VOLUME 1, 2018

ISACA® Journal

THE FUTURE OF DATA PROTECTION

+

CLOUDIFYING THREATS—UNDERSTANDING CLOUD APP ATTACKS AND DEFENSES

THE MACHINE LEARNING AUDIT—CRISP-DM FRAMEWORK

APPLYING AI IN APPLICATION SECURITY

ISACA.ORG

ISACA®

GET CERTIFIED.
GET AHEAD.

CISA® CRISC™ CISM® CGEIT®

SEE WHAT'S NEXT, NOW

No matter your role in IS/IT—audit, security, cyber security, risk, privacy or governance, these credentials are designed for forward-thinking professionals across a variety of industries. ISACA® certifications are not just any certification, they are the ones that can get you ahead!

2018 EXAM REGISTRATION NOW OPEN

Choose your certification and exam prep that best suits your needs—get started today!

www.isaca.org/GetCertified-Jv1

Register
EARLY
– to –
SAVE

6th Annual
European
Compliance & Ethics Institute

25–28 March 2018 | *Frankfurt, Germany*

SCCE

- Hear from top compliance & ethics professionals from Europe and around the world
- Learn the latest and best solutions for compliance & ethics challenges, including anti-corruption, data protection, and risk management
- Build your professional network
- Earn the continuing education units you need, and take the Certified Compliance & Ethics Professional - International (CCEP-I)® exam

EuropeanComplianceEthicsInstitute.org | *lizza.catalano@corporatecompliance.org*

ISACA Journal

Online-Exclusive Features

Do not miss out on the *Journal*'s online-exclusive content. With new content weekly through feature articles and blogs, the *Journal* is more than a static print publication. Use your unique member login credentials to access these articles at *www.isaca.org/journal*.

Online Features
The following is a sample of the upcoming features planned for January and February 2018.

Auditing Big Data in the Enterprise
Joshua McDermott, CISA, CEH, CISSP, PMP

Toward Encrypted and Private Databases
Josh Joy

When What Is Lost Is Lost Forever
William Emmanuel Yu, Ph.D., CRISC, CISM, CISSP, CSSLP

Discuss topics in the ISACA® Knowledge Center: *www.isaca.org/knowledgecenter*
Follow ISACA on Twitter: *http://twitter.com/isacanews*; Hashtag: #ISACA
Follow ISACA on LinkedIn: *www.linkedin.com/company/isaca*
Like ISACA on Facebook: *www.facebook.com/ISACAHQ*

The *ISACA® Journal* seeks to enhance the proficiency and competitive advantage of its international readership by providing managerial and technical guidance from experienced global authors. The *Journal's* noncommercial, peer-reviewed articles focus on topics critical to professionals involved in IT audit, governance, security and assurance.

Read more from these *Journal* authors...

Journal authors are now blogging at *www.isaca.org/journal/blog*. Visit the *ISACA Journal* blog, Practically Speaking, to gain practical knowledge from colleagues and to participate in the growing ISACA® community.

ISACA®
3701 Algonquin Road, Suite 1010
Rolling Meadows, Illinois 60008 USA
Telephone
+1.847.660.5505
Fax +1.847.253.1755
www.isaca.org

Managing Availability in the Multi-Modal Era

2018 marks the 20th year of Steven Ross' Information Security Matters column in the *ISACA® Journal*. This column has remained one of the most popular columns in the *ISACA Journal* through the years. ISACA® is deeply grateful to Steve for his time, his expertise and his talent as an author and a valued colleague. The *ISACA Journal* looks forward to many more years of successful collaboration with Steve.

Do you have something to say about this article?

Visit the *Journal* pages of the ISACA® website *(www.isaca.org/journal)*, find the article and click on the Comments link to share your thoughts.

http://bit.ly/2BSAoNN

A few columns back, I wrote about the security of multi-modal IT environments,[1] in which applications and infrastructure are operated in colocation (colo) sites, as Internet-based services; in the cloud, as managed services; and in proprietary data centers—all at the same time. In that article, I dropped a rather heavy hint that I would address disaster recovery in multi-modal environments and I will do that, but I would like to expand a bit on availability management first. Nothing has changed regarding multi-modal availability management. When something stops working, it has to be fixed and brought back up. Nothing has changed…except so much is different.

Finding Fault

Outages may be caused one of two[2] ways: by logical failure of data, software or infrastructure, or by physical disruption (i.e., a disaster) affecting equipment or networks. When a system or many systems go down, it is not always immediately evident where the cause lies. Thus, when systems flatline, the first task for IT operations personnel is simply to find out what happened and where. When all systems were operated in a single data center on an organization's own premises, this was relatively simple. With those systems in many locations, figuring out what is going wrong is considerably more difficult.

Organizations with more than one data center already face this problem, but at least those organizations own all the sites. Things get more complicated when multiple owners are involved. If, for an example, a department uses an Internet-based service, an outage might be traced to an organization's own data center, to its telecommunications carrier(s), to the servicer's data center(s) or to its carriers. In the immediate aftermath of an outage, it is often the case that the people responsible for each are also trying to figure out what is wrong. Until who is at fault can be established, each one blames all the others.

The Virtual Console

Managing availability in a multi-modal environment is challenged by the relative obscurity of all the components in that environment. This raises the importance of a virtual console[3] that enables visibility into all of an organization's systems, wherever they may be or regardless of who owns them. When one component fails, that might be a stand-alone event or it might have a knock-on effect on other components. For example,

Steven J. Ross, CISA, CISSP, MBCP
Is executive principal of Risk Masters International LLC. Ross has been writing one of the *Journal's* most popular columns since 1998. He can be reached at stross@riskmastersintl.com.

a cloud-based application might generate data that are used by another application running in a colo data center. Without a virtual console that enables an organization to see and to manage both simultaneously, there is a high likelihood of downstream problems. And, if both cannot be recovered to a common point in a concerted manner, those problems may be reflected in an overall loss of control over the data.

> "WHEN AN ORGANIZATION MOVES DATA AND SOFTWARE TO A COLO, IT LOSES CUSTODY BUT RETAINS AGENCY."

Or is "loss of control" the correct phrase? To a great extent, organizations that institute multi-modal architectures[4] have already lost a degree of control over information resources and, thus, over the availability of those resources. When an organization moves data and software to a colo, it loses custody but retains agency. In the case of managed services, it may keep custody, but lose the ability to initiate, execute and control its resources. Thus, the *availability* of data (and the recovery of access to the data when access is interrupted) must be distinguished from the *use* of the data and their recovery following an outage. Where an organization runs its own data center, it is responsible for the availability of both data and software in the event of a disruption. This may not be the case when some or all of custody and use are given to third parties. All of which is a convoluted way of saying that managing availability in a multi-modal environment is quite complicated.

Responsibility and Accountability

Note that I said *responsibility* for availability. Responsibility can be assigned, but *accountability* for availability remains with the owner of the relevant resources. In some ways, this is just another case of the ongoing discussion of control over outsourcing.[5] Simply put, the owner of a resource is accountable for its availability even if it has chosen to "hire" someone else to carry out the tasks involved. The distinction only seems to be an issue in organizations where the prevalent culture leads to retention of ownership along with evasion of responsibility.[6] When contemplating a multi-modal architecture, organizations must consider both maintaining the availability of information resources as well as which entities will have the access and the tools to provide availability.

This seems confounding only because we view the whole range of making information resources available—ownership, accountability, responsibility, access, recovery, security, operations, *et al*—from the perspective of IT-the-way-it-used-to-be. An apt analogy might be houses-the-way-they-used-to-be. A few centuries ago, as the pioneers spread across the plains,[7] if you wanted a house, you built it. You owned it and the land beneath it, too, by dint of the fact that you had built a house on it. If you wanted heat in your house, you chopped down some trees. If you wanted food, you grew it or killed it. If you wanted water, you dug a well. If you wanted your house to be there when you went away, well, you did not go away very often. Today, most of us have "outsourced" our heat, food and water. We may own our houses, or we may have occupancy, but not ownership of either the house or the land (i.e., rentals).

We understand that the responsibility and accountability for continued existence of houses are shared explicitly by the owner and the occupant, who may or may not be the same. The lines of demarcation are established in contracts and laws. The same is true of IT in a multi-modal environment. What, after all, is a service level agreement (SLA)

in an IT contract but a commitment by one party to make information resources available and by the other to accept limited periods in which they are not? Managing availability in a multi-modal environment requires a great deal of attention to details, which are being defined by the multi-modal pioneers of our day. Perhaps we are all pioneers now, but we will become settlers someday.

Endnotes

1 Ross, S.; "Information Security in the Multi-Modal Era," *ISACA® Journal*, vol. 5, 2017, *www.isaca.org/Journal/archives/Pages/default.aspx*

2 Actually, there are three if you include downtime caused by cyberattacks, yet another broad hint for future consideration.

3 *Op cit,* Ross. In my research for the previous article, I found that Nintendo uses this term for some of its gaming products. Obviously, I do not mean the term that way and will no longer make excuses for using it. However, I now find that IBM uses the same term for some of its AIX systems, and it is also used with regard to some Unix-based operating systems. I wish there were a more appropriate term for a single console that provides visibility into numerous unintegrated platforms and networks, but I cannot think of one. Suggestions are welcome, but until someone comes up with something better, I am sticking with "virtual console."

4 I have been referring to multi-modal *environments* and here use the word *architectures*. They are not the same thing, with the former implying operations and the latter design. To the extent that organizations operate by design—and there are many exceptions—I think the terms can be used interchangeably.

5 Ennals, R.; *Executive Guide to Preventing Information Technology Disasters*, Springer Verlag, London, 1995. Gouge, I.; *Shaping the IT Organization—The Impact of Outsourcing and the New Business Model*, Springer Verlag, London, 2003. The literature on responsibility and accountability with regard to IT outsourcing is voluminous. These are two enlightening examples.

6 *Ibid.*, p. 97

7 Very American, I know, but the image is a good one.

Backup and Recovery

Do you have something to say about this article?

Visit the *Journal* pages of the ISACA® website *(www.isaca.org/journal)*, find the article and click on the Comments link to share your thoughts.

http://bit.ly/2k6Vlwr

When I sat for (and passed!) my Certified Information Systems Auditor® (CISA®) examination back in 2005, one of the key task statements was "Evaluate the adequacy of backup and recovery provisions to ensure the resumption of normal information processing in the event of a short-term disruption and/or the need to rerun or restart a process."[1] By the time I came to update the *CISA® Review Manual* for the 2016 job practices, an equivalent task read "Evaluate IT continuity and resilience (backups/restores, disaster recovery plan [DRP]) to determine whether they are controlled effectively and continue to support the organization's objectives."[2] Although the wording has changed, the message from ISACA® is clear—backup and recovery are still key controls.

I find this interesting since, due to technological improvements—most notably virtualization and the cloud—IT auditors often receive pushback when seeking assurance on IT continuity and resilience. So, how should an IT auditor respond when presented with these newer technological solutions? In this column, Tommie Singleton previously advocated principles for backup and recovery.[3] I believe it is worth reviewing these principles, considering the changes in technology.

Ian Cooke, CISA, CGEIT, CRISC, COBIT Assessor and Implementer, CFE, CPTE, DipFM, ITIL Foundation, Six Sigma Green Belt
Is the group IT audit manager with An Post (the Irish Post Office based in Dublin, Ireland) and has 30 years of experience in all aspects of information systems. Cooke has served on several ISACA® committees and is a current member of ISACA's CGEIT® Exam Item Development Working Group. He is the community leader for the Oracle Databases, SQL Server Databases, and Audit Tools and Techniques discussions in the ISACA Knowledge Center. Cooke supported the update of the *CISA Review Manual* for the 2016 job practices and was a subject matter expert for ISACA's CISA and CRISC Online Review Courses. He is the recipient of the 2017 John W. Lainhart IV Common Body of Knowledge Award for contributions to the development and enhancement of ISACA publications and certification training modules. He welcomes comments or suggestions for articles via email (Ian_J_Cooke@hotmail.com), Twitter (@COOKEI), or on the Audit Tools and Techniques topic in the ISACA Knowledge Center. Opinions expressed are his own and do not necessarily represent the views of An Post.

Backup Principle 1—Perform Regular Backups

The principle for regular data backups is to back up data daily. That backup could be to media (e.g., tape or external hard drive), or it could be to a remote location via the cloud (i.e., the Internet). If an enterprise is backing up to media, this principle recommends that backups be conducted to a different media for end-of-week and end-of-month backups (this daily, weekly and monthly set of backups is known as "grandfather-father-son").[4]

I think we can all agree that with virtual machine (VM) replication, the first sentence in this section should now end with "at least daily." Indeed, with VM replication you can create an exact copy of the VM on a spare host and keep this copy in sync with the original VM. So, the principle of (at least) daily backups still stands.

However, replication alone can, in certain circumstances, increase the risk. If the VM or the data is somehow corrupted, this corruption will automatically be replicated to the other machine(s). Therefore, it makes perfect sense to maintain separate VM backups. Since we cannot be sure when a data corruption might be discovered, it also makes sense to maintain grandfather-father-son backups.

A word of caution: The administrator of the VM backups should not have Internet access since this would expose the backups to ransomware. All such users should have a second, unprivileged account that does not have access to the VMs, but does have access to the Internet for troubleshooting purposes.

Backup Principle 2—Test Backup Process Reliability

The next concern is whether the backup process is reliable. Therefore, upon using a new backup methodology or technology, management should provide a means to test the data afterward to ensure that the process is actually recording all of the data onto the target backup device.[5]

It is a widespread practice to back up entire VMs, but it is possible to exclude data to reduce the size of the VM backup or replica and decrease the load on the network. This principle, therefore, still stands, as an IT auditor should still validate what is being backed up. Furthermore, as there is a risk of corruption during the backup process, an IT auditor should ensure that a health check is periodically performed. This typically means scheduling a cyclic redundancy check (CRC) for metadata and a hash check for VM data blocks in the backup file to verify their integrity. The health check provides assurance that the restore point is consistent and that it will be possible to restore the data.

Backup Principle 3—Use Secure Storage

Another concern is where the backup is stored. If it is stored onsite, and if the entity suffers a pandemic event such as a fire, the event would destroy the operational data and the backup data. Thus, the backup principle for storage is to use a location that is at a safe distance from the entity's location. The cloud automatically provides this element.[6]

This principle, obviously, still stands and, yes, the cloud automatically provides this element. However, there is also an element of risk when backing up to the cloud—sensitive enterprise data may be accessible by the cloud provider and/or other third parties who share the cloud. It is, therefore, vital to ensure that VMs backed up to the cloud are encrypted. The type of encryption, key management procedures, etc., should all, of course, be verified.[7]

Backup Principle 4—Perform Test Restores

Additionally, management should provide a test for restoring the backup at least once a year. That test should be documented, even if it is just a screenshot showing the data restored.[8]

Ah, the test restore! For some reason, many IT departments just do not see the value. Anyone who is an IT auditor has heard one, if not several, variations of "Test restores are not needed because we use clustering/have a failover solution/use VM replication/use storage area network (SAN) replication/use log shipping and, therefore, we just would not restore that way."

> "IT MAKES SENSE TO PERFORM TEST RESTORES TO ENSURE THAT THE CORRECT DATA ARE BEING BACKED UP, THAT THE DATA ARE, IN FACT, RESTORABLE AND THAT THE ENTERPRISE KNOWS HOW TO RESTORE IT."

However, it has already been noted that replication can increase the risk of data loss if data are corrupted and/or sabotaged. Depending on the configuration, the same can be said for each of the above solutions. Regardless, restoring from backup simply provides an additional option in a real disaster scenario. Therefore, it makes sense to perform test restores to ensure that the correct data are being backed up, that the data are, in fact, restorable and that the enterprise knows how to restore it. This principle still stands.

Recovery Principle 1—Identify and Rank Critical Applications

The principles of developing a business continuity plan/disaster recovery plan (BCP/DRP)[9] include a step to identify the critical applications and rank them in importance of operations. This list becomes strategically valuable if ever needed in providing

the recovery team with a blueprint of how to restore application software.[10]

In a previous column, I advocated categorizing applications in terms of confidentiality, integrity and availability.[11] The suggested availability categorization provides the list of ranked critical applications and is required regardless of the technology used to restore the applications. Therefore, this principle still stands.

> "THE HEART OF A BCP/DRP IS TO PROVIDE A BACKUP MEANS OF PROVIDING THE ESSENTIAL COMPONENTS OF COMPUTER OPERATIONS."

However, the application list should be adjusted to consider application interfaces and data flows. For example, application A may be more business critical than applications B and C, but may require data from application C, which should, therefore, be restored first. Also, multiple applications may share a single VM. If applications do share a VM, an enterprise may be restoring a lower-priority application alongside a high-priority application.

