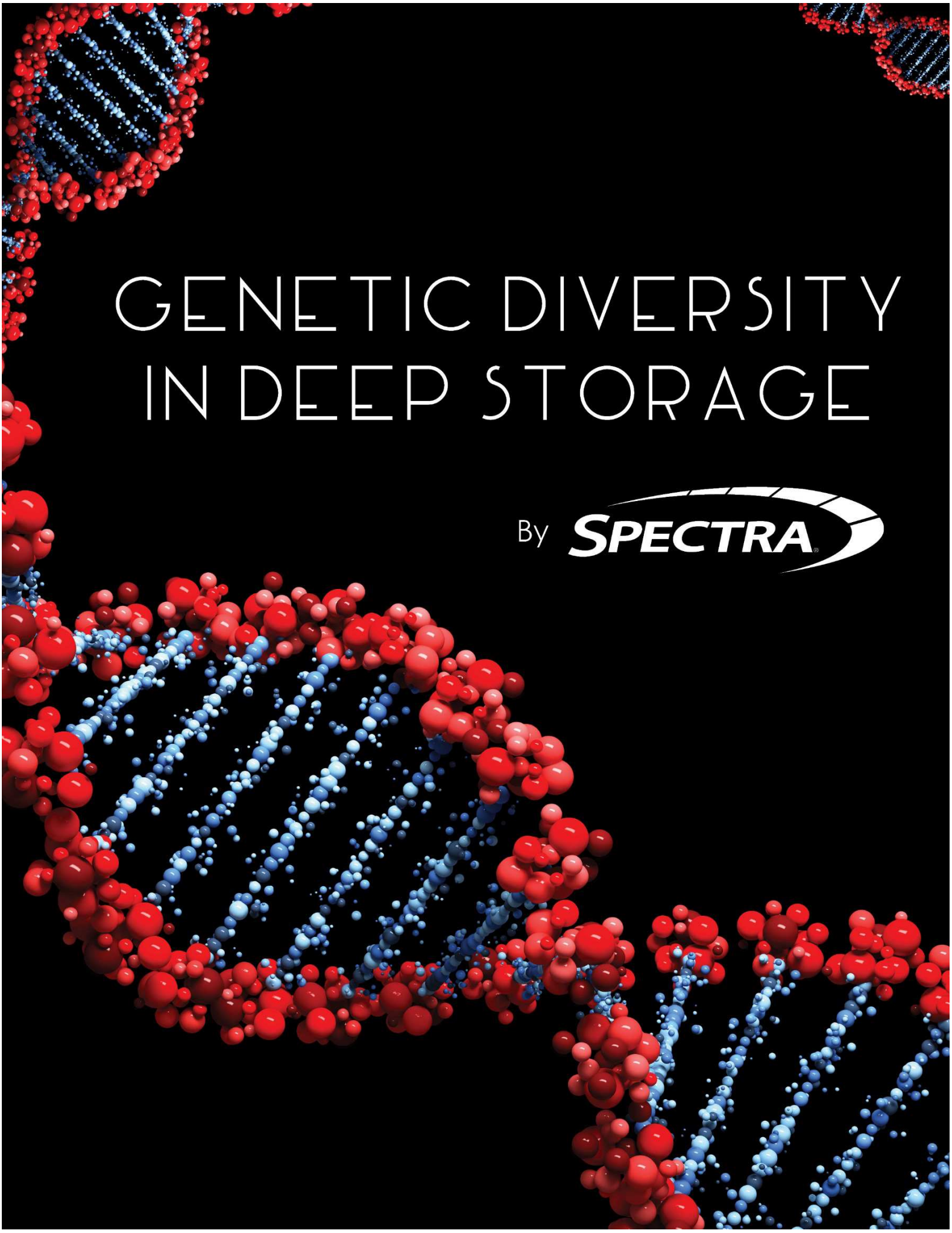
GENETIC DIVERSITY IN DEEP STORAGE

By SPECTRA



SPECTRA

DEEP STORAGE EXPERTS

Cyber Endgames

Genetic Diversity on the Front Lines of IT Defense

July 2015

800-833-1132 • 303-449-6400 • (Fax) 303-939-8844 • 6285 Lookout Road • Boulder, CO 80301 USA SPECTRALOGIC.COM

