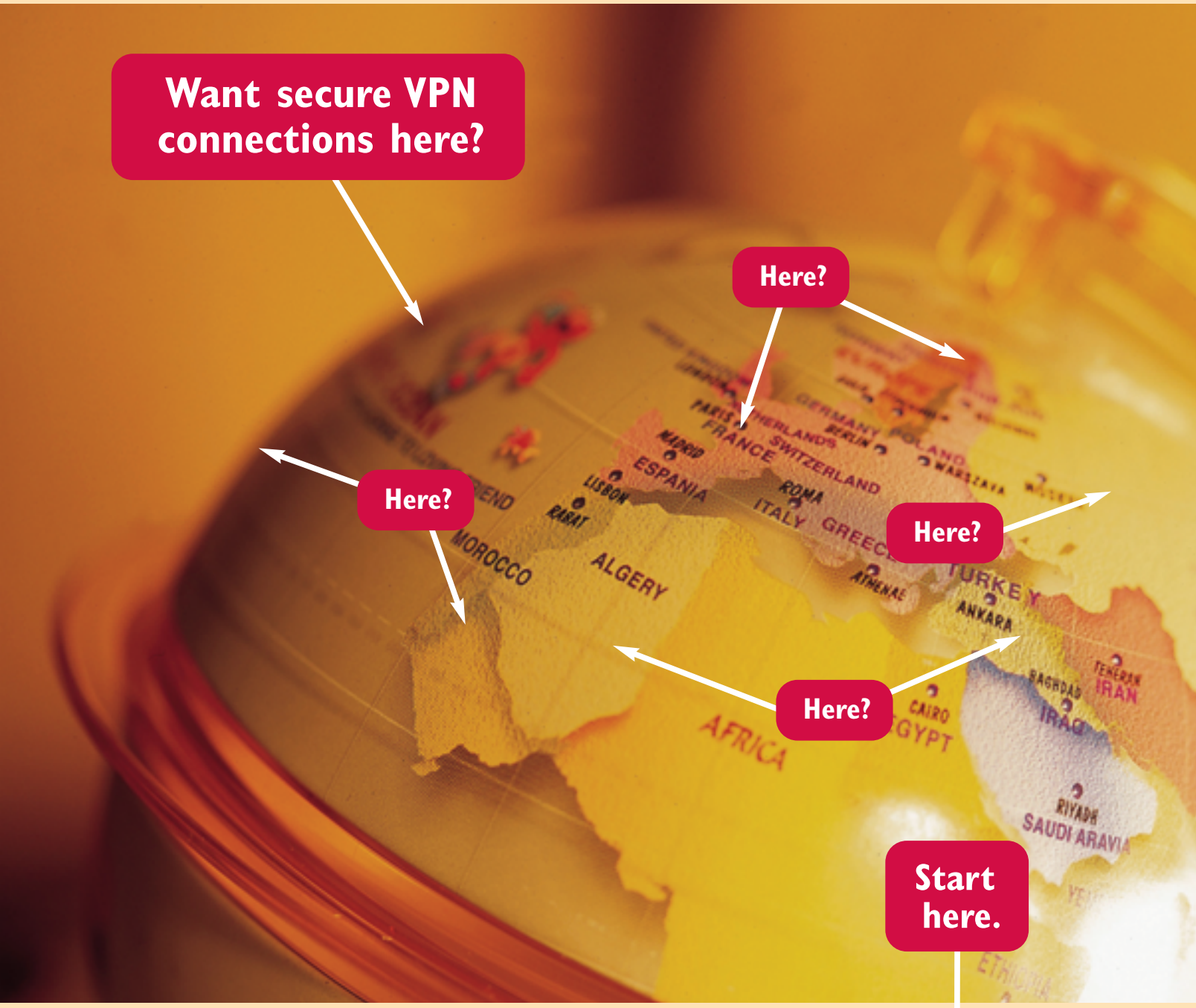


Forum XWall™ Web Services Firewall - Reinventing Security

Security should never be an inhibitor to new opportunity: Forum XWall™ Web Services Firewall has been enabling Fortune 1000 companies to move forward with XML Web services confidently. Forum XWall regulates the flow of XML data, prevents unwanted intrusions and controls access to critical Web services.

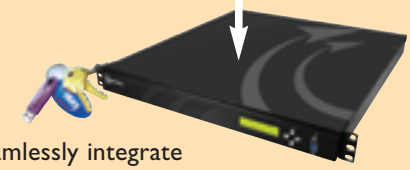
Visit us at www.forumsys.com to learn more about how you can take your next leap forward without increasing the risks to your business.

FORUM SYSTEMS™ – THE LEADER IN WEB SERVICES SECURITY



iKey iGate

Easy remote access doesn't have to mean easy hacker access!



A remote access VPN with SSL is the easiest way to get everyone up and working quickly and securely. With SafeNet's iGate server protection, you also have the option of adding iKey USB authentication. Together they seamlessly integrate to form the industry's only high assurance SSL VPN platform, combining integrated 3A security and application level control with powerful user authentication. Plus an access and control interface that gives you instant authorization and authentication. So if you'd like the ease of a remote access VPN with SSL—without the security worries—call SafeNet today.

Call 1-800-696-5308 to be SafeNet sure.
www.safenet-inc.com/igate

SafeNet®
The Foundation of Information Security

APPLICATIONS - AUTHENTICATION - REMOTE ACCESS - ANTI-PIRACY - LICENSE MANAGEMENT - VPN/SSL