Recovery Principle 2—Create a Recovery Team With Roles and Responsibilities

Another principle, and obvious need, is to create a recovery team. The team should include all the functions and roles necessary to quickly and completely restore computer operations. There should be a document that identifies the team members, their respective roles and the steps each would take in restoring operations.[12]

A team is still required to perform the actual recovery, so this principle still stands. My concern with virtual and/or cloud technologies is that the recovery is left to the IT operations team responsible for the VMs. I believe it is still vital to confirm the recovery point, that all expected transactions are present and that all interfaces are running correctly—all of which still require the input of application experts and business users.

Recovery Principle 3—Provide a Backup for All Essential Components of Computer Operations

The heart of a BCP/DRP is to provide a backup means of providing the essential components of computer operations. The site should include a building, electricity, furniture and other basic needs for housing the computer operations. Typically, the site follows the same principle as storage of backup data in that it is located a safe distance from the entity's facility, but not too far to reach in a timely manner if it is necessary to recover operations.[13]

Regardless of the technology, this principle still stands. The cloud may negate the need for an enterprise to have its own physical secondary data center, but there must still be a location for staff to access the restored services. For example, this article is stored on a Microsoft OneDrive and I am accessing it from my office at home. If my home and laptop were destroyed (perish the thought!), I would need a new location from which to access the article.

Recovery Principle 4—Provide for Regular and Effective Testing of the Plan

Principles of backup and recovery suggest that the most important step is to provide a full test of the

BCP/DRP at some regular interval to ensure that it actually works and to improve the plan to be more efficient and effective.[14]

There is no doubt that VMs and the cloud can largely automate application recovery while enhancing both recovery times and recovery points. However, in a disaster situation there is still very much a need for human actions and operations. This is the very definition of a process[15] and, where processes exist, there is a need to confirm that they produce the desired results. The only way to do this is to regularly test and document the results of the BCP/DRP. Therefore, this principle very much still stands. Further, regardless of the technology in use, the BCP/DRP can always be enhanced. Testing allows for this while also allowing the IT auditor to provide assurance.

Conclusion

This column could well have been titled "What Every IT Auditor Should *Still* Know About Backup and Recovery." Innovative technologies such as VMs and the cloud help the efficiency and effectiveness of backup and recovery plans, but they do not replace the need to plan, document, or test and test again. The technologies still rely heavily on human intervention and, while that is the case, assurance can be provided only by applying good principles to backup and recovery.

Endnotes

1 ISACA, *CISA Review Manual 2005*, USA, 2004
2 ISACA, *CISA Review Manual 26th Edition*, USA, 2016
3 Singleton, T.; "What Every IT Auditor Should Know About Backup and Recovery," *ISACA® Journal*, vol. 6, 2011, *www.isaca.org/Journal/archives/Pages/default.aspx*
4 *Ibid*.
5 *Ibid*.
6 *Ibid*.
7 ISACA, *Assessing Cryptographic Systems*, USA, 2017, *www.isaca.org/Knowledge-Center/Research/Documents/Assessing-Cryptographic-Systems_res_eng_0817.pdf*
8 *Op cit,* Singleton
9 Business continuity plans and disaster recovery plans are different and separate processes, but in this article, they will be referred to as one unit.
10 *Op cit,* Singleton
11 Cooke, I.; "Doing More With Less," *ISACA Journal*, vol. 5, 2017, *www.isaca.org/Journal/archives/Pages/default.aspx*
12 *Op cit,* Singleton
13 *Ibid*.
14 *Ibid*.
15 Merriam Webster defines "process" as "(a) series of actions or operations conducing to an end." *www.merriam-webster.com/dictionary/process*

Building Tomorrow's Leaders, Today

Sandy Fadale, CRISC, CISM, CGEIT
Is a senior security consultant with Mariner Security Solutions. She has more than 25 years of in-depth IT experience in the field of enterprise computing with an emphasis on information security, which includes IT security, application development and business continuity. Prior to Mariner Security Solutions, Fadale was a senior manager at Bell Aliant, a manager with Ernst & Young LLP, and a manager with Visteon Corporation in its information security and risk advisory practices. She also served in the US military in telecommunications, utilizing various encryption techniques. She has been the president of the ISACA® Atlantic Provinces Chapter since 2008. Fadale teaches Certified in Risk and Information Systems Control™ (CRISC™), Certified Information Security Manager® (CISM®), and Certified in the Governance of Enterprise IT® (CGEIT®) certification review courses and is a subject matter expert who has assisted in the creation of the 2012, 2013 and 2014 editions of the ISACA *CRISC® Review Manual*.

Q: How do you think the role of the information security professional is changing or has changed?

A: It has long been said that employees are the biggest threat, but I do not believe that has ever been so true as it is now. I believe information risk management is one of the most important skills a security professional needs to possess today to provide value. From small start-ups to large organizations, information assets are leaving our organizations and the ability to understand where they are and how to secure them is extremely challenging.

Gone are the days we would issue edicts (back in the late 1980s and '90s). We are now advisors and we need to understand governance and compliance, privacy, metrics and data analytics, as well as business consulting skills.

Protecting information, no matter where it is located, requires a different way of thinking. Information security professionals who are used to concentrating on technology need to change their focus to business processes and data. Cloud computing and mobile devices are controlling this evolution; they are requiring that security professionals spend more time on governance and providing advice to organizations than on direct operational responsibilities for cloud and mobile environments.

Q: What leadership skills do you feel are critical for professionals to be successful in the field of information security?

A: I believe analytical skills are critical for information security professionals. Analytical skills refer to the ability to collect and analyze information, solve problems, and make decisions. These strengths can help increase and benefit an organization's productivity.

Employers look for employees who use clear, logical steps and excellent judgment to understand an issue from all angles before executing an action.

There are five skills that fall under analytical skills that I believe are important to master: communication, creativity, critical thinking, data analysis and research capabilities.

Q: What is the best way for someone to develop those skills?

A: Having a mentor to teach you how to fine-tune analytical skills is really helpful. Do not feel as though you have ever mastered this area. Always continue to refine it.

But how do you go about selecting a mentor? Find someone to emulate, and study that person. Then ask him/her to be a mentor. Ensure that the person understands the area in which you want to grow and why you chose him/her. Continuously evaluate your growth and your mentored relationship. It may be awkward at first, but let the relationship develop organically. As the relationship grows and you are being challenged, do not run away.

Q: What advice do you have for information security professionals as they plan their career paths and look at the future of information security?

A: This is not a cut-and-dried topic. It is extremely complex and usually depends on a combination of technical skills, nontechnical skills and personal interest. I do not believe that just anyone can be a technical security professional. One must understand networking, route switching, common hacking techniques plus many other areas. I have built my career on the people, process, governance and risk management side of the house. So, those starting a career must really understand which path they want to take: the super-cool "How do hackers think?" technical side or the governance, risk and compliance side.

How people gain knowledge is as individual as the individual. For example, you can go to university and specialize. Once university is completed and you are in a junior role, you can then pursue certification (and there are many of those). That said, however, just having a certification

and textbook knowledge does not automatically make an effective security professional. I worry that recruiters and human resources (HR) departments often rely too heavily on paper-based qualifications, given the pressing need to fill open positions.

Information security as a career choice is hugely rewarding. Regardless of the discipline chosen, security requires life-long learning and constant change. Security professionals never grow bored.

Q: What do you think are the most effective ways to address the cyber security skills gap and, especially, the lack of women in the cyber security workspace?

A: Part of the problem stems from the lack of information about careers in information security. This issue traces its roots all the way down to parents and school counselors not knowing about the full range of opportunities or, at best, reducing the field to looking only to the specialty of hacking. Often, parents and students are told that the only way to follow a security career path is to go through a traditional computer science program or a networking program, then switch into security. This may have been the reality 10 years ago, but it is no longer the case: An increasing number of schools are offering graduate and even undergraduate courses feeding directly into information security careers.

Organizations need to retain employees as the opportunities for switching jobs for more money are numerous in this field. Retention techniques include providing opportunities for continuing education and professional development and setting a clear path for development.

Information security professionals can work together to begin to make a difference in the cyber security talent shortage. Efforts can range from retraining existing employees, to recruiting high schoolers into specific educational pathways, to bringing together the government sector, private sector and academia to share amazing opportunities for employment and growth in the ever-expanding field of cyber security.

Q: You served in the US military. How did that experience shape your professional experience as a civilian?

A: When I was in the military in the late '70s, I got a taste of security as I set up encrypted communications channels using 16-bit encryption, and I knew this was the career for me. Security was not a "thing" after I got out of the military as a disabled veteran in 1980. However, in 1986, I went back to school and graduated with a computer science degree and then a security administrator position happened to come up in 1989 and I took it. The rest is history.

I think I was shaped by the military in my discipline and respect for people. I was raised that your word is everything and a handshake is binding, and the military reinforced that.

Q: What has been your biggest workplace or career challenge and how did you face it?

Working in a truly global organization several years ago, I was faced with understanding privacy laws, security laws and regulations around the globe. Security policies could not be written at a parent company and pushed out to its affiliates because countries' laws differed.

We had an incident in Italy that we needed to investigate. During the investigation, I received a call from our legal department asking what I was doing, so I told them. I was promptly advised that Italy has some of the strictest privacy laws in the world and I should not be able to see what I had seen. We ended up building a small data center there so information would remain in-country.

I had to update the policies to better reflect the laws around the world.

1 What is the biggest security challenge that will be faced in 2018?

Continued cyberattacks from Russia and China

2 What are your three goals for 2018?

- Obtain my Certified Information Systems Auditor® (CISA®) certification
- Build Mariner Security Services (MSS) Training product line into a well-respected training offering in Atlantic Canada
- Help grow the MSS business

3 What is your favorite blog?

Krebs on Security @briankrebs

4 What is on your desk right now?

Coffee (decaffeinated), two monitors, wireless keyboard, wireless mouse and mouse pad, a light and my client notes notebook

5 Who are you following on Twitter?

Green Party Canada, Digital Forensics, Peter Morin, Paul Jauregui, TrendLabs, Think Progress, Kaseya, The IoT, Dark Reading, among others

6 How has social media impacted you professionally?

It allows me to have the latest trends at my fingertips.

7 What is your number-one piece of advice for other information security professionals, especially women?

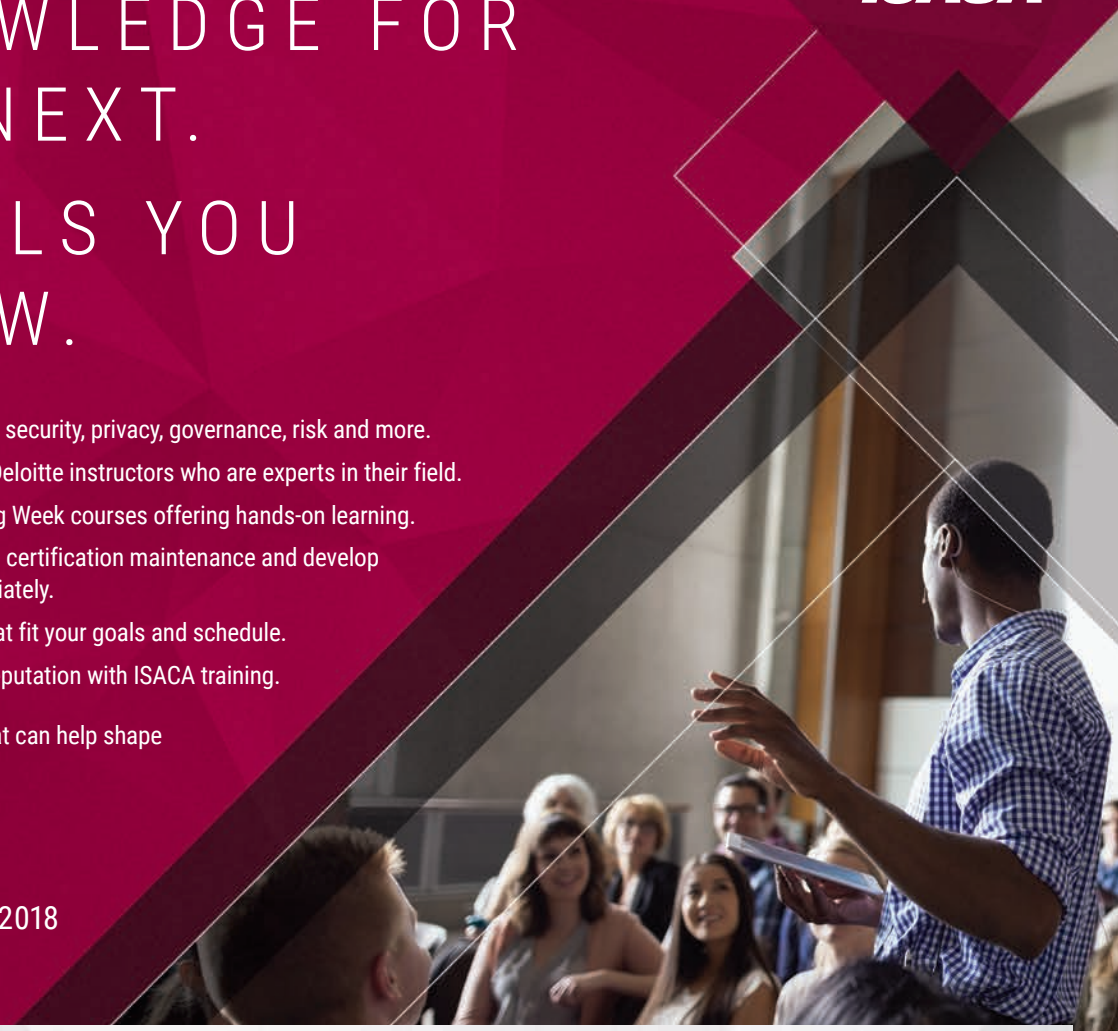
Be confident in what you know. There is a fine line between assertion and aggression—do not cross it.

8 What do you do when you are not at work?

Spend time with my wife outdoors, camping, playing Pickleball and practicing yoga

ISACA®

THE KNOWLEDGE FOR WHAT'S NEXT. THE SKILLS YOU NEED NOW.



- Dive deep into IS audit, security, cyber security, privacy, governance, risk and more.
- Interact with experienced ISACA® or Deloitte instructors who are experts in their field.
- Save time with focused, 4-day Training Week courses offering hands-on learning.
- Earn up to 32 CPEs per course toward certification maintenance and develop real-world skills you can apply immediately.
- Choose the Training Week courses that fit your goals and schedule.
- Build your expertise and boost your reputation with ISACA training.

Develop career-enhancing expertise that can help shape your future role.

SEE WHAT'S NEXT, NOW

REGISTER TODAY AT
WWW.ISACA.ORG/TRAINING2018

PREPARE FOR YOUR NEXT ROLE, NOW.

Gain new tools and techniques as you advance or refresh your knowledge.

ISACA TRAINING COURSES

TUITION:
ISACA Members US $2,295 | Non-Members US $2,495

CISM Bootcamp: 4-day Exam Prep

COBIT 5: Strategies for Implementing IT Governance

Cybersecurity Fundamentals 4-day Cram Course

Foundations of IT Risk Management

Fundamentals of IS Audit & Assurance

Governance of Enterprise IT

RSA 2018 TRAINING COURSES

TUITION:
ISACA Members and Non-Members US $1,200

CISM 2-day Cram to the Max Course

CSX Cybersecurity Fundamentals 2-day Workshop

ISACA/DELOITTE TRAINING COURSES

TUITION:
ISACA Members US $2,495 | Non-Members US $2,695

Cloud Computing: Seeing through the Clouds—What the IT Auditor Needs to Know

Healthcare Information Technology

Information Security Essentials for IT Auditors

Internal Audit Data Analytics & Automation

An Introduction to Privacy and Data Protection

Network Security Auditing

Taking the Next Step—Advancing Your IT Audit Skills

For details on discounts, deadlines, registration, cancellation and more,
VISIT ISACA.ORG/TRAINING18JV1

FEATURE

Cloudifying Threats

Understanding Cloud App Attacks and Defenses

Cloud applications (apps) and services have revolutionized business productivity and efficiency by providing a robust and flexible environment in which to share and transfer data. Businesses are becoming more dependent on the cloud as the trend of adopting cloud apps is growing at an exponential rate. End users do not have a choice, as cloud apps are being shipped to them as default software in the hardware devices. For example, mobile devices are shipped with default cloud apps. Additionally, enterprise cloud apps are being used as storage solutions to host, manage and share data. That being said, every technology is susceptible to abuse and exploitation, and cloud apps are no exception.

Hackers and agents of foreign nations are increasingly exploiting cloud apps to perform nefarious operations that could potentially result in significant financial losses and compliance-related fines, in addition to loss of reputation to individuals and enterprises alike.