Contents

ABSTRACT

Humankind is the only species which builds upon knowledge handed down from previous generations. Survival of governments, corporations, societies, and even humankind all depend on information. Today's reliance on *existing* information and the rapid advancement of *new* information has led to the call for better efforts in digital preservation, the preservation of knowledge. As organizations seek to build more robust data repositories, preserving information for the future must not be hindered or destroyed by a single point of failure. Still we see such failure points being introduced into some of the largest data repositories on earth through failure to create genetic diversity in data storage mediums. This paper examines the new set of threats endangering information today and suggests simple methods to assure the survival of information after such attacks as well as the organizations which depend on that information.

PART ONE: FAMINE, GENETICS & TODAY'S IT

LESSONS FROM THE U.K.

GENETICS AND THE ROYALS

In 1845 Great Britain was a superpower, but its reigning monarch, Queen Victoria, faced a grave challenge. For generations, her royal ancestors practiced cousin-to-cousin marriages (in fact, her husband, Prince Albert, was her first cousin). As a result of genetic homogeneity, she suffered from hemophilia, as did other members of the royal family. Compared to their genetically diverse subjects, they lived their lives at heightened risk because even a small wound could result in profuse bleeding or death.

GENETICS AND FAMINE

Closed genetics and homogeneity foster weakness

During Victoria's reign, another kind of inbreeding plagued Ireland. It appeared not in royal form, but in the humble soil of Irish farmlands. In the early 19th century, Ireland's population and poverty rate increased rapidly, and potatoes became the country's primary – and in many cases, only – food source. By 1840, the Irish relied on a single species, the lumper potato, for food. All such potatoes were genetically identical. When the blight arrived, virtually every potato that might have matured in Irish fields instead turned to slime. The subsequent, nearly decade-long Great Famine killed at least a million people; another two million left the country in desperation.

TAKEAWAYS FOR A DIGITAL AGE

The impaired health of the royal family and the devastation of the Great Famine demonstrate one indisputable fact: If, for whatever reason, we ignore the imperative for genetic diversity, disaster follows. With diversity comes strength, resiliency and perpetuity.

Genetic diversity builds strength, resiliency and perpetuity

Today we are turning a blind eye to a lack of genetic diversity in a new realm – the cyber world of our own creation. The data storage used by our institutions, and upon which our daily lives rely, is now falling prey to genetic homogeneity. Like the royal family, our storage methods are too closely related, that is, they lack sufficient diversity. Like the lumper potatoes of the Great Famine, our information is vulnerable to eradication by a single attack.

Our inaction could result in a digital calamity sufficient to propel western civilization into a new Dark Age.

DATA CENTERS: DIVERSITY BY DESIGN

Until recently, homogeneity was not an issue for the information age. The era of IT came into being such a short time ago, it typically thrived on an influx of new, disparate technologies and rapid rates of change. The sheer *volume* of change seemed to assure a diversity of technologies and IT methodologies. However, as information and IT grow at exponential rates, we stand closer to challenges, dangers and potential disasters that we never before considered.

THE OLD DAYS: BACKUP WITH DIVERSE TECHNOLOGIES

In 1995, a data center could store and process volumes of information that tallied in gigabytes per week. Today we might measure weekly volumes of data in hundreds of terabytes, petabytes and beyond.

In some foundational ways, however, the storage approaches of a 1995 data center have been weakened or negated. For example, information was stored across a range of media (such as various kinds of tapes, hard drives and other backup systems).

Though often overlooked, the 3-2-1 rule assures genetic diversity.