Before trying to understand potential cloud threats, IT departments need to have complete visibility into the channels through which data flow between users and cloud apps that exist outside of the network and perimeter security defenses. While the threats posed by shadow IT and shadow data[1] are real and persistent, enterprises are not staffed or equipped to determine how their users are accessing and transmitting the enormous flow of confidential data to and among different cloud apps. To determine this, enterprises must first gain visibility into all cloud apps being accessed by their network users before trying to understand the risk that malware and user activity pose to confidential company data.

Cloud apps have faced a wide variety of threats over the last couple of years. Google Drive has been hit by a number of phishing attacks where HTML/JavaScript(JS)[2] and OAuth[3] functionalities were abused to steal user account credentials. Dropbox, OneDrive and other cloud apps have been used to distribute malware[4, 5] to user systems. Configuration errors in cloud storage apps such as Amazon Web Services (AWS) have led to unintentional data exposure, causing security breaches that severely impacted affected organizations. Data leakage via AWS buckets[6, 7] is a grave threat to enterprises, as a small error could result in broad exposure of sensitive data. Finally, inherent design and security issues in cloud apps[8] have been regularly exploited by hackers to execute large-scale exploits. Overall, threats to cloud apps are real and enterprises must fully understand them, the potential impact to their business, and how to defend against attacks on cloud apps to protect user confidentiality and compliance-related data.

Do you have something to say about this article?

Visit the *Journal* pages of the ISACA® website *(www.isaca.org/journal)*, find the article and click on the Comments link to share your thoughts.

http://bit.ly/2k6sIQ1

Aditya K. Sood, Ph.D.
Is a director of cloud security at Symantec. Sood has research interests in cloud security, malware automation and analysis, application security, and secure software design. He has worked on a number of projects pertaining to product/appliance security, networks, mobile and web applications. He has authored several papers for various magazines and journals including IEEE, Elsevier, CrossTalk, ISACA®, Virus Bulletin and USENIX. His work has been featured in several media outlets including Associated Press, Fox News, *The Register, The Guardian, Business Insider*, CBC and others. Sood has been an active speaker at industry conferences and presented at BlackHat, DEFCON, HackInTheBox, RSA, Virus Bulletin, OWASP and many others. He is also an author of *Targeted Cyber Attacks*, published by Syngress.

Rehan Jalil
Is a senior vice president of cloud security at Symantec. He was the founder of the cloud services security company Elastica, which was acquired by Bluecoat. Previously, he was president of WiChorus (Tellabs subsidiary) and senior vice president of Tellabs. Prior to that, he was the chief architect at Aperto Networks, where he led the development of broadband wireless silicon and carrier-grade systems. At Sun Microsystems, he contributed to the development of one of the industry's earliest advanced multicore, multithreaded processors for throughput computing and graphics applications.

Cloud Apps: Threat Model and Actors

There are three types of threat actors who circumvent existing cloud security controls and trigger attacks against cloud apps:

- **Risky employees**—"To err is human,"[9] and employees are no exception to this rule. Employee mistakes regularly result in data exposure via cloud apps. When sharing documents containing sensitive information, employees often unwittingly overshare them by granting access to an overly broad audience. Some examples include:
  - An employee shares a confidential document publicly. This results in broad access, and anyone with a shared URL to the document can access the content and freely use or abuse it.
  - An employee unwittingly allows another user to download a confidential file directly from the cloud app by not specifying the access controls when the document is shared.
- **Malicious insiders**—Disgruntled employees can cause serious damage to enterprises by exploiting their position as insiders to circumvent security protocols and destroy or exfiltrate confidential data. Examples of suspicious activities performed by malicious insiders include:
  - Excessive downloads of confidential files from cloud apps
  - Excessive deletion of confidential files hosted on cloud apps
  - Broadly sharing a large number of files publicly to be accessed remotely
  - Accessing cloud apps at unusual times or for abnormally long durations
  - Accessing confidential data that they typically do not access as part of their normal job functions
  - Performing excessive printing or screen capture actions on documents stored in cloud apps
- **Hackers and state actors**—Sophisticated attackers can target cloud apps and associated users to steal data and perform unauthorized operations. Attacks can be launched directly, when attackers target a cloud app itself, or indirectly, when attackers target the users of a cloud app to gain access to their cloud accounts. Some examples of direct and indirect attacks include, but are not limited to:
  - Direct attacks:
    - Launching brute-force attacks against cloud apps
    - Discovering and exploiting inherent vulnerabilities in cloud apps

> "INHERENT DESIGN AND SECURITY ISSUES IN CLOUD APPS HAVE BEEN REGULARLY EXPLOITED BY HACKERS TO EXECUTE LARGE-SCALE EXPLOITS."

  - Indirect attacks:
    - Sending phishing emails containing malicious attachments or URLs to a web portal that could steal credentials via social engineering attacks
    - Installing malware on an end-user machine and stealing cloud app credentials via man-in-the-browser (MitB) attacks

Once the account is compromised by either direct or indirect attacks, attackers can easily exfiltrate data using multiple methods.

Cloud Apps: Threats

There are several types of cloud threats exemplified by real-world case studies, which include, but are not limited to, distributing malware via cloud apps resulting in drive-by download, account hijacking, broad sharing of sensitive documents, leaking sensitive information via the documents hosted on cloud apps, and abusing functionalities and features of cloud apps to trigger phishing attacks.

Credential Stealing and Account Hijacking

Attackers target end users to steal their enterprise cloud app credentials to perform nefarious operations. Credentials for cloud apps can be stolen in multiple ways. Examples of commonly used attack techniques include:

- **Phishing attacks**—This is the most widely used attack method deployed by attackers. It uses social-engineering techniques to trick users into providing account credentials. Attackers have used variations of phishing attacks to target end users and steal their credentials for specific cloud apps. These have been broadly categorized into three types based on how the phishing web pages are deployed and distributed:
  - Phishing pages deployed on cloud apps—This is one of the most advanced techniques, in which phishing pages are hosted on the cloud apps themselves. Attackers abuse the inherent HTML/JS-supporting functionality of cloud apps to host phishing web pages. Once the web pages are shared publicly, the associated URL, embedded in a phishing email, is sent to users. This attack is hard to defend against because the phishing pages are hosted on legitimate cloud app domains that use HTTPS. An average user will typically assume that the web pages are legitimate. As a result, users supply credentials that are transmitted via an HTTP GET/POST request to an attacker-managed domain. **Figure 1** shows a real-world attack scenario using phishing pages hosted on Google Drive. Note that Google Drive has since deprecated its HTML/JS support, but several other cloud apps still support this functionality.

Figure 1—Phishing Web Pages Hosted on Google Drive With HTTPS

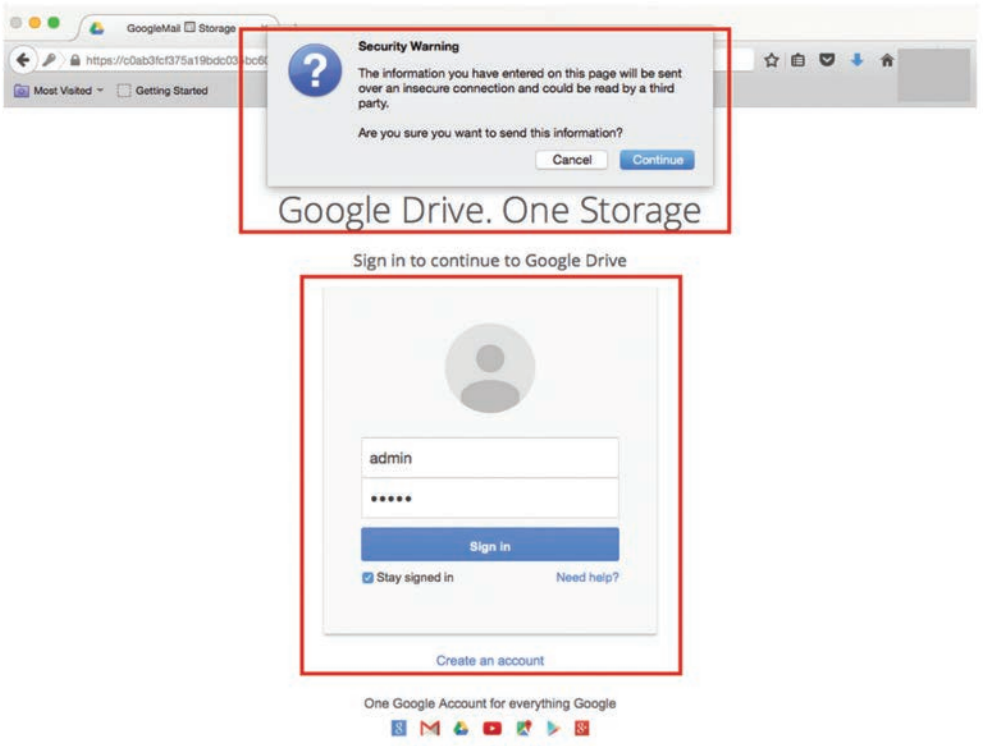


Figure 2—Phishing Web Page for Office 365 Is Encoded in the data:text/html Handler

– Phishing pages deployed as attachments—Attackers can abuse the functionality of data URLs supported by their browser. It is possible to encode data in a "data:text/html" handler, and when allowed to open in the browser's address bar, it renders the content. Attackers are using this trick to encode phishing web pages in the data handler and pass them as attachments in phishing emails. When the user opens the attachment, the content (data handler) is opened in the browser address bar and it renders the decoded phishing web page. **Figure 2** shows an example of this variant.
– Phishing pages deployed on noncloud app domains—This is the most widely used phishing technique, in which web pages of legitimate cloud apps are cloned, updated and deployed on noncloud app domains. Attackers select a domain that may look legitimate but is not. These attacks are executed in conjunction with social-engineering tactics to trick users into revealing their cloud app credentials. **Figure 3** shows an example of a phishing attack in which a web page similar to the official login page for Office 365 is deployed on the non-cloud app domain.

- **Man-in-the-browser (MitB) attacks**—MitB attacks are advanced exploits in which end-user systems are first infected with sophisticated malware, such as a bot, which is then enabled to perform advanced operations in the compromised system. The bot actually snoops communication taking place between the user's browser and the cloud app.

Figure 3—Phishing Web Page for Office 365 Account Deployed on Noncloud App Domain

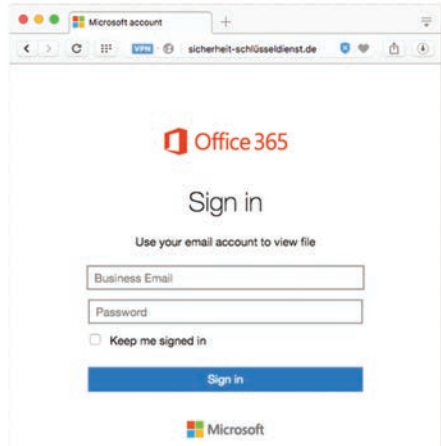
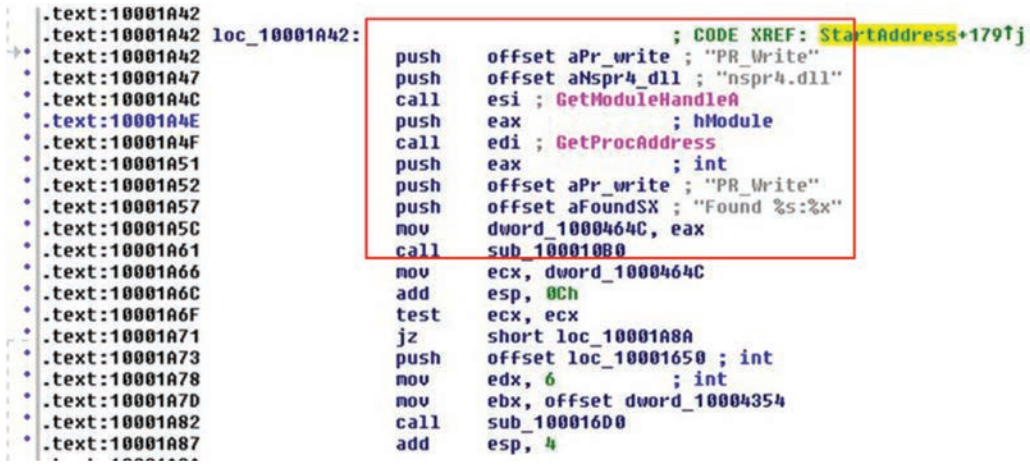


Figure 4—Code Highlights the Mozilla Firefox Function Hooked by the Malware to Steal Credentials Via the Browser



The bot injects unauthorized code into the browser process and logs the cloud app credentials entered by the user. This attack is different from a standard keylogging attack, as the attack model is different. MitB attack mode is currently deployed in a majority of botnets. **Figure 4** shows the reverse-engineered code from a malware binary highlighting the "Pr_Write" function in the "NSPR4.DLL" library. The library is hooked by the bot to steal data entered by the user in the HTML forms opened in the Mozilla Firefox browser. Primarily, the bot hooks the critical functions imported from the libraries in the browser process to dump the credentials in the HTTP GET/POST requests.

- **Man-in-the-cloud (MitC) attacks**—MitC[10] attacks are similar to MitB attacks. The difference is that tokens are stolen instead of account credentials. Tokens are used heavily in cloud apps as authentication mechanisms for transmitting data to cloud app application program interfaces (APIs) from authorized resources. Malware residing in the end-user system is capable of hijacking the communication channel. This is done by either hooking the cloud agent functions or using social-engineering attacks to inject attacker-supplied unauthorized synchronization tokens so that valid and unexpired tokens can be extracted to gain access to users' accounts. Primarily, the MitC malware exploits the file synchronization services for installing additional malware, exfiltrating data and performing command and control (C&C) operations. The attack method is different, but the end result is the same: gaining access to user accounts.

Malware Distribution

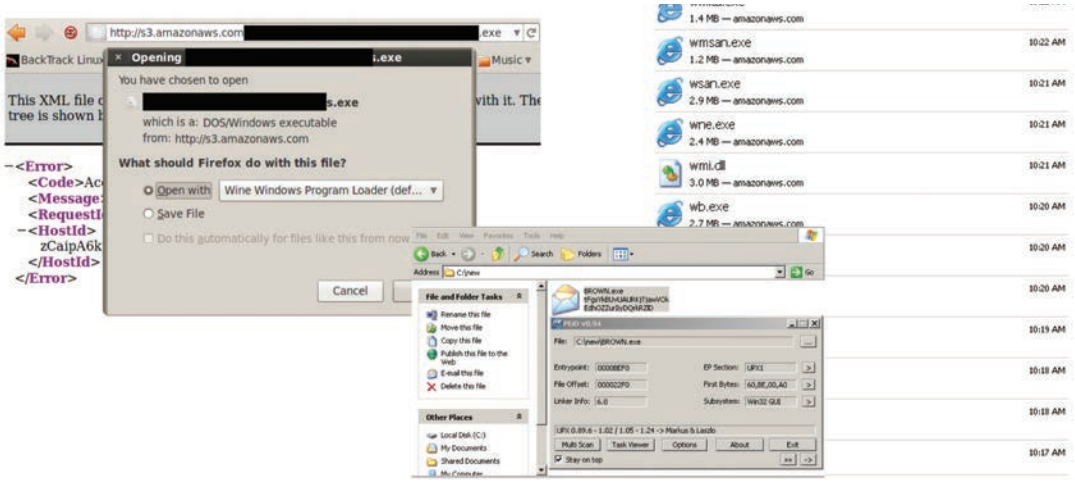
Cloud app storage functionality has been abused by attackers to distribute malicious files to end users. Malware is distributed through the cloud when attackers use stealthy techniques to upload malicious files on cloud apps or share malicious files publicly by configuring global access rights. As a result, malicious files are now ready to be shared with or distributed to end users by a cloud app's URL. To infect end users, attackers can:

- Distribute the direct cloud apps' URLs that reference malicious files to end users either by a third-party platform or via an embedded link in a phishing email
- Conduct a stealthy drive-by download attack in which the cloud app URL that references a malicious file is embedded in a third-party website in an HTML iframe or obfuscated JavaScript. When a user visits the third-party website, the cloud app URL is rendered in the browser, which downloads the file onto the end user's system. Attackers can opt to use advanced techniques to perform this operation covertly.

Overall, the basic idea for attackers is to weaponize cloud app storage functionality by using apps as malware delivery platforms. **Figure 5** shows a malicious executable (MZ header) file (Zeus bot) successfully uploaded to Google Drive. **Figure 6** shows malicious executables hosted on the AWS Simple Storage Service (S3) buckets.

Figure 5—Malicious Executable File Successfully Uploaded on Google Drive

Figure 6—Malicious Executables Hosted and Distributed via AWS S3 Buckets



Data Exfiltration and Leakage

Data exfiltration is the process of stealing and stealthily transmitting data from compromised systems to unauthorized locations on the Internet. Since enterprise cloud apps store sensitive data in the cloud, they are vulnerable to security breaches that result in the leakage of data due to human error or hackers. Data can be exfiltrated or leaked from the cloud apps in multiple ways, including:

- Users of enterprise cloud apps can share sensitive documents with a broad audience by making documents public through configuring access rights in an insecure manner, e.g., sharing sensitive files publicly via Google Drive, Box or other similar sharing sites. Amazon S3 buckets have been under the radar because multiple instances have been noted where sensitive data were disclosed via S3 buckets.
- Users can upload files containing sensitive data such as personally identifiable information (PII), payment card industry (PCI) information, and protected health information (PHI) on cloud apps and share those files in an insecure manner with other users.
- Attackers can validate and verify sensitive files hosted in compromised cloud accounts and exfiltrate the data by making those files public and downloading them onto an unauthorized server, and by sending files as attachments via emails using compromised user accounts.

- Malicious code installed on end-user systems can be directed to steal files from the folders specific to an enterprise cloud app agent that is used to sync files with the cloud servers. The malware can easily encrypt the data and exfiltrate it via either HTTP/HTTPS or other protocol channels.

Figure 7 highlights the disclosure of sensitive documents via Amazon S3 buckets.

Cloud App Vulnerabilities

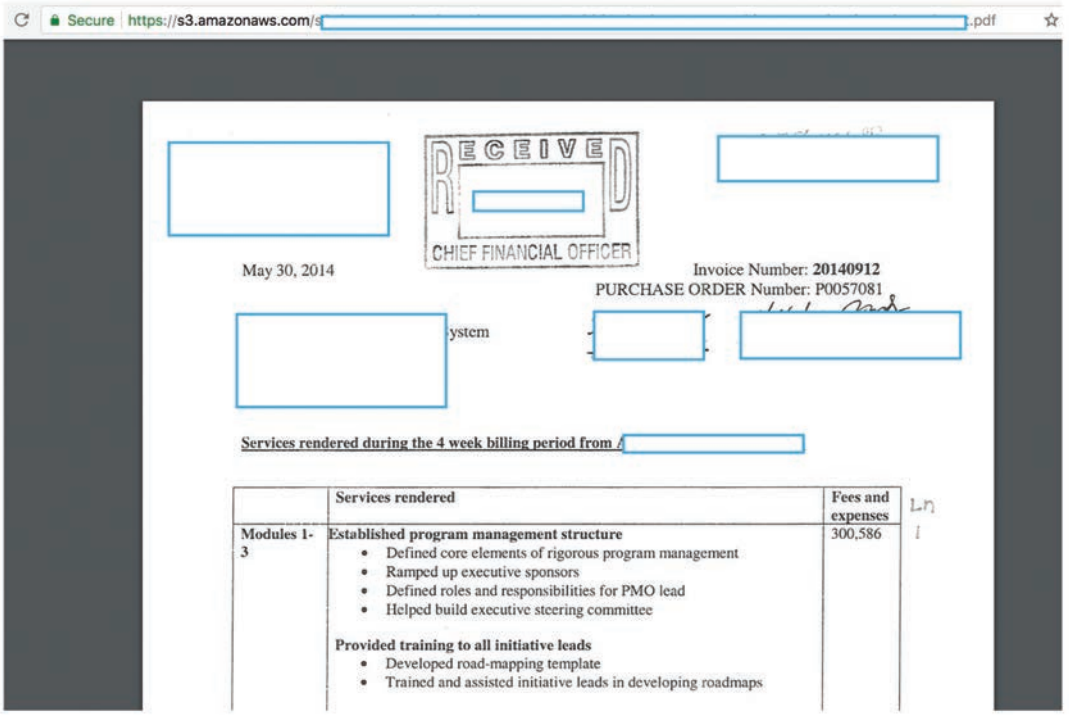
Security vulnerabilities that exist in cloud apps could be exploited by attackers to launch large-scale attacks. Cloud apps such as Google Drive, Box and Dropbox provide services to a large number of customers and exploiting inherent vulnerabilities could lead to serious widespread problems. Under responsible disclosure guidelines, security researchers disclose vulnerabilities to organizations in a secure fashion. In a similar vein, attackers are also continuing to look for vulnerabilities in cloud apps. The difference is that attackers will not disclose those vulnerabilities, but rather will exploit them to steal sensitive data or to abuse the cloud app service with unauthorized operations.

In addition to security vulnerabilities, design flaws can also contribute to the abuse and exploitation of cloud apps. Design flaws occur as a result of poor development choices that are approved without appropriate security reviews before actual components are designed. Cloud apps have been found to be vulnerable to poor single sign-on (SSO) implementations, insecure authentication mechanisms and other security-deficient design decisions. Application and infrastructure teams should take the necessary steps to avoid bugs early in the development stage and to rectify configuration errors during deployment.

GDPR Compliance and Security Breaches

The European Union General Data Protection Regulation (GDPR)[11] is a data protection regulation that requires organizations to protect the personal

Figure 7—Sensitive Documents Exposed via Amazon AWS S3 Bucket



data of the users and privacy of EU citizens. A recent poll indicates that 80 percent of US companies need to stay compliant with GDPR.[12] GDPR has taken significant steps to push data controllers and data processors under an obligation to comply with data protection requirements, which include but are not limited to providing timely notifications of security incidents; sharing data with the end users if requested, including the right to be forgotten; implementing robust security solutions to monitor the data flow; and preventing security incidents. With such obligations, data controllers and data processors are subjected to direct enforcement by the supervisory legal authorities and monetary fines as part of compensation claims by data subjects for any damages caused by security breaches if the organizations fail to stay compliant with GDPR. Generally, GDPR has given a new dimension to security and compliance by allowing data subjects to stay in control of their personal data. Data processors and controllers are required to stay more vigilant and proactive in handling sensitive information of data subjects.

> "GENERALLY, GDPR HAS GIVEN A NEW DIMENSION TO SECURITY AND COMPLIANCE BY ALLOWING DATA SUBJECTS TO STAY IN CONTROL OF THEIR PERSONAL DATA."

There are a few critical points of GDPR that pertain to security. The most relevant articles related to security in GDPR are:

- Article 33 of the GDPR details the requirements that need to be followed by data processors and controllers when implementing technical and security controls to ensure that data stay secure and private. The controls must guarantee the security, availability, confidentiality and integrity of data, including system resiliency. The expectation is to achieve stable and secure systems with maximum availability.
- Article 34 of the GDPR sets forth the requirement for data processors and controllers to report data breaches to the supervisory authority. Article 35 covers the rules requiring full disclosure of personal data breaches to the data subjects. Articles 34 and 35 are interdependent. GDPR has made stringent requirements for data controllers and processors to report security breaches within a specific time period, which is expected to be no later than 72 hours after data controllers and processors have become aware of the incident. Some flexible scenarios have been discussed for breach reporting, but, overall, this is a big leap toward ensuring that data processors and controllers are responsible for personal data. At the same time, data controllers have the responsibility to report security breach information to the data subjects without any undue delay. This has entitled data subjects to expect clear and prompt communication from data controllers if their data are stolen or leaked during a security breach. GDPR has significantly increased the responsibilities of data controllers and processors.

This article has discussed a number of attacks that can result in security breaches. Now the most important question is, "What should be done if data controllers and processors fail to adhere to the GDPR guidelines specified in articles 34 and 35?" If data controllers fail to comply with GDPR articles 34 and 35 of breach notification and disclosure, they are subject to financial penalties that could be as high as four percent of their organization's global (worldwide) annual revenue of the prior financial year or up to US $23.2 million, whichever is higher.

Security breaches in cloud apps could be a result of inherent cloud threats. As a result, enterprises can suffer financial losses by failing to adhere to compliance requirements. To avoid financial repercussions, it is essential to combat threats against cloud apps to provide a secure, safe and compliant environment.

Recommendations and Countermeasures

The following are the recommended countermeasures essential to defending against threats to cloud apps:

- The enterprise environment should be audited up-front to detect shadow data and shadow IT in the network. This is an essential step because it helps administrators discover the different types of cloud apps used, their relative risk, and how exactly end users and devices transact data with those cloud apps. With the adoption of bring your own device (BYOD), this becomes even more critical.
- Best practices dictate that all files should be scanned when they are uploaded to and downloaded from the cloud. Such scans ensure that files transferred to cloud apps do not contain any inherent threats. Engaging an active threat detection service that integrates with cloud apps is recommended. This helps prevent the distribution or syncing of files containing malware to large groups of users.

> “USER BEHAVIOR MODELING AND ANALYSIS HELPS TO DISCOVER ANOMALIES IN USER BEHAVIOR WHILE USERS ARE INTERACTING WITH CLOUD APPS.”

- Content inspection (CI) technologies that can scan file content for confidential data, such as personally identifiable information (PII), payment card industry (PCI) information, and US Health Insurance Portability and Accountability Act (HIPAA) information, before they are uploaded to cloud apps should be adopted. This helps prevent data exfiltration attempts.
- Files hosted in cloud apps should be checked for sharing and access rights. This is an important step to ensure these files are not shared too broadly. The cloud app security solution should provide administrators with the ability to scrutinize how users are sharing files with each other, thereby ensuring that unauthorized users do not gain access to confidential data.
- User behavior modeling and analysis helps to discover anomalies in user behavior while users are interacting with cloud apps. User behavior models can be designed using techniques such as supervised or unsupervised machine learning, contextual analysis, natural language processing, and others to ensure that anomalies are detected and to enable administrators to act proactively so attacks or threats can be prevented. With user behavior modeling, attacks such as unauthorized access and brute forcing can be detected easily.
- Policy enforcement is one of the preventive steps that helps avert the distribution of threats by actively blocking them once they are identified. This functionality helps administrators enforce policies in enterprise cloud apps that identify malware, prevent the leakage of sensitive data and restrict the sharing of specific files.
- Cloud infrastructure and associated cloud apps should undergo rigorous security assessments, comprising penetration testing, vulnerability assessment, configuration reviews, source code reviews, etc., to ensure that infrastructure and applications are free from security vulnerabilities and that supporting networks are sufficiently hardened. Having a strong security posture enables enterprise cloud app providers to provide an environment that is robust and secure.
- For compliance, organizations should deploy a cloud app security solution that not only helps them achieve compliance, but that also provides security controls to monitor, detect and prevent threats that reside in, and are distributed by, cloud apps. This also helps the security operations center (SOC) manage and circumvent threats that originate from cloud apps.

Conclusion

Because the technology world has encountered an exponential increase in the usage of cloud apps and security breaches via the cloud, it is necessary for enterprises to have a complete, platform-driven approach to obtain visibility into cloud apps; communication is essential so that risk and threats can be mitigated and remediated respectively. With the robust requirements listed by upcoming regulations, such as GDPR, the importance of a cloud app security solution cannot be ignored. For combating security breaches and threats in the cloud apps, a robust cloud security solution is the demand of the time to which every organization has to adhere.

Endnotes

1 Cloud Threat Labs and Symantec CloudSOC, Shadow Data Report, 2016, *https://insidecybersecurity.com/sites/insidecybersecurity.com/files/documents/mar2017/cs2017_0097.pdf*

2 Korolov, M.; "Google Drive Phishing Is Back With Obfuscation," *CSO*, 28 July 2015, *https://www.csoonline.com/article/2953190/vulnerabilities/google-drive-phishing-is-back-with-obfuscation.html*

3 Franceschi-Bicchierai, L.; "Someone Hit the Internet With a Massive Google Doc Phishing Attack," Motherboard, 3 May 2017, *https://motherboard.vice.com/en_us/article/53nzxa/massive-gmail-google-doc-phishing-email*

4 Talbot, D.; "Dropbox and Similar Services Can Sync Malware," *MIT Technology Review*, 21 August 2013, *https://www.technologyreview.com/s/518506/dropbox-and-similar-services-can-sync-malware/*

5 Sood, A.; "Cloud Storage Apps as Malware Delivery Platforms (MDP): Dissecting Petya Ransomware Distribution via Dropbox," *Symantec Connect*, 30 March 2016, *https://www.symantec.com/connect/blogs/cloud-storage-apps-malware-delivery-platforms-mdp-dissecting-petya-ransomware-distribution-dro*

6 Cameron, D.; K. Conger; "GOP Data Firm Accidentally Leaks Personal Details of Nearly 200 Million American Voters," Gizmodo, 19 June 2017, *http://gizmodo.com/gop-data-firm-accidentally-leaks-personal-details-of-ne-1796211612*

7 Constantin, L.; "Cloud Storage Error Exposes Over Two Million Dow Jones Customer Records," *Forbes*, 17 July 2017, *https://www.forbes.com/sites/lconstantin/2017/07/17/cloud-storage-error-exposes-over-two-million-dow-jones-customer-records/#17fc5c83199f*

8 Mimoso, M.; "Office 365 Vulnerability Exposed Any Federated Account," *Threatpost*, 28 April 2016, *https://threatpost.com/office-365-vulnerability-exposed-any-federated-account/117716/*

9 Pope, A.; *An Essay on Criticism, Part II*, UK, 1711

10 Imperva, *Hacker Intelligence Initiative Report: Man in the Cloud Attacks*, USA, 2015, *https://www.imperva.com/docs/HII_Man_In_The_Cloud_Attacks.pdf*

11 EUGDPR.org, "GDPR Portal," *www.eugdpr.org/eugdpr.org.html*

12 Ashford, W.; "GDPR Fines May Affect Almost 80% of US Firms, Poll Shows," *Computer Weekly*, 8 November 2017, *www.computerweekly.com/news/450429701/GDPR-fines-may-affect-almost-80-of-US-firms-poll-shows*

Mistakes Happen—Mitigating Unintentional Data Loss

日本語版も入手可能
www.isaca.org/currentissue

Nobody intends to lose data, but it happens. Often. Immature processes, inadequate tools and unaware users reveal themselves as the weak links. The growing saturation of technology in all areas of business and personal life diversifies the sources of unintentional data loss. Ever-changing laws continue to increase the risk and cost of noncompliance when unintentional data losses occur. The confluence of these factors raises the stakes for all security professionals. After reviewing today's landscape, security professionals can plan to protect tomorrow's data from risk.

Changing Landscape

Unintentional data loss occurs when confidential information leaves a corporation's boundaries without explicit approval by authorized personnel. Continuing innovation in areas such as the cloud and Internet of Things (IoT) reduces boundaries. Entrepreneurial companies start up new services, transmitting and storing data every day. The move to the cloud and the increasing usability of cloud services are accelerating the risk of potential data loss. End users purchase applications such as cloud enterprise resource planning (ERP) solutions without involvement from IT or IT security teams. Without end-user recognition, these actions create shadow IT departments. Shadow IT increases the risk of data transmission and storage outside of organizational standards and controls.

Corporations recognize the significant risk inherent in protecting data. For example, almost 20 percent of the risk factors listed in Alphabet Inc.'s (Google's parent company) 31 December 2016 10-K reflected data security risk.[1] Examples include:

- If security measures are breached resulting in the improper use and disclosure of user data, or if services are subject to attacks that degrade or deny the ability of users to access products and services, products and services may be perceived as not being secure. Users and customers may curtail or stop using the enterprise's products and services, and the enterprise may incur significant legal and financial exposure.
- Intellectual property rights are valuable, and any inability to protect them could reduce the value of an enterprise's products, services and brand. A variety

Do you have something to say about this article?

Visit the *Journal* pages of the ISACA® website *(www.isaca.org/journal)*, find the article and click on the Comments link to share your thoughts.

http://bit.ly/2AnuPtm

Mike Van Stone, CISA, CISSP, CPA
Is the director of internal audit at Ionic Security. His diverse experience includes technology, financial, operational and compliance auditing with companies ranging from Fortune 5 to a growing technology start-up. He led international, integrated and cosourced audit teams on four continents and in two languages. Van Stone has published a disaster recovery and business continuity article for Infragard's Atlanta (Georgia, USA) chapter and presented on the following topics during ISACA® Atlanta Chapter meetings and annual conferences: auditing the cybercontrols behind physical security, auditing the software development life cycle and integrated auditing.

Ben Halpert
Is vice president of risk and corporate security at Ionic Security Inc. He focuses on educating and empowering today's digital citizens in the workplace, at schools and at home. He also champions cyberethics education for children from preschool through high school via Savvy Cyber Kids, a nonprofit organization he founded in 2007. As a trusted voice on cyber and physical security issues, Halpert has made numerous conference, radio and TV appearances. He has been featured in newspapers and magazines; and has published many articles on smart technologies, data privacy and cloud computing.

Enjoying this article?

- Read *What Does It Mean To Me? GDPR Data Protection Impact Assessments*. *www.isaca.org/gdpr-dpia*

of new and existing laws could subject the enterprise to claims or otherwise harm the business.

- Privacy concerns relating to technology could damage an enterprise's reputation and deter current and potential users from utilizing its products and services.