If one media failed, another supplanted it. And while backup and archiving processes were still pretty much in their infancy, a couple of basic principles remain valid today. Data backups were kept in a variety of super-secure locations – not just in a 100-mile-radius cloud, but throughout the world. Fire, flood, earthquake, war and other disasters might strike and destroy a few selected locations, but not all of them. The data center also followed the discipline of "Grandfather-father-son," a method of rotating tapes in backup and making sure a set is sent off-site. Similarly, the "Tower of Hanoi" rotation used a more complex method of pegging disk volumes to tape sets in backup. Both approaches benefited from creating genetic diversity in storage medium

In this manner, the data center of 20 years ago incorporated an ingenious genetic-diversity design. First, use a mix of media and locations to maximize information security. Next, recognize, understand and plan for worst-case scenarios. Finally, always prepare to rebuild from the ground up – using the information you rescued from disaster.

Though often overlooked, there's a simple, modern day version of data protection which plays off of previous approaches, the "3-2-1 Rule." Always have three copies of your data on two different mediums with one copy off-site. It's a simple guarantee of both genetic diversity and disaster recovery. But as we will see… it's not just consolidation *within* data centers that has started to threaten the strength of our data storage approaches.

STORAGE HOMOGENEITY: THE WEAK LINK OF CONSOLIDATION

We tend to think of storage in terms of media (disk vs. tape, for example) or brand (IBM, Dell, Spectra Logic, EMC, and so on). What isn't immediately apparent is the consolidation that has occurred in the manufacturing of the disk and tape devices themselves. For instance, there were more than 70 manufacturers of tape drives in the 1990s; today there are roughly three. Disk drive consolidation is about the same – from dozens in the 1980s and 1990s to just three today. And only two of them are players in the enterprise storage market (Figure 1).

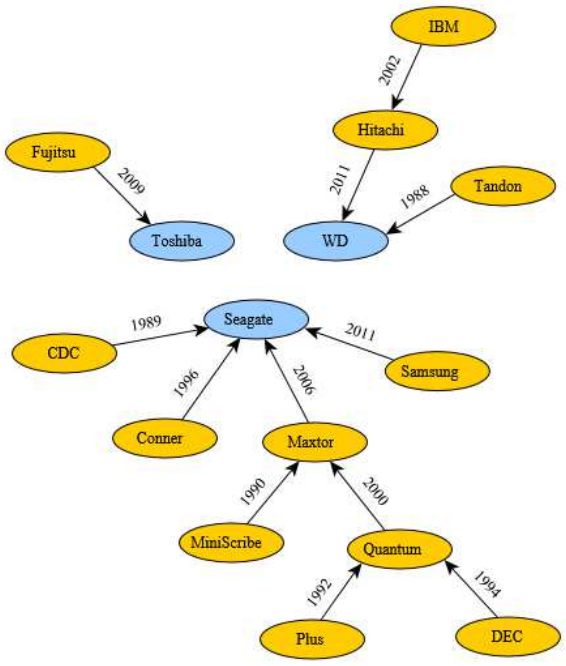


Figure 1. Data storage consolidation

We cannot control consolidation in manufacturing. It's an unfortunate fact today that malware designed to attack disk firmware must succeed against only two manufacturers in order to wipe out roughly 90 percent of all enterprise disk drives in existence. Likewise, our tape choices will most likely be between LTO and TS. What we *can* control is whether we further consolidate to all disk – or even all tape – for archive storage. While consolidation is generally considered a positive in the world of storage, too much consolidation comes at the cost of losing the robustness required to protect the information which drives organizations.

Deduplication: Consolidating More than Just Data

A form of data reduction, deduplication saves space and cuts cost by recording only selected, non-redundant information. While data deduplication has become a standard backup tool in many data centers, there are issues associated with deduplication flexibility and its non-linear approach to information storage when it's deployed as an archive.