Data Loss Laws and Regulations

As reflected in Alphabet Inc.'s 10-K, data protection laws and regulations continue to evolve with the changing landscape. Countries, states, regions and industries across the world focus on data security. Key laws, regulations and frameworks include the Canadian Personal Information Protection and Electronic Documents Act (PIPEDA), US Health Insurance Portability and Accountability Act (HIPAA), Payment Card Industry Data Security Standard (PCI DSS), US Gramm-Leach-Bliley Act (GLBA), European Union (EU) General Data Protection Regulation (GDPR), and the Basel Accords.

Noncompliance can be costly. For example, the US Department of Health and Human Services (DHHS) reported HIPAA program-to-date settlements of US $67,210,982 as of 28 February 2017.[2] However, GDPR noncompliance has the potential to surpass the fines of any of those programs. The GDPR was approved by the EU Parliament on 14 April 2016.[3] While enforcement does not begin until 25 May 2018, companies must plan for the changes now to avoid the potentially costly penalties. Key changes resulting from GDPR include increased territorial scope, clearer terms, 72-hour breach notification requirements, rights to be forgotten and privacy by design.[4]

The EU Data Protection Directive 95/46/EC set an unclear territorial scope, but its replacement, GDPR, applies to all organizations processing the personal data of data subjects residing in the European Union, regardless of the organization's location. EU's Parliament approved GDPR penalties up to the greater of US $20 million or four percent of the violating organizations's global revenue. With such severe penalties, GDPR will drive focus on implementing "appropriate technical and organizational measures in an effective way in order to meet the requirements of this regulation and protect the rights of data subjects."[5]

Organizations often establish organizational standards and controls to support compliance with laws and regulations, but GDPR will drive organizations to include data protection from the onset of system design to avoid data protection failures. In addition, data-centric controls that may have been overlooked in the past will need to be revisited by control owners.

Unintentional Data Loss

Internal employees and contractors are often the source of data loss. **Figure 1** lists by calendar year, DHHS's top five issues in investigated cases closed with corrective action.[6] Notably, technical safeguards barely make the list on an annual basis.

Figure 1—Top Five Issues in Investigated Cases Closed With Corrective Action

Year	Issue 1	Issue 2	Issue 3	Issue 4	Issue 5
2015	Impermissible Uses & Disclosures	Safeguards	Administrative Safeguards	Access	Technical Safeguards
2014	Impermissible Uses & Disclosures	Safeguards	Administrative Safeguards	Access	Technical Safeguards
2013	Impermissible Uses & Disclosures	Safeguards	Access	Administrative Safeguards	Minimum Necessary

Source: Department of Health and Human Services Office for Civil Rights, "Top Five Issues in Investigated Cases Closed with Corrective Action, by Calendar Year," USA. Reprinted with permission.

In addition to the emerging challenges, longstanding challenges such as employee loss of computers, Universal Serial Bus (USB) drives and mobile devices put data at risk. Employees enter incorrect email addresses, fail to lock their computers, inappropriately share data with unauthorized partners, use unsecured peer-to-peer messaging, and take high-resolution office pictures with sensitive equipment or data in the background. In this new landscape with opaque boundaries and more threats of unintentional data loss, security professionals must prepare for the needs of a datacentric future to meet contractual and legal protection requirements.

> IN THIS NEW LANDSCAPE WITH OPAQUE BOUNDARIES AND MORE THREATS OF UNINTENTIONAL DATA LOSS, SECURITY PROFESSIONALS MUST PREPARE FOR THE NEEDS OF A DATACENTRIC FUTURE TO MEET CONTRACTUAL AND LEGAL PROTECTION REQUIREMENTS.

Mistakes happen. However, there are important lessons to be gained from the plethora of well-publicized data protection failures. The following are selected examples of unintentional data losses and their impact:

- **Shadow IT**—A physician's attempt to disconnect his personal server from a hospital's network exposed 6,800 patients' data to Internet searches. The DHHS settled the claim for US $4.8 million.[7]
- **Stolen laptop**—An April 2016 theft of a California Correctional Health Care Services unencrypted laptop from an employee's car resulted in the loss of the personal health information (PHI) of 400,000 inmates.[8, 9]
- **Lost USB device**—The Alaska Department of Health and Human Services lost a USB device containing Medicaid beneficiaries' health information, resulting in a US $1.7 million fine.[10]
- **Lost mobile phone**—The 2013 loss of a mobile phone compromised the protected health information of 412 patients at six nursing homes, which resulted in a DHHS fine of US $650,000.[11]
- **Social media, data distribution, third-party data rights**—A UK watchdog group determined that, over a three-year period, the UK National Health Service had at least 50 instances of data being posted on social media; at least 236 instances of data being shared inappropriately via email, letter or fax; and at least 251 instances of data being inappropriately shared with a third party.[12]
- **Email**—In April 2014, an employee at a risk advisor and insurance brokerage firm accidentally sent a spreadsheet to a large group of employees containing employees' names, email addresses, birthdates, Social Security numbers, employee identification numbers, office locations and medical insurance plan details. The company paid for two years of identity theft protection for 4,830 people.[13]
- **Asset disposal**—A New York-based health plan provider returned photocopy machines to a lessor without wiping 344,579 individuals' data stored on the machine.[14]
- **Instant messaging and online disclosure**—A software, data and media company accidentally uploaded more than 10,000 confidential private instant messages of traders to a public website during testing.[15]
- **Clicking on email malware**—Banking employees clicked on malware called Carbanak, allowing hackers to gain access to data necessary to steal more than US $300 million, per Kaspersky Lab.[16]

With an ever-growing list of methods through which

> COLLECTIVELY, THE GOVERNANCE PRACTICES FORM DEFENSE-IN-DEPTH STRATEGIES TO ADDRESS DATA CONFIDENTIALITY, INTEGRITY AND AVAILABILITY.

data have been unintentionally lost, this is just a partial inventory. Many of these incidents highlighted a lack of control maturity or user awareness. Organizations must strive to continuously improve processes and controls protecting data.

Common Data Protection Controls and Considerations

Most IT processes and controls have a direct impact on data security. **Figure 2** lists several key COBIT® 5 data protection governance practices.[17]

Collectively, the governance practices form defense-in-depth strategies to address data confidentiality, integrity and availability. These layers are intended to address internal, external, intentional and unintentional data protection threats. Some common data protection controls and considerations for data security planning follow:

- **Data classification**—Classification policies and procedures drive access management decisions and data protection controls. Classification categories are often influenced by laws, regulations and frameworks. By classifying data, practitioners can empower organizations to grant access on a need-to-know basis. Most organizations struggle in identifying data and data locations due to their sheer volume, exacerbated by the presence of shadow IT. The challenges inherent in the process were made clear in the photocopying machine asset disposal data loss example previously mentioned. Organizations produce volumes and volumes of data every day and are often reliant on individuals to ensure data are consistently and correctly marked. Identifying, classifying and protecting data by classification are covered under COBIT® processes IDs such as APO01.06, BAI08.02, BAI08.03, DSS05.02, DSS05.03, DSS05.04 and DSS06.06.

Figure 2—COBIT 5 Data Protection Governance Practices

Practice ID	COBIT Practice Name
APO01.06	Define information (data) and system ownership.
APO07.06	Manage contract staff.
BAI08.02	Identify and classify sources of information.
BAI08.03	Organise and contextualise information into knowledge.
DSS05.01	Protect against malware.
DSS05.02	Manage network and connectivity security.
DSS05.03	Manage endpoint security.
DSS05.04	Manage user identity and logical access.
DSS05.06	Manage sensitive documents and output devices.
DSS05.07	Monitor the infrastructure for security-related events.
DSS06.03	Manage roles, responsibilities, access privileges and levels of authority.
DSS06.05	Ensure traceability of information events and accountabilities.
DSS06.06	Secure information assets.

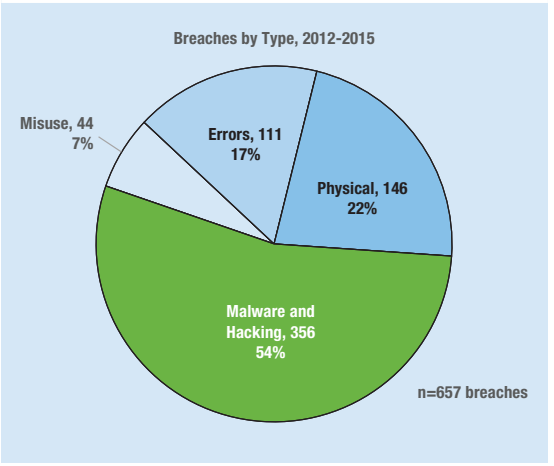
Source: ISACA®, *COBIT® 5 Implementation—Supplemental Tools and Materials*, USA, 2013. Reprinted with permission.

- **Data loss prevention (DLP)**—Ultimately, when marked or identified through rules, DLP tools can provide another layer of protection. Establishing data loss prevention rule sets requires careful planning and testing. If the rules are too restrictive, the tool becomes an impediment to business; if the rules are too permissive, data security is negatively impacted. DSS06.06 addresses the application of data classification policies and procedures to protect information assets through tools and techniques.
- **Encryption**—Organizations often apply encryption to data classified as sensitive. Data can be encrypted by multiple types of tools. For the previously noted data loss examples caused by asset loss, hard disk encryption may have mitigated the risk. However, hard drive encryption would not have helped with the examples in which the data left the hard disk and, especially, the network boundaries. File-level encryption tools add another layer of protection, but these tools have historically been hampered by insiders with knowledge of keys and passwords who retain access to the file. Additionally, difficulties in scaling key management effectively can limit enterprise-level implementations. As a result, end users often use consumer-focused file-level data encryption products that remain outside the scope of IT's centralized management processes, thereby creating further control weaknesses.

End users can often be the root cause of data losses. For example, the 2016 California Data Breach Report findings shown in **figure 3** reflect a common theme.[18] The errors category, totaling 17 percent of the breaches between 2012 and 2015, represents incidents stemming from anything insiders (employees or service providers) unintentionally do or leave undone that exposes personal information to unauthorized individuals. The chart also includes a physical category. This category includes both intentional thefts and unintentional hardware losses. COBIT covers encryption of information in storage and in transit according to its classification in DSS05.02 and DSS05.03.

- **User awareness**—Organizations build end-user awareness through policies, procedures, training and even phishing simulation tests. However, malicious actors continue to target employees to gain access to sensitive data. Building a security culture within an organization is necessary to drive higher compliance with procedures and gain full acceptance of security tools and monitoring. Former US White House Military Office Chief Information Security Officer Steve Pugh shared his method for gaining acceptance in this way: "I trust you, but I don't trust the packets coming from your machine."[19] This can be an effective cultural enabler to implement controls necessary to combat both unintentional and intentional data leakage. DSS05.01 emphasizes the importance of communicating malicious software awareness and enforcing prevention procedures and responsibilities.

Figure 3—California Data Breach Report Findings



Source: Harris, K.; "California Data Breach Report 2012-2015," California Department of Justice, USA, February 2016. Reprinted with permission.

- **Access management**—Properly addressing access management can be a difficult challenge. Most organizations deal with a constant flow of employees and contractors, a growing list of internally and externally hosted applications, and significant complexities in segregation of duties. Restricting access to files unintentionally distributed outside of an organization's network poses a significant obstacle to true end-to-end management

of access to data. Contractor access agreements, access reviews, access principles, access privilege assignment and segregation of duties are addressed in multiple processes, including APO07.06, DSS05.04, DSS05.06 and DSS06.03.

- **Logging and monitoring**—Logging and monitoring help detect security events and provide evidence to trigger the incident management process. Key challenges include monitoring third-party cloud applications accessible to employees from outside the corporate network and insufficient automation to effectively manage the large volume of log data. DSS05.07 concentrates on logging, level of log detail and log review.

> "FIREWALLS ARE STILL AN IMPORTANT PART OF ANY DEFENSE-IN-DEPTH STRATEGY, BUT THEIR SCOPE OF MITIGATION IS NOW MORE LIMITED."

- **Antimalware**—This software can often be required by contracts or industry regulations. However, over the years bad actors have become more savvy and intentionally develop hacks to evade common malware tools. This cat-and-mouse game continues to play out daily, but long gone are the days when antimalware software alone provided comprehensive endpoint and network protection. COBIT documents the standard to install and activate malicious software protection tools with updated definition files to filter incoming traffic and protect against unsolicited information in DSS05.01.
- **Firewalls**—These network security systems historically served as the castle walls effectively protecting data, but now, more than ever, data are leaving corporate networks by design and flowing to the cloud, partner connections and the multitude of consumer-managed devices introduced through bring-your-own-device (BYOD) implementations within organizations. In this changing landscape, firewalls are still an important part of any defense-in-depth strategy, but their scope of mitigation is now more limited. COBIT process DSS05.02 focuses on implementing network filtering mechanisms to control inbound and outbound traffic in accordance with policies.

The majority of the current data protection controls form part of the defense-in-depth strategy, but many of these controls have limitations when it comes to addressing the current threat landscape, resulting in unmitigated risk. Looking forward, additional solutions should be evaluated to further lessen the risk of unintentional data loss.

The Future of Data Protection

Ultimately, technical security controls should focus on protecting data. This means knowing and dynamically changing policy by controlling who is accessing sensitive data at the point of consumption and even when data leave the enterprise's boundaries. Organizations should build capabilities to analyze data usage inside and outside of their network to identify anomalies of access based on the role of an individual trying to access specific data. Additionally, data encryption should be decoupled from applications to leverage the scalability of the cloud and enable data protection at creation, in use, at rest and in transit. These steps will serve to bind security, access management and analytics to data.

Conclusion

The cost of noncompliance and loss of trust due to unintentional and intentional data loss are high; incidents are too frequent. Leaders in the organization must focus on developing a data protection strategy that provides the control, scalability, visibility and flexibility needed to meet the needs of the ever-changing threat and regulatory landscape. These security characteristics enable those empowered to mitigate risk to prevent, identify and mitigate unintentional data loss.

Endnotes

1 Alphabet Inc., 2016 10-K Form, 2017, *https://abc.xyz/investor/pdf/20161231_alphabet_10K.pdf*

2 Department of Health and Human Services (DHHS), "Enforcement Highlights," USA, 31 March 2017, *https://www.hhs.gov/hipaa/for-professionals/compliance-enforcement/data/enforcement-highlights/index.html*

3 European Union, "General Data Protection Regulation (GDPR) Portal: Site Overview," *www.eugdpr.org/eugdpr.org.html*

4 European Union, "GDPR Key Changes," *www.eugdpr.org/key-changes.html*

5 *Ibid*.

6 Department of Health and Human Services (HHS), "Top Five Issues in Investigated Cases Closed With Corrective Action, by Calendar Year," USA, 16 September 2016, *https://www.hhs.gov/hipaa/for-professionals/compliance-enforcement/data/top-five-issues-investigated-cases-closed-corrective-action-calendar-year/index.html*

7 Department of Health and Human Services, "Data Breach Results in $4.8 Million HIPAA Settlements," USA, 7 May 2014, *https://www.hhs.gov/hipaa/for-professionals/compliance-enforcement/examples/new-york-and-Presbyterian-hospital/index.html*

8 California Correctional Health Care Services, "Potential Breach of Patient Health Information," USA, 13 May 2016, *www.cphcs.ca.gov/docs/press/Release%20-%20Potential%20Breach%20PHI.pdf*

9 Department of Health and Human Services, Office for Civil Rights, "Breach Portal: Notice to the Secretary of HHS Breach of Unsecured Protected Health Information," USA, *https://ocrportal.hhs.gov/ocr/breach/breach_report.jsf*

10 Department of Health and Human Services, "Alaska DHSS Settles HIPAA Security Case for $1,700,000," USA, *https://www.hhs.gov/hipaa/for-professionals/compliance-enforcement/examples/alaska-DHSS/index.html*

11 Department of Health and Human Services, "Business Associate's Failure to Safeguard Nursing Home Residents' PHI Leads to $650,000 HIPAA Settlement," USA, *https://www.hhs.gov/hipaa/for-professionals/compliance-enforcement/agreements/catholic-health-care-services/index.html?language=es*

12 Smitheringale, S.; "New Report: Patient Confidentiality Broken 6 Times a Day," Big Brother Watch, 14 November 2014, *https://www.bigbrotherwatch.org.uk/2014/11/new-report-patient-confidentiality-broken-6-times-day/*

13 New Hampshire Office of the Attorney General, "Re: Notification of Data Security Incident," USA, 17 April 2014, *www.doj.nh.gov/consumer/security-breaches/documents/willis-north-america-20140417.pdf*

14 Department of Health and Human Services "HHS Settles With Health Plan in Photocopier Breach Case," USA, 7 August 2013, *https://www.hhs.gov/hipaa/for-professionals/compliance-enforcement/examples/health-plan-photocopier-breach-case/index.html*

15 Seward, Z. M.; "Bloomberg Accidentally Posted Private Terminal Messages Online," *Quartz*, 13 May 2013, *https://qz.com/84004/bloomberg-accidentally-posted-private-terminal-messages-online/*

16 Sanger, D. E.; N. Perlroth; "Bank Hackers Steal Millions via Malware," *The New York Times*, 14 February 2015, *https://www.nytimes.com/2015/02/15/world/bank-hackers-steal-millions-via-malware.html?_r=0*

17 ISACA, *COBIT® 5 Implementation—Supplemental Tools and Materials*, USA, 2 December 2013, *www.isaca.org/COBIT/Documents/Forms/DispForm.aspx?ID=7294*

18 Harris, K.; "California Data Breach Report 2012-2015," California Department of Justice, February 2016, *https://oag.ca.gov/sites/all/files/agweb/pdfs/dbr/2016-data-breach-report.pdf?*

19 Pugh, S.; Data Protection Update Course, The Proscenium, Atlanta, Georgia, USA, 22 September 2016

FEATURE

Applying AI in Application Security

Do you have something to say about this article?

Visit the *Journal* pages of the ISACA® website *(www.isaca.org/journal)*, find the article and click on the Comments link to share your thoughts.

http://bit.ly/2AWQFDL

Kiran Maraju, CEH, CISSP
Has 18 years of information security experience and is involved in multiple network, web application and mobile security vulnerability assessments. He also has experience with penetration testing, security code review, software development life cycle security, network security, database security and wireless security assessments. Maraju is working as a specialist leader for Deloitte. Prior to that, he worked as a scientist for the Indian Space Research Organization (ISRO).

The use of artificial intelligence (AI) in cyber security will help organizations enhance existing application security capabilities. Application security covers the security of web or thick client and mobile applications that pass through various phases of the security development life cycle, e.g., security design and security coding. Various AI areas such as machine learning and expert systems can be leveraged to improve application security to derive, predict or apply inferences to forecast security threats, identify security vulnerabilities and identify the security coding remediation guidance.

The following AI areas can be applied to application security:

- **Machine learning**—Decision-tree learning (DTL) during threat identification
- **Expert systems**—Forward chaining and backward chaining for security code review and code review guidance

Security auditors can make use of these techniques to automate the attack threat identification and code review process. The process involves developing various decision support inference rules for various application security vulnerabilities, applying the decision and inference rules to expert systems, and training the same systems using an algorithm with the various application security attack scenarios and attack paths. Security auditors can identify and infer the possibility of successful attacks by providing inputs to these machine-learning-based application security expert systems. **Figure 1** describes the relationship between DTL and expert systems in this process.

Machine Learning DTL—Application Security Attack Threat Identification

DTL is a form of inductive learning that uses a training set of examples to create a hypothesis that makes general conclusions. It also determines best attribute paths of attack (threats) if possible. During decision-tree development, a set of training examples is divided into smaller subsets and an associated decision tree is developed incrementally. A decision tree covering the training set is returned after completing the learning process. **Figure 2** describes threat decision trees.

Cross-site scripting (XSS) is a web application security vulnerability where input data submitted to the web application are not validated properly for malicious inputs. Attackers use this XSS vulnerability to hijack sessions and use cookies and for page defacement attacks.[1]

The following is a typical XSS scenario with training examples, including attack paths with different attributes available such as input data validation, output data validation and data type. The tainted data type values with possible data types are returned to the end user (reflective XSS), saved in the server (persistent XSS) or transferred to a different system as input. By applying entropy and gain values, the next-best-attribute decision tree with an entropy value selected as root node and with all subsequent branch nodes and root nodes is constructed as depicted in **figure 3**.

The decision-tree method is a type of supervised learning that involves attributes, training sets,

Figure 1—DTL, Forward Chaining and Backward Chaining

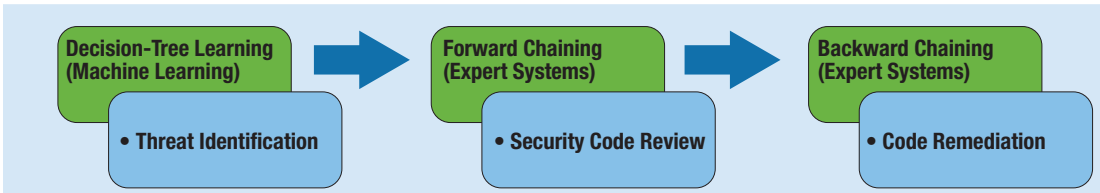
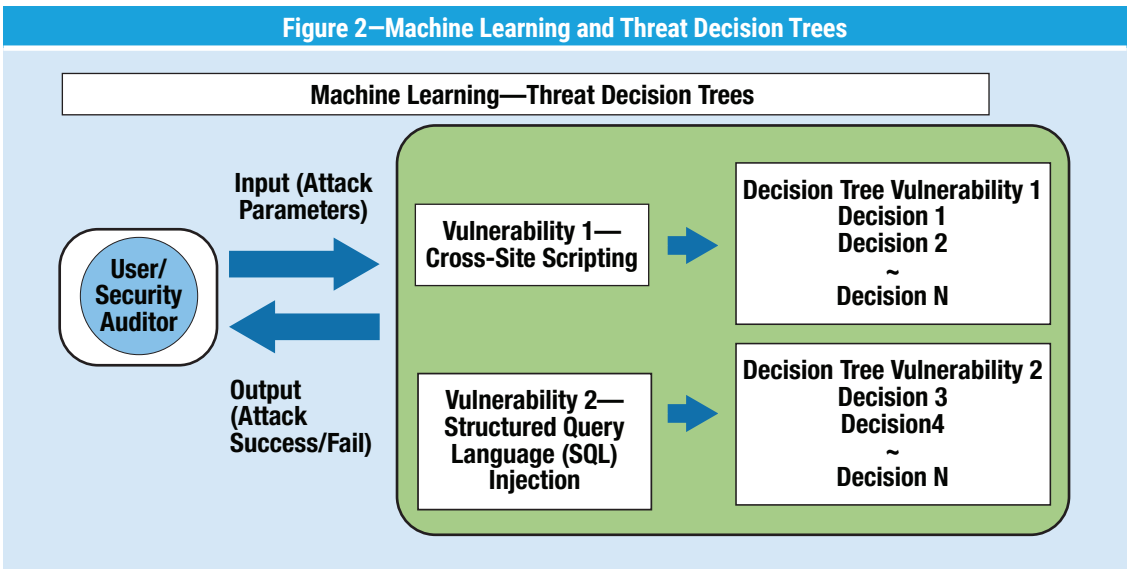


Figure 2—Machine Learning and Threat Decision Trees



and determining the best attribute according to entropy and gain with algorithms such as Iterative Dichotomiser 3 (ID3).[2]

The information gain (G), G(S,A) where A is an attribute and S is a sample of training examples

p_+ is the positive examples in S
p_- is the negative examples in S

- Entropy of S: Average optimal number of bits to encode information about certainty/uncertainty. Entropy (E) is the minimum number of bits needed to classify an arbitrary example as yes or no.

$$Entropy(S) = -p_+ \log_2 p_+ - p\text{-}\log_2 p_-$$

- Gain(S,A): Reduction in entropy after choosing attribute A

$$Gain(S, A) = Entropy(S) - \sum_{v \in Values^{*}A} \frac{|S_v|}{|S|} Entropy(S_v)$$

Therefore, the entropy of the training data, E(S), can be represented as E ([3+, 9-]) because out of the 12 training examples, three of them are attack success and nine of them are attack fails. **Figure 4** describes attack paths and their outcomes.

The previous attack paths are provided as training sets to the decision support system, and the input is passed through this decision-support system. This is depicted in **figure 3** as the XSS attack decision tree. When the input data validation is not performed, this will check for output data validation and, if output data validation is not performed, it will check for data type and provide the outcome as attack success (i.e., XSS attack is possible) or attack fail (i.e., XSS attack is not possible). For attack path P13, input data validation is performed with the decision-tree outcome attack fail. This means that XSS is not possible using the above training sets.

Figure 3 is the XSS vulnerability decision tree. Similar types of decision trees can be created for various application security vulnerabilities to correlate, identify and predict the security threats once these decision trees are constructed for various security vulnerabilities. These decision support system rules can be leveraged as inference rules for the application security expert systems.

Application Security Expert Systems

Expert systems are capable of interpreting the provided input, advising and deriving a solution. They also provide suggested alternatives to the problem and predict the results. **Figure 5** illustrates the components of an expert system—knowledge base and inference engine. The knowledge base comprises information that is factual and heuristic (guessable), including rules and facts. If-then-else rules are a part of the knowledge representation. Information acquired from security experts is fed to the expert systems in the form of if-then-else rules to develop security expert systems. Facts are the assertions.

Figure 3—XSS Attack Tree

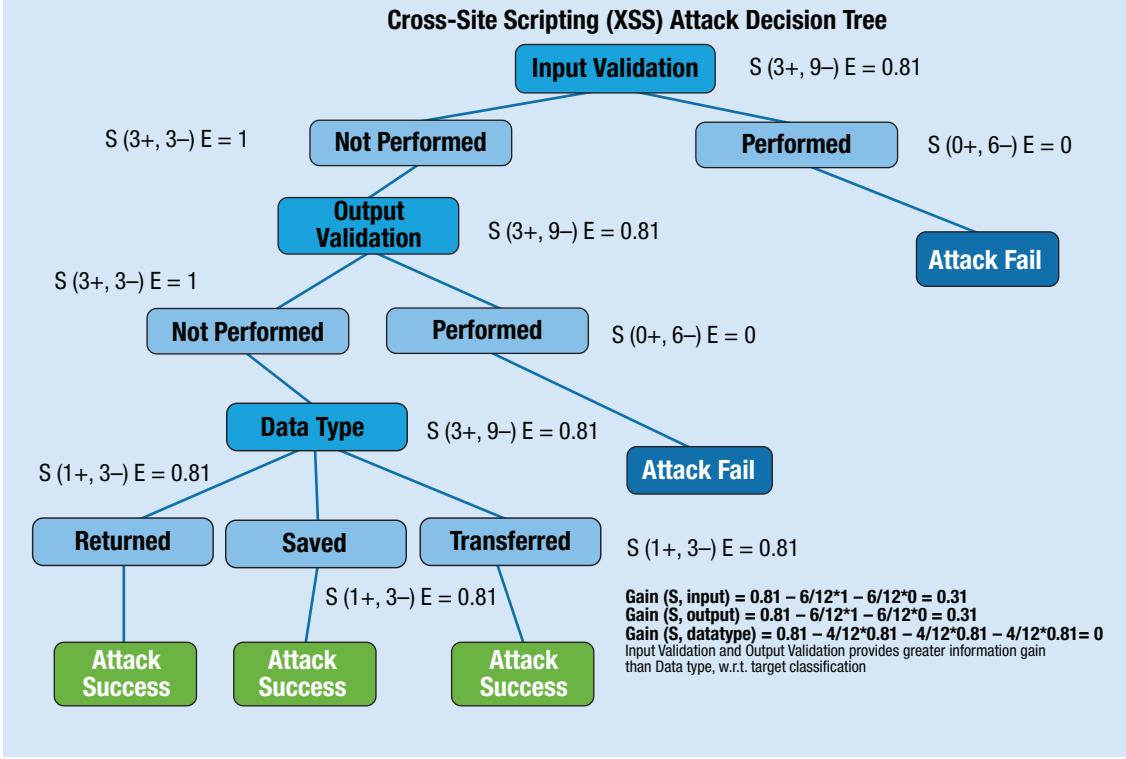
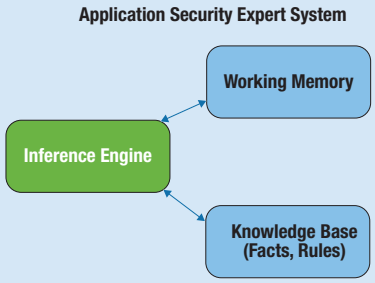


Figure 4—Attack Paths and Outcomes

Attack Path	Input Data Validation	Data Type	Output Validation	Result (XSS Injection Attack)
P1	Not Performed	Returned	Not Performed	Success
P2	Not Performed	Returned	Performed	Fail
P3	Not Performed	Saved	Not Performed	Success
P4	Not Performed	Saved	Performed	Fail
P5	Not Performed	Transferred	Not Performed	Success
P6	Not Performed	Transferred	Performed	Fail
P7	Performed	Returned	Not Performed	Fail
P8	Performed	Returned	Performed	Fail
P9	Performed	Saved	Not Performed	Fail
P10	Performed	Saved	Performed	Fail
P11	Performed	Transferred	Not Performed	Fail
P12	Performed	Transferred	Performed	Fail
P13	Performed	Returned	Not Performed	??

Figure 5—Expert System Components



Security Code Review Inference Engine—Expert Systems Forward Chaining

The security code review inference engine with forward chaining[3] can be used to predict values, i.e., deriving what can happen next. This will help the security code review engines to actually determine the type of attack. **Figure 6** details the SQL injection[4] vulnerability, with the forward chaining inference rules using if-then-else rules by matching the various conditions.

SQL injection is a web application security vulnerability where input data submitted to web applications are not validated properly for malicious inputs. Using SQL injection vulnerability attacks, attackers will inject malicious SQL commands to the back-end database and exfiltrate database details. **Figure 7** describes the security code review inference engine and forward chaining rules.

Forward Chaining Inference Rules for SQL Injection Vulnerability

The following are examples of the facts and rules used to demonstrate the forward chaining inference rules for SQL injection vulnerability.

Facts:

- **Fact 1:** X is a URL parameter.
- **Fact 2:** X contains a special character.
- **Fact 3:** X is not white-list input validation processed.
- **Fact 4:** X is passed through a database call.

Rules

- **Rule 1:** If X is a form parameter and X contains special characters, then X is a tainted input.
- **Rule 2:** If X is a URL parameter and X contains special characters, then X is a tainted input.
- **Rule 3:** If X is a cookie parameter and X contains special characters, then X is a tainted input.
- **Rule 4:** If X is a file import and X contains special characters, then X is a tainted input.
- **Rule 5:** If X is an HTTP header and X contains special characters, then X is a tainted input.
- **Rule 6:** If X is a tainted input and input is not white-list input validation processed, then X is an unvalidated input.

Figure 6—SQL Injection Vulnerability With Forward Chaining

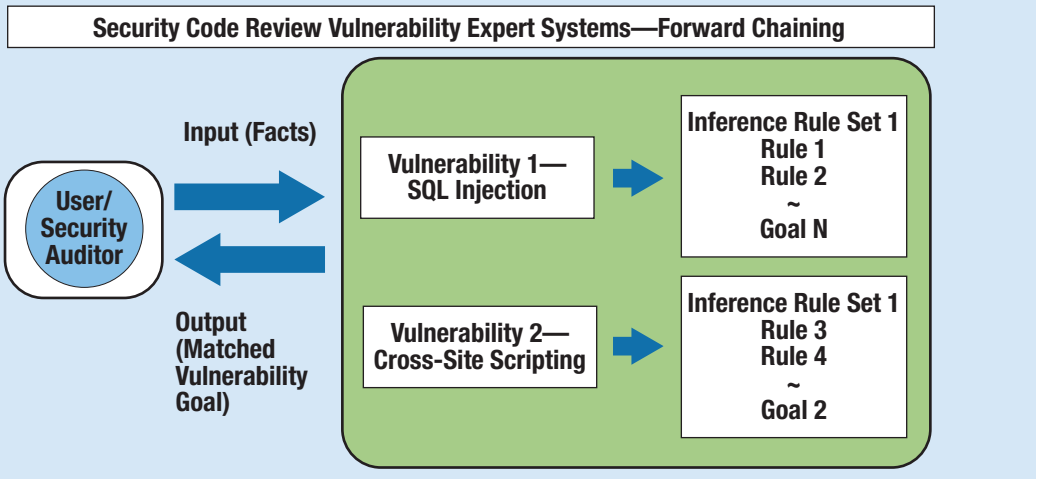
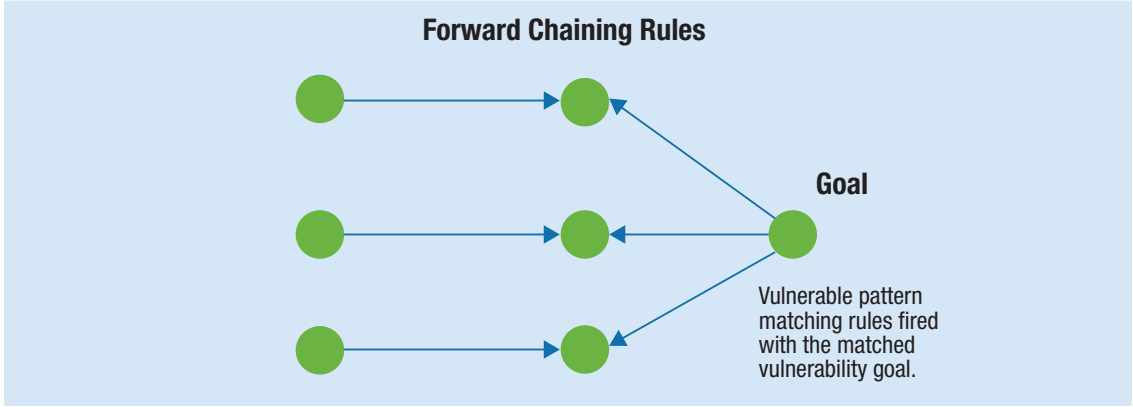


Figure 7—Security Code Review Inference Engine



- **Rule 7:** If X is a tainted input and X is not blacklist input validation processed, then X is an unvalidated input.
- **Rule 8:** If X is a tainted input and not an escaping input, then X is an unvalidated input.
- **Rule 9:** If X is an unvalidated input and X is not processed in prepared statements (with parameterized queries), then X is a potential SQL injection input.
- **Rule 10:** If X is an unvalidated input and X is passed through a database call, then X is a potential SQL injection input.