Deduplication has its place and limitations

Used in conjunction with other media, deduplication has a strong place in business backup plans. Backup is a frequently produced copy (or series of copies) of your most recently created information. An archive (whether it exists on hard drive, tape or DVD) is a perfect replication of grouped system information created at specified moments in time. In its truest form, an archive is information that's been removed from primary storage. By archiving information from primary storage, backup windows can be significantly reduced. Archives don't suffer from the constant replication that befalls primary data. Therefore, archives do not benefit from being "deduplicated." The extra cost of a deduplication appliance, often justified in backup, goes largely unused in archiving. And because deduplication stores exclusively on disk storage, it's become a major contributor to storage homogeneity. Yet another example of losing genetic diversity in storage mediums.

FURTHER CONSOLIDATION THROUGH THE CLOUD

There is no denying the cloud's convenience and ease. *Forbes* magazine,[1] citing the "International Data Group's 2014 Enterprise Cloud Computing Study,"[2] notes that nearly 70 percent of businesses already use cloud technology, and an additional 18 percent plan to do so (Figure 2).

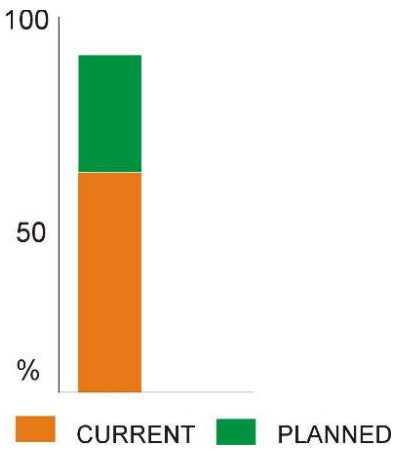


Figure 2. Business use of cloud IT

While accurate, reports such as these often fail to address both types of cloud - "Compute" vs. "Storage." Compute cloud services offer Software as a Service (SaaS) applications or other computing services. Storage cloud services (such as Amazon Glacier) offer remote data storage for information that is typically static. For long-term storage of large data sets, however, storage cloud services seldom deliver on the promise of low cost.

More data + fewer baskets = less security

When you evaluate archives in the > 300 TB range, you'll find that cloud storage costs exceed those of local infrastructure in just two and a half years. That's for the cloud storage cost alone. It doesn't take into account the cost of retrieval or the WAN bandwidth to store and retrieve the data. Further, cloud storage sacrifices genetic diversity: Most clouds store information on just one medium. Also, the cloud's concentration of information makes it an easy target for attack. This raises the question: At what point does cloud storage become false economy? BAYCOR CEO George Baker is among the country's most respected advocates for the protection and preservation of critical national infrastructures. Baker comments, "With the proliferation of cloud computing, *more data* [emphasis added] is being placed in *fewer baskets*, and that reliance on failover sites has *reduced physical security*."[3]

Today the availability and ease of IT and online services have changed the world socially, politically and economically: It's an exciting new frontier that few of us

---

[1] "Cloud Data Center Trends To Watch For In 2015,"*Forbes,* http://www.forbes.com/sites/huawei/2015/04/16/cloud-data-center-trends-to-watch-for-in-2015/

[2] "IDG Enterprise Cloud Computing Study 2014," IDG Enterprise, http://www.idgenterprise.com/report/idg-enterprise-cloud-computing-study-2014

[3] "Experts: Electromagnetic Interference Threat to Uptime is Real," Data Center Knowledge, http://www.datacenterknowledge.com/archives/2014/05/01/electromagnetic-interference-threat-uptime-real/

would willingly reject. At the same time, we have also opened a Pandora's Box of potential pitfalls, problems and catastrophes, of which most of us remain unaware. Here are a few of the more prominent, present hazards faced both by data centers and end-users:

- Despite its convenience, cloud computing deposits larger amounts of information in fewer receptacles, which compromises physical security. The decreased geographic diversity of secure storage locations imperils information integrity and survival.
- As cost-per-gigabyte plummets, data centers increasingly rely on hard drives for storage. The more dependence on a single type of storage media, the greater the likelihood of information loss.
- Sophisticated IT has become readily available to criminals internationally. Armed with little more than laptops and connectivity, players from places such as Belarus, Pakistan, China, Russia and North Korea – as well as the U.S. and Europe – pose immediate, potentially deadly threats. And our IT homogeneity makes us vulnerable targets.