Figure 8 details the rules matching the current working memory, conflict set, rule fired and the next cycle after a rule has fired.

Application Security Vulnerability Remediation Guidance—Expert Systems Backward Chaining

The application security remediation guidance inference engine with backward chaining can be used for diagnosis of the values, i.e., deriving what happened. This application security remediation guidance expert system helps the developer to determine various possible sub-goals (solutions) to fix the vulnerabilities. **Figure 9** describes backward chaining expert systems that provide guidance for security vulnerabilities with inputs as goals and facts and outcomes as possible matched sub-goals.

Figure 10 details the application security remediation guidance rules for SQL injection remediation vulnerability where rules 14, 15 and 16 will be sub-goals. These sub-goals can be treated as possible solutions that a developer can implement to remediate the SQL attack, i.e., escape the input or implement blacklist/white-list input validation for special characters.

Backward Chaining Inference Rules for Identifying SQL Injection Vulnerability Recommendations

The goal is SQL injection, and the rules are:

- **Rule 11:** If X is a potential SQL injection input, then X is an unvalidated input.

Figure 8—Working Memory, Conflict Set and Rule-Fired Cycles

Cycle	Working Memory	Conflict Set	Rule Fired
0	URL parameter, special characters, tainted input	2	2
1	URL parameter, special characters, tainted input, input not whitelist input validation processed	2, 6	6
2	URL parameter, special characters, tainted input, input not whitelist input validation processed, X is passed through a database call	2, 6, 10	10
3	URL parameter, special characters, tainted input, input not whitelist input validation processed, X is passed through a database call, SQL injection	2, 6, 10	Halt

Figure 9—Application Security Vulnerability Remediation Guidance Expert Systems, Backward Chaining

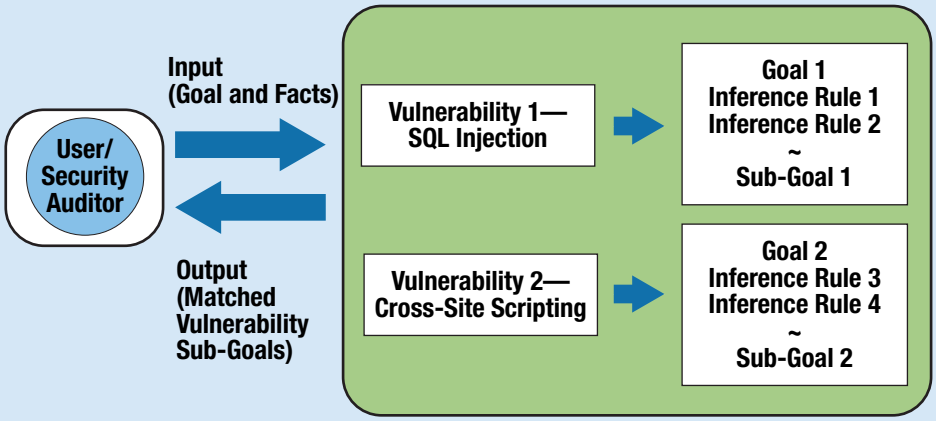
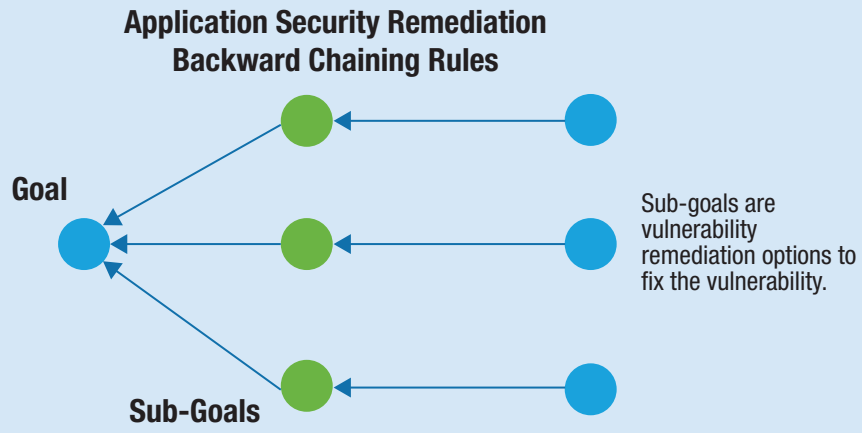


Figure 10—Application Security Remediation and Sub-Goal Development



- **Rule 12:** If X is a potential SQL injection input, then X is not processed in prepared statements (with parameterized queries).
- **Rule 13:** If X is a potential SQL injection input, then X is passed through a database call.
- **Rule 14:** If X is an unvalidated input, then X is a tainted input and not an escaping input (sub-goal).
- **Rule 15:** If X is an unvalidated input, then X is a tainted input and X is not blacklist input validation processed (sub-goal).
- **Rule 16:** If X is an unvalidated input, then X is a tainted input and X is not white-list input validation processed (sub-goal).
- **Rule 17:** If X is a tainted input, then X is from HTTP request header and X contains special characters.
- **Rule 18:** If X is a tainted input, then X is from file import and X contains special characters.
- **Rule 19:** If X is a tainted input, then X is from cookie parameter and X contains special characters.

Figure 11—SQL Injection Working Memory, Conflict Set and Rule-Fired Cycles			
Cycle	**Working Memory**	**Conflict Set**	**Rule Fired**
0	SQL injection input	11, 12, 13	11
1	SQL injection input, unvalidated input	11, 12, 13, 14, 15, 16	14
2	SQL injection input, unvalidated input, tainted input and not escaping input	11, 12, 13, 14, 15, 16	15
3	SQL injection input, unvalidated input, tainted input and not escaping input, not white-list input validation processed, not blacklist input validation processed	11, 12, 13, 14, 15, 16	16
4	SQL injection input, unvalidated input, tainted input and not escaping input, not white-list input validation processed, not blacklist input validation processed	11, 12, 13, 14, 15, 16	Halt

- **Rule 20:** If X is a tainted input, then X is from URL parameter and X contains special characters.
- **Rule 21:** If X is a tainted input, then X is from parameter and X contains special characters.

Figure 11 details the rules matching the current working memory, conflict set, rule fired and the next cycle after a rule has fired.

Working memory (WM) = ([unvalidated input] [tainted input] [not escaping input] [not white-list input validation processed] [not blacklist input validation processed]) comprises the various methods (sub-goals) that can be considered as possible solutions to remediate the SQL injection vulnerability.

Conclusion

Decision-tree machine learning and application security expert system techniques can be leveraged to automate decision-making to determine the next best attributes to use to identify the attack paths to classify/identify security threats, security vulnerabilities and code remediation guidance. This can be achieved by identifying and providing all possible attack scenarios to DTL and application security expert systems. Training sets are incrementally developed to create hypotheses to derive conclusions. The application security expert systems with forward and backward chaining can also be used to determine the security vulnerabilities, i.e., deriving consequences based on possible antecedents (matched rules), and can also be used for advising security vulnerability coding remediation solutions to fix the vulnerabilities.

Endnotes

1 Open Web Application Security Project, Cross-Site Scripting (XSS), *https://www.owasp.org/index.php/Cross-site_Scripting_(XSS)*
2 Quinlan, J. R.; "Induction of Decision Trees," *Machine Learning 1*, p. 81–106, Kluwer Academic Publishers, USA, 1986, *www.hunch.net/~coms-4771/quinlan.pdf*
3 Al-Ajlan, A.; "The Comparison Between Forward and Backward Chaining," *International Journal of Machine Learning and Computing*, vol. 5, iss. 2, 2015, p. 106–113, *www.ijmlc.org/vol5/492-A14.pdf*
4 Open Web Application Security Project, SQL Injection, *https://www.owasp.org/index.php/SQL_Injection*

FEATURE

Big Data Deidentification, Reidentification and Anonymization

日本語版も入手可能
www.isaca.org/currentissue

Big data seems indeterminate due to its constant use in intellectual data science fields and science, technology and humanities enterprises. There is a growing need to understand what big data can do for society at large. Not only can it improve human life by innovating speedier medical releases in the marketplace, but it can also utilize computing power to analyze large data sets and improve the efficiency of current technologies.

The use of big data is possible only with the proper dissemination and anonymization of publicly accessible data. To facilitate and administer the implementation of controls around the subject of big data, one must truly understand the concepts of deidentification, reidentification and anonymization. One famous study demonstrated that 87 percent of the American population can be uniquely identified by their gender, ZIP code and date of birth.[1] This illustrates the idea that anonymization, while practical, requires further study and due diligence. It is important that personal data that have been anonymized are anonymized correctly before being used as part of a publicly available big data set. Auditing professionals who work with big data, deal with global privacy implications and handle sensitive research data require the knowledge and technical aptitude to audit the big data space to stay relevant. Almost all enterprises are now taking on big data projects, and staying compliant with growing regulatory risk requirements is causing internal compliance, risk and audit functions at these enterprises to demand auditors with these necessary skill sets.

Deidentification, Reidentification and Anonymization

It is critical to reflect that the Data Protection Directives (DPD) definition of personal data is personal information relating to an identified or identifiable natural person.[2] It is possible for the controller or a third party to identify the data subject at hand, directly or indirectly, by referencing the data subject's identification number or through one or more concepts specific to the physical, mental, economic, cultural or social identity of the data subject. Therefore, it is important to consider the deidentification, reidentification and anonymization of data in big data sets when considering data use for enterprise projects and external-facing studies.

Deidentification is the altering of personal data to establish an alternate use of personal data so it is next to impossible to identify the subject from which

Do you have something to say about this article?

Visit the *Journal* pages of the ISACA® website *(www.isaca.org/journal)*, find the article and click on the Comments link to share your thoughts.

http://bit.ly/2BesQYI

Mohammed J. Khan, CISA, CRISC, CIPM
Is a global audit manager at Baxter, a global medical device and health care company. He works with C-suites across audit, security, medical device engineering (cyber) and privacy offices. He has spearheaded multinational global audits and assessments in several areas, including enterprise resource planning systems, global data centers, cloud platforms (AWS, SFDC, etc.), third-party manufacturing and outsourcing reviews, process re-engineering and improvement, global privacy assessments (EUDD, HIPAA, GDPR), and medical device cyber security initiatives in several markets over the past five years. Most recently, he has taken on further expertise in the area of medical device cyber security. Khan has previously worked as a senior consultant for Ernst & Young and Deloitte and as a technology expert for global ERP/supply chain systems at Motorola. He frequently speaks at national and international conferences in the space of data privacy, cyber security and risk advisory.

the data were derived. **Figure 1** is an example of deidentification where the column "Student Name" is removed.

Figure 1—Deidentification of Personal Data

~~Student~~ Name	Graduating Year	Grade Average	Number of Classes Failed
~~Mark Smith~~	1996	B	1
~~Langley Michael~~	1989	C	2
~~Noah Sandoh~~	2003	A	0

Reidentification is the method of reversing the deidentification by connecting the identity of the data subject. For example (building on the previous example), one could use LinkedIn to determine that Mark Smith graduated high school in 1996. This allows for the reidentification of Mark Smith's record (it is the only one showing a graduating year of 1996), thereby revealing his grade average and number of classes failed.

Anonymization is the ability for the data controller to anonymize the data in a way that it is impossible for anyone to establish the identity of the data.

Figure 1 could be anonymized as shown in **figure 2** (using the techniques of generalization, noise addition and permutation, which will be explained).

Figure 2—Anonymization

~~Student Name~~	Graduating Year (Generalization)	Grade Average (Noise Addition: +/- Random Value A to F)
~~John Doe~~	~~1996~~ 1990s	~~B~~ B+
~~Jane Doe~~	~~1989~~ 1980s	~~C~~ C-
~~John Smith~~	~~2003~~ 2000s	~~A~~ A-

European and US Laws Related to Data Anonymization Concepts

As mentioned earlier, the DPD definition of personal data is information relating to an identified or identifiable person. Specifically, Article 2(a) of the DPD states

> *"Personal data" shall mean any information relating to an identified or identifiable natural person ('data subject'); an identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity.*[3]

Directive 95/46/EC refers to anonymization in Recital 26 to exclude anonymized data. Recital 26 signifies that to anonymize any data, the data must be stripped of sufficient elements such that the data subject can no longer be identified. The e-Privacy Directive (Directive 2002/58/EC) also refers to "anonymization" and "anonymous data" very much in the same regard.[4]

> "ANONYMIZATION IS THE ABILITY FOR THE DATA CONTROLLER TO ANONYMIZE THE DATA IN A WAY THAT IT IS IMPOSSIBLE FOR ANYONE TO ESTABLISH THE IDENTITY OF THE DATA."

The US Department of Health and Human Services (HHS) enforces the US Health Insurance Portability and Accountability Act (HIPAA) and establishes specific and strict standards for deidentification of covered health data or protected health information (PHI).[5] The deidentification standard requires that PHI must remove all 18 specified patient identifiers[6]

and apply statistical or scientific principles to validate the reidentification of the deidentified data prior to using it for big data purposes.

Methods of Pseudonymizing and Anonymizing Data

Pseudonymization is the process of deidentifying data sets by replacing all identifying attributes, that are particularly unique (e.g., race, gender) in a record with another. However, the data subject owner in this case (the owner of the original data set) can still identify the data directly, allowing for reidentification. For example, if one were to eliminate all identifying data elements and leave an internal numerical identifier it would make reidentification impossible for a third party, but very easy for the data controller. Thus, such identifiers, that is, all pseudonymized data, are still personal data.

> "ANONYMIZATION IS ESSENTIALLY THE DESTRUCTION OF IDENTIFIABLE DATA; THEREFORE, IT IS VIRTUALLY IMPOSSIBLE TO REESTABLISH THE DATA TOGETHER."

The pseudonymized data are not normally supposed to be used as test data; they must be anonymized. One can rely on randomly generated data from some key sites that specialize in such use.[7] Pseudonymization reduces the linkage of data sets with the original identity of the data subject, thereby avoiding any legal issues with the deidentification and anonymization of personal data prior to releasing it into the big data space. The implementation of pseudonymization to secure the data from being identifiable at the data-subject level requires basic guidelines including:

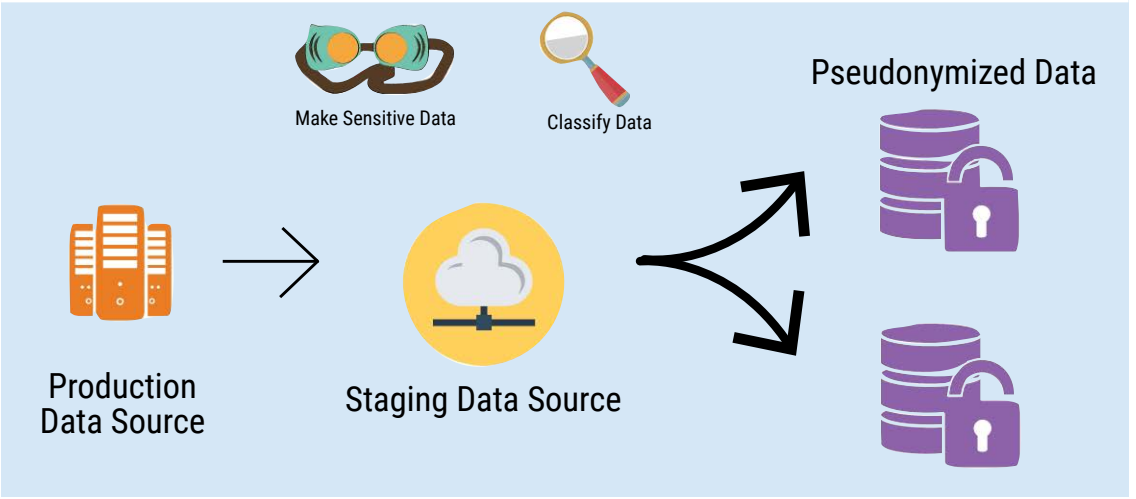
- Eliminating the ability to connect data sets to other data sets, making identification of anonymized data uniquely identifiable
- Storing the encryption key securely and separately from the encrypted data
- Data protection using administrative, physical and technical security measures

Figure 3 demonstrates how pseudonymization works.