THE GAMBLE AND THE STAKES

A recent study by IDC[4] reveals that 42 percent of data in the U.S. that requires security has none (Figure 3). On a worldwide basis, the number jumps to 47 percent. The question is not if – but when – a data catastrophe will occur.

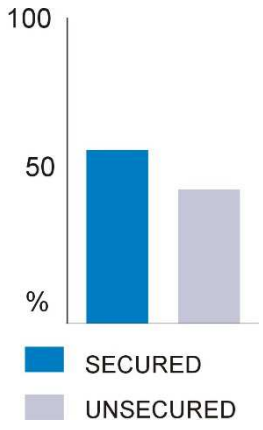


Figure 3. Status of sensitive IT data in the U.S.

What is at stake? We depend on IT and connectivity for a huge range of vital services, most of which we take for granted. They include, among others:

- Communications
- Energy
- The power grid
- Air and other transportation
- National defense
- Commerce and finance
- Distribution and inventory systems

[4] "The Digital Universe in 2020 . . . ," International Data Corporation, http://www.emc.com/collateral/analyst-reports/idc-the-digital-universe-in-2020.pdf

- Health and medical systems

In order to function, all such entities depend on access to data; the cost of data loss is staggering. A recent study by the Aberdeen Group[5] places the average cost of downtime due to data loss at more than $163,000 per hour; larger companies report hourly costs of more than $680,000. Keep in mind that these companies would eventually regain their data access. Can we even calculate the value of information that is permanently, irretrievably lost?

Our stewardship of data and information determines our future. If our cache of essential data is destroyed, rebuilding would prove impossible. And that is why we must always maintain a secure, incorruptible archive of information and knowledge.

THERE ARE COMPANIES WHO STILL UNDERSTAND THE IMPORTANCE OF GENETIC DIVERSITY IN STORAGE

Google is viewed as one of the most successful and cutting edge companies in the computing world. In 2011, a small portion of Gmail users logged into their email accounts only to find them empty. Google explained in a letter to users, "…in some rare instances software bugs can affect several copies of the data. That's what happened here. Some copies of mail were deleted… To protect your information from these unusual bugs, we also back it up to tape. Since the tapes are offline, they're protected from such software bugs." Because Google had the information on non-disk storage, they were successful in restoring 100% of the deleted data.*

Oddly enough, many of the blogs and press around the Gmail incident focused on disbelief that one of the most well-known names in compute was using tape. While some may feel that certain technologies and/or methods are "out of style," assuring business continuance for your organization and 100% recoverability of data for your users is never out of style.

At Google, a simple storage software update introduced the unexpected bug which caused the Gmail problem. Obviously, it remains critically important to protect data from traditional threats such as user error, software bugs and natural disaster. Unfortunately, a greater threat faces data centers and all those who depend on digital information for their organizations and even day-to-day lives. That's where genetic diversity in storage plays an even greater role.[6]

---

[5] "Downtime and Data Loss: How Much Can You Afford?" Aberdeen Group, http://v1.aberdeen.com/launch/report/perspective/8623-AI-downtime-disaster-recovery.asp

6 * Source: http://gmailblog.blogspot.com/2011/02/gmail-back-soon-for-everyone.html)

PART TWO: THE AGE OF ATTACK

In May 2015 former CIA Acting Director Michael Morrell said the "great war" against all types of terrorists would continue "for as far as I can see."[7] A few days later, Homeland Security Secretary Jeh Johnson observed that the U.S. had entered a new environment in which, "Because of the use of the Internet, we could have little or no notice in advance of an independent actor attempting to strike."[8]

So it was that, within the space of a few days, two of the most informed, credible terrorism experts in the U.S. delivered chilling remarks concerning prospects for the future of national security. Within 48 hours, however, their stories had dropped from the news cycle and so, presumably, from public attention.

Those of us who are entrusted with safeguarding the information used to run our organizations remain keenly aware of the challenges we face. In the past, preparation for and recovery from information-related disasters typically focused on natural events, such as fire, flood and human error. Now *intentional* attacks rank well at the top of our disaster survival and rebuilding plans.

Consider some of the more notable, recent attacks on U.S. and other western interests. The perpetrators include nation-states, independent actors, criminal syndicates and cyber terrorists of unknown origin. In at least one instance (that of altering hard drive firmware with timed malware), the National Security Agency itself may be the culprit. Extortion, data destruction, hacking, intimidation and espionage number among the crimes. A summary of a few recent, destructive – and probable – cyber crimes informs our view of threats to IT today.

INFECTED HARD DRIVE FIRMWARE

We don't know how many infected hard drives exist, where they are or when they may activate.

In early 2015 Reuters reported, "The U.S. National Security Agency has figured out how to hide spying software deep within hard drives made by Western Digital, Seagate . . . and other top manufacturers, giving the agency the means to eavesdrop on the majority of the world's computers . . ." [9]

Some of the malware dates to the year 2001 and may be scheduled to activate 15 years or more in the future; the NSA declined comment. The

---

[7] "CIA veteran Morell: ISIS' next test could be a 9/11-style attack," *USA Today,* http://www.usatoday.com/story/news/politics/2015/05/10/michael-morell-cia-the-great-war/27063655/

[8] "DHS secretary: Lone wolf attackers could 'strike at any moment'," *The Hill,* http://thehill.com/homenews/sunday-talk-shows/241562-dhs-secretary-lone-wolf-attackers-could-strike-at-any-moment

[9] "Russian researchers expose breakthrough U.S. spying program," Reuters, http://www.reuters.com/article/2015/02/16/us-usa-cyberspying-idUSKBN0LK1QV20150216

malware has been found in at least 30 countries. Because the malware is firmware-based, it continues to re-infect even after a drive has been wiped completely and reformatted.

Implications are profound, as this is the first known instance of malware-infected firmware in hard drives. Even the simplest firmware contains thousands of lines of code. If attackers target firmware, the few lines of code it would take stand little chance of ever being discovered. The very real possibility of embedded malware set to activate at some unknown future time is considered one of the most imposing threats to data storage as we know it today. Literally every hard drive in use could be wiped out simultaneously – including those storing archives, backups or disaster recovery information.

SONY: SYSTEMATIC DATA DESTRUCTION

Initial reports on the 2014 North Korean cyber attack against Sony Pictures Entertainment focused mainly on the ways in which the studio had been compromised and embarrassed by leaks of their internal documents and new movies. However, as noted in the *Financial Times* and other publications, "large quantities of [Sony's] data were systematically destroyed."[10] "Mountains of documents had been stolen, internal data centers had been wiped clean, and 75 percent of the servers had been destroyed."[11]

Among the most destructive attacks launched against a corporate entity

". . . Tenable Security, a US-based cyber security company, claims . . . this shows hackers are moving from stealing data to destroying it." Tenable's CEO states, "I really believe the people doing these attacks will move from exfiltration to pure destruction of data."[12]

Sony represents the first known case of intentional data destruction by hackers working remotely. In this high-profile case, among the largest so far against a corporate entity, perpetrators destroyed sole master copies of feature films, eradicated business records and leaked personal correspondence to the public.

As is the case with most organizations that experience a cyber breach, Sony initially downplayed the attack, though it rates among the largest and most effective ever mounted against a corporate entity.

---

[10] "Sony cyber attack reveals hackers changing their stripes," *Financial Times*
http://www.ft.com/cms/s/0/1c967b94-7c0d-11e4-a7b8-00144feabdc0.html

[11] "Sony Cyberattack, First a Nuisance, Swiftly Grew Into a Firestorm," *The New York Times,*
http://www.nytimes.com/2014/12/31/business/media/sony-attack-first-a-nuisance-swiftly-grew-into-a-firestorm-.html

[12] "Sony cyber attack reveals hackers changing their stripes," *Financial Times*
http://www.ft.com/cms/s/0/1c967b94-7c0d-11e4-a7b8-00144feabdc0.html

Saudi Aramco and RasGas: Compromising world energy suppliers

In 2012 Saudi Aramco, the world's largest oil exporter and key player in global energy markets, experienced a cyberattack of tremendous scale and impact. A self-replicating virus (known as Shamoon or Disttrack) infected and shut down at least 30,000 of its Windows-based PCs. It appears ". . . likely that some drilling and production data were lost. . ." [13] The virus caused random data deletion on hard drives and impacted the company's business processes.

Researchers at Symantec observed that Shamoon corrupts files, overwrites the master boot record and replaces data with image files.

A similar attack occurred a few days later at RasGas, a Qatar liquid-petroleum gas supplier. According to BBC reports, the company was forced to take down its website and shut down corporate email systems, though RasGas claims no adverse effect on its production.

Disrupted energy supplies often lead to global economic downturn, as well as political instability. Though both RasGas and Saudi Aramco deny the severity of the cyber attacks they suffered, analysts believe the attacks were substantial, resulting in serious damage and slow recovery. Also, "with the increasing computerisation of critical infrastructure services, the energy and utility industries have never been more vulnerable to cyber attacks" [14]

Intentional Electromagnetic Interference (IEMI)

Build your own IEMI weapon for as little as $200

IEMI is the "intentional malicious generation of electromagnetic energy introducing noise or signals into electric and electronic systems thus disrupting, confusing, or damaging these systems for terrorist or criminal purposes." [15] Weaponized IEMI may be delivered in a variety of forms and sizes. Electromagnetic pulse (EMP) is a type of IEMI.

---

[13] "The Cyber Attack on Saudi Aramco," International Institute for Strategic Studies, https://www.iiss.org/en/publications/survival/sections/2013-94b0/survival--global-politics-and-strategy-april-may-2013-b2cc/55-2-08-bronk-and-tikk-ringas-e272

[14] Ibid.

[15] *"Electromagnetic compatibility (EMC) - Part 2-13 . . . ,"* *International Electrotechnical Commission,* https://webstore.iec.ch/publication/4131

Improvised IEMI devices can be built from instructions and materials available online. Prices range from about $200 to $2500, depending on weapon size and capabilities. "[Small] mobile devices [in a pickup truck] . . . can damage or destroy electronics circuitry up to 600 feet away, and corrupt and disrupt the data of commerce over two miles away, then drive away undetected."[16] Depending on its power and accuracy, an IEMI weapon can destroy entire power grids; or all the electronics within a targeted region, campus or building; or just the smartphone on your desk.

IEMI mitigation and data survival

New, purpose-built data centers shield against IEMI by using a combination of metals within their inner and outer walls.[17] Some data centers use IEMI-resistant enclosures, such as Faraday cages. However, the vast majority of U.S. data and electronics remains vulnerable to IEMI attack. Different media types respond differently to IEMI. For instance, information on a hard drive would be eradicated, but data on tape or optical media would survive.

RANSOMWARE

To decrypt your data, just pay the ransom – and hope for the best.

As an executive at a mid-sized American business, you enter your office at 10:00 A.M. to review some new Excel worksheets, but they cannot be found. Neither can you access any files of any type -- on your system or on the company network. Everyone else at your firm has the same problem: Your company's data has been ransomed– encrypted by crime-syndicate hackers. Because it sets keys in the Windows registry, the crypto-ransomware restarts every time an infected computer boots.

By noon, a ransom note appears on your screen. To regain your data, the message states, you must deposit a specified amount of money into a bitcoin account or other, untraceable digital wallet. Should you fail to comply, your company's information will remain encrypted and unusable. Additionally, you may also need to replace all the affected hardware – or, at least, the hard drives.

CryptoLocker, which first appeared in 2013, is the best known and most powerful of ransomware programs. Sophos, a developer of cyber security hardware and software, notes, "Sadly, there's not much you can do to get your

---

[16] "Intentional Electromagnetic Interference . . . ," *Mission Critical*, http://www.missioncriticalmagazine.com/ext/resources/MC/Home/Files/PDFs/WP_IntentionalElectromagneticInterference.pdf

[17] "New data center protects against solar storms and nuclear EMPs," Computerworld, http://www.computerworld.com/article/2606378/new-data-center-protects-against-solar-storms-and-nuclear-emps.html

files back except to pay the ransom – the encryption is too strong to crack."[18] While software exists to *remove* malware associated with CryptoLocker, as of yet there is no way to *decode* CryptoLocker-encrypted files – other than paying the hackers and hoping for the best. Often, the culprits fail to unlock the data after the ransom is paid or fail to even follow up allowing users to pay the ransom.

GENETIC DIVERSITY: COMMON-SENSE RESPONSE…

As demonstrated in the previous examples of contemporary cyber attacks, our data centers face immediate, grave challenges which, in power and scope, are unlike anything we have encountered in the past. We can, however, still win cyber battles – and even win the greater war.

Tape is great because it's not disk

Take a closer look at prevailing, popular technologies, many of which we accept as a matter of course. For example, both in the cloud and the data center, it may seem most quick, efficient and expedient to depend on hard drive-based technology almost exclusively. In reality, however, it is vital that we diversify the media upon which we store essential, archival information. It is rare that any of the attacks mentioned above would be able to destroy or lock information on both disk *and* tape mediums – leave one or the other as refuge.

As Google's Raymond Blum recently stated in discussing Google's use of tape, "… tape is great because it's not disk." Together with prudent information-management strategies, an investment in varied, robust media and technologies holds our best hope for information survival and rebuilding. In fact, most of the cyber attacks cited in this paper could be reduced or eliminated through just such an approach.

Even with no knowledge of genetics at the time, Queen Victoria was wise enough to theorize the dangers of limited diversity. She wrote to her daughter Vicky, "I do wish one could find some more black eyed Princes and Princesses for our children! For that constant fair hair and blue eyes makes the blood so lymphatic… it is not as trivial as you may think, for darling Papa—often with vehemence said: 'We must have some strong blood.'" [19]

---

[18] "Anatomy of a ransomware attack . . . ," Sophos Ltd., https://blogs.sophos.com/2015/03/03/anatomy-of-a-ransomware-attack-cryptolocker-cryptowall-and-how-to-stay-safe-infographic/

19 *Source: http://sciencecases.lib.buffalo.edu/cs/files/hemo.pdf

In nature, the survival of the fittest stems from hybrid vigor and genetic diversity – the same elements we must incorporate into our data centers. We need to reconsider and rethink the ways we equip, organize, safeguard and use data centers and the information they store.

TO LEARN MORE

If you would like to learn more about how to keep your information safe from failure, please visit Spectralogic.com. We'll show you how a few easy, cost-effective measures can bring more strength and security to your organization. We look forward to hearing from you.

Deep Storage Experts

Spectra Logic develops deep storage solutions that solve the problem of long term storage for business and technology professionals dealing with exponential data growth.

Dedicated solely to storage innovation for more than 35 years, Spectra Logic's uncompromising product and customer focus is proven by the largest information users in multiple vertical markets globally.

Spectra enables affordable, multi-decade data storage and access by creating new methods of managing information in all forms of deep storage—including archive, backup, cold storage, cloud, and private cloud.

For more information, please visit http://www.spectralogic.com.

SPECTRA.

DEEP STORAGE EXPERTS

303-449-6400 • 800-833-1132 • (Fax) 303-939-8844 • 6285 Lookout Road • Boulder, CO 80301 USA • spectralogic.com