Anonymization is achieved when the data can no longer be used to identify a natural person by using "all the means likely reasonable to be used by the controller or by any other person."[8] Compared to pseudonymization, anonymization of data is irreversible. It is virtually impossible to reestablish the anonymized data once the links between the subject and the subject's records are broken and anonymized. Anonymization is essentially the destruction of identifiable data; therefore, it is virtually impossible to reestablish the data together.

For example, every day, John attends a yoga class at the same yoga studio and, on his way, he buys a donut from the store next to the studio. John also uses the same method of payment and, once a week, he uses the pay phone next to the donut store to call his wife to let her know he is going to pick her up a donut to bring home. Even if the data owner of the previous example has "anonymized" John's personally identifiable data (e.g., name, address, phone number), the behavior he displays can possibly be used to directly identify him. Hence, it is important to anonymize his data by stating facts through grouping. For instance, "10 people went to the yoga studio and purchased donuts every day from the store next to the studio" and "20 people called from the pay phone one day out of the week." This data now is anonymized, since one can no longer identify John's predictable pattern of behavior. The solution of anonymizing data truly prevents the owner of the data and enterprises using the data to identify individual data sets. Randomization

Figure 3—Pseudonymization



changes the accuracy of the data by removing the unique identifier between the data and the individual. There are two methods to perform this technique:

- **Noise addition**—Alters the attributes by adding or subtracting a different random value for each record (e.g., adding a different random value between A+ and C- for the grade of the data subject)

> "RANDOMIZATION CHANGES THE ACCURACY OF THE DATA BY REMOVING THE UNIQUE IDENTIFIER BETWEEN THE DATA AND THE INDIVIDUAL."

- **Permutation**—Consists of swapping the values of attributes from one data subject to another (e.g., exchanging the incomes of data subjects with failed grades of data subject A with data subject B)

Conclusion

Big data will exponentially grow and, as studies show, "A full 90 percent of all the data in the world has been generated over the last two years."[9] The use of big data to capitalize on the wealth of information is already happening, and this can be seen from the daily use of technology platforms such as Google Maps or predictive search patterns while on a website. As auditors, it is important to understand the basic concepts of big data to properly address personally identifiable data with anonymizing or deidentifying. Growing regulations around data usage, including specific changes to the regulatory and privacy landscape in both Europe and in the United States, will require careful technical and legal frameworks. As data keep exponentially increasing, while new regulations requiring data owners to properly protect the identity of their data subjects emerge, it is more important than ever to carefully tread through such topics in big data to better the technology and innovations that will come with the use of big data.

Endnotes

1 Sweeney, L.; "Simple Demographics Often Identify People Uniquely," Data Privacy Working Paper 3, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA, 2000, *https://dataprivacylab.org/projects/identifiability/paper1.pdf*

2 Office of the Data Protection Commissioner, "EU Directive 95/46/EC—The Data Protection Directive," European Union, *https://www.dataprotection.ie/docs/EU-Directive-95-46-EC-Chapter-1/92.htm*

3 *Ibid.*

4 Data Protection Working Party, "Opinion 05/2014 on Anonymisation Techniques," Article 29 Data Protection Working Party, European Union, 10 April 2014, *http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf*

5 Department of Health and Human Service, "45 CFR Subtitle A (10–1–10 Edition)," USA, *https://www.gpo.gov/fdsys/pkg/CFR-2010-title45-vol1/pdf/CFR-2010-title45-vol1-sec164-502.pdf*

6 Department of Health and Human Services, "Guidance Regarding Methods for Deidentification of Protected Health Information in Accordance With the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule," USA, *https://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/index.html*

7 Mockaroo, Realistic Data Generator, *www.mockaroo.com*

8 Office of the Data Protection Commissioner, "Anonymisation and Pseudonymisation," European Union, *https://www.dataprotection.ie/docs/Anonymisation-and-pseudonymisation/1594.html*

9 Dragland, A.; "Big Data—For Better or Worse," SINTEF, *https://www.sintef.no/en/latest-news/big-data-for-better-or-worse/*

FEATURE

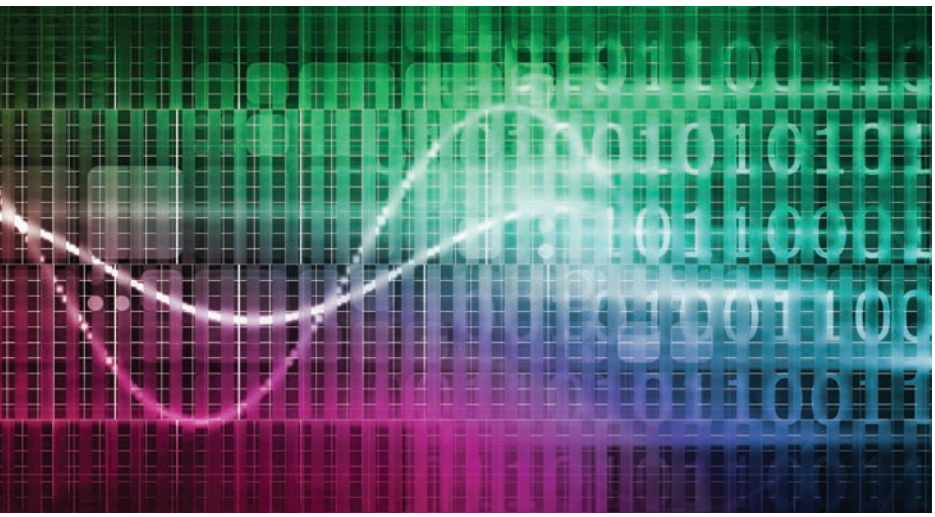
The Machine Learning Audit—CRISP-DM Framework

Do you have something to say about this article?

Visit the *Journal* pages of the ISACA® website *(www.isaca.org/journal)*, find the article and click on the Comments link to share your thoughts.

http://bit.ly/2BIz7I6

Machine learning is revolutionizing many industries, from banking to manufacturing to social media. This mathematical optimization technique is being used to identify credit card fraud, tag individuals in photos and increase e-commerce sales by recommending products. Machine learning can be summarized as a computer recognizing patterns without explicit programming. For example, in traditional software engineering, the computer must explicitly be programmed via control statements (e.g., if this event happens, then do this), necessitating that the engineer design and implement the series of steps the computer will perform to complete the given task. However, when dealing with mass amounts of correlated data (two or more variables moving together or away from each other, e.g., the relationship between temperature and humidity), human intuition breaks down. With advances in computing power, the abundance of data storage and recent advances in algorithm design, machine learning is increasingly being utilized by corporations to optimize existing operations and add new services, giving forward-thinking, innovative companies a durable competitive advantage. This increased usage helps establish the need for machine learning audits.[1, 2, 3] However, a standard procedure for how to perform a machine learning audit has yet to be created. Using the Cross Industry Standard Process for Data Mining (CRISP-DM) framework may be a viable audit solution.



Andrew Clark
Is a principal machine learning auditor for Capital One. At Capital One, he establishes approaches for auditing machine learning algorithms and prototypes ways to use machine learning for audit process optimization. He has designed, built and deployed a continuous auditing infrastructure across 17 global manufacturing subsidiaries for a publicly traded manufacturing conglomerate and built a consolidated data mart off the American Institute of Certified Public Accountants Audit Data Standards.

The Machine Learning Audit

There are many possible ways to approach a machine learning audit, ranging from a standard software development life cycle (SDLC) approach to a full code review with mathematical assumption inquiries. As in many areas in life, the Pareto principle,[4] also known as the 80/20 principle, can be applied to show that 80 percent of the benefit is yielded by 20 percent of the work. For example, auditors can provide a high level of assurance on the efficacy of a specific machine learning algorithm by examining a modified, domain-specific version of the SDLC of planning, defining, designing, building, testing and deployment,[5] assuming a traditional waterfall approach. In many cases, the maturity level of data science work flows is below that of traditional software engineers, but a general process should still be covered in their analysis. The CRISP-DM model, arguably the industry standard for how machine learning is conducted by practitioners (even if they have not explicitly followed the framework), follows the same principles, but is modified to the needs of the machine learning process.[6] The steps are:

1. Gain an understanding of the business
2. Gain an understanding of the data
3. Prepare the data

4. Complete modeling
5. Evaluate
6. Deploy

By following the CRISP-DM approach, a level of assurance can be obtained by a high-level review, with more assurance provided if subject matter experts examine each step in more depth. It is important to note that the CRISP-DM framework needs to be modified for more targeted auditing and is more crucial than ensuring that the proper steps of work are documented. That is the reason for proposing the CRISP-DM. For the data preparation, modeling and evaluation stages, a thorough evaluation usually requires a significant knowledge of programming, databases, linear algebra, probability theory, statistics and calculus, but, arguably, one of the most important steps in a machine learning audit can be conducted without these skills, assuming the assistance of the audited, by relying on the traditional auditing technique of sampling.

In the machine learning community, there is considerable interest in making interpretable machine learning models, where individuals can understand that a given classification was made (e.g., this radiology report shows cancer, vs. a noncancerous growth) purposefully. This is important since, in many domains such as medicine, individuals will not trust an algorithmic result unless he/she understands why the prediction was made. Several frameworks exist for auditing machine learning algorithms, such as the Local Interpretable Model-Agnostic Explanations (LIME)[7] and FairML[8] frameworks; however, these frameworks provide only an interpretation of the model weights and not a risk-based holistic understanding of the machine learning process. This is where the CRISP-DM approach comes into effect. It should be noted that the LIME and FairML frameworks can be utilized in conjunction with the CRISP-DM framework in the evaluation stage to assist the auditor in understanding the model.

When a machine learning model is fully trained and put into production, it receives data from one set of attributes at a time or as a data stream, depending on the use case. For this example, assume a discrete model with a single set of attributes given to the model one at a time. In either case, after examining what the input parameters are, the auditor could derive a pseudo set of data to feed into the algorithm and examine the predicted outcomes for characteristics that would help to expose any potential biases, e.g., a loan prediction model discriminating against a racial group by zip code alone. By feeding in data over a gamut of possibilities, assurance over the potential biases of the performance of a model can be obtained without fully explaining how or why the algorithm makes a certain prediction. Even with a subject matter expert, a globally interpretable model evaluation method does not currently exist for certain models (e.g., support vector machines and neural networks). By evaluating the output of this sampling series (which should be repeated multiple times with the same input data to ensure consistency), practical accuracy can be determined compared to the mathematical accuracy used when models are being trained by data scientists (the performance, i.e., the accuracy of the model and its ability to meet the business requirements are less complex to ascertain).

“IN THE MACHINE LEARNING COMMUNITY, THERE IS CONSIDERABLE INTEREST IN MAKING INTERPRETABLE MACHINE LEARNING MODELS, WHERE INDIVIDUALS CAN UNDERSTAND THAT A GIVEN CLASSIFICATION WAS MADE PURPOSEFULLY.”

Business Understanding

The business understanding section should be relatively straightforward from an audit perspective, but can be challenging during the development of a model. This section addresses what the business use case is and, with the help of a domain expert, what attributes of the use case should be included in the model such as income amount, job title, education level. In sophisticated environments, when other types of models have already been used, either as software or mental decision models, this step can be a little easier than starting from

scratch. As the CRISP-DM framework is iterative, the business understanding section will be revisited often during a large project.

Data Understanding

Data understanding is an incredibly important step of the CRISP-DM framework. Without understanding the nature and idiosyncrasy of the data, an accurate model cannot be constructed. However, there is more to this step than meets the eye, since most data, besides categorical variables, have an inherent scale to them, such as Celsius, Fahrenheit, kilometers, miles, etc.

Another important consideration is where the data are housed. Different stores of data have different considerations such as the given schema of a relational database. Without a thorough understanding of the data, strong algorithmic models and their subsequent audits cannot be accomplished.

> “WITHOUT A THOROUGH UNDERSTANDING OF THE DATA, STRONG ALGORITHMIC MODELS AND THEIR SUBSEQUENT AUDITS CANNOT BE ACCOMPLISHED.”

The auditor needs to be vigilant at this stage to understand all the variables and ensure that these variables are not in conflict or introducing biases. Correlation and covariance matrices could be examined at this stage to understand how the variables correlate and vary in response to one another.

Data Preparation

Once the data scientist understands what the use case is, how the data have been collected and other details, preprocessing the data into a usable form for modeling is required. For relational data, there may not be much “wrangling” required to get the data into an amendable structure. However, with unstructured text, such as log files and website scrapped data, the preprocessing stage can be very time consuming. Techniques such as regular expressions (regex) may be required to separate text strings such as extracting an IP address from a log file. The following regex command is able to parse out an IP address, 172.16.254.1, for example:[9]

```
\b\d{1,3}\.\d{1,3}\.\d{1,3}\.\d{1,3}\b
```

In most instances, the data need to be scaled so that all the features or dimensions have the same scale. Usually z-score standardization is used, yielding a mean x=0 and a standard deviation s=1 of:

$$z = \frac{x - \bar{x}}{s}$$

For instance, continuing with the Celsius vs. Fahrenheit example, using a set of Celsius values, C = {10, 30, 25, 15} and a set of Fahrenheit values F= {80, 37, 52, 47}, one can scale them by calculating their means:

$$\bar{x} = \frac{x_1 + x_2 + x_3 + x_4}{n}$$

and their standard deviations:

$$s = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2}$$

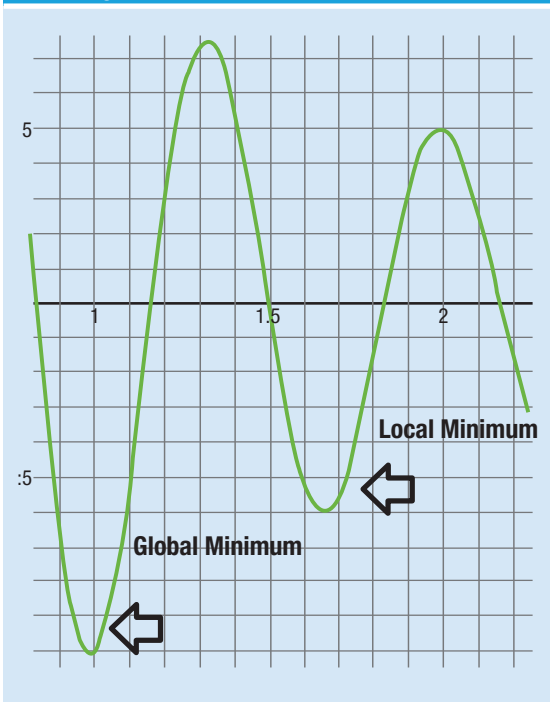
After standardization, the Celsius and Fahrenheit scores are: $C_z = \{-1.1, 1.1, 0.5, -0.5\}$ and $F_z = \{1.4, -0.9, -0.1, -0.4\}$. When the data are standardized, the individual scales are removed, i.e., if the Fahrenheit values are first converted to Celsius and then standardized, the same result will be achieved.

Modeling

Modeling is the vital component of machine learning. It is commonly assumed that data scientists and machine learning engineers spend much of their time modeling; however, in most machine learning projects, modeling is one of the shorter steps, at least for the initial implementation. Many different knobs can be tweaked in different directions to refine the performance of an

algorithm. However, many data scientists use intuition followed by brute-force grid search techniques in practice to try out all the available hyper-parameters (parameters set prior to training that are not learned, within a given range of values, yielding the best output). Depending on the number of hyper-parameters tried and the complexity of the algorithm, this can be a very computationally intensive task. If algorithms have not been tuned, then the models are most likely not fully optimized. This usually implies that they have not reached their global minima, but a lack of model tuning does not endanger an algorithm's understandability. See **figure 1** for a visual representation of the difference between a local and global minimum.

Figure 1—Global and Local Minimums



Recent developments in the field of automated machine learning[10] are increasingly used not only to tune the hyper-parameters of models, but to select the specific algorithm itself. When using automated machine learning, the data scientist and auditor need to be vigilant and examine the model selected to ensure that the degree of interpretability required for the given use case is met. This means that the business must explain why each decision was made, as in the case of companies' subject to the "right to explanation" clause of the European Union's (EU) General Data Protection Regulation (GDPR).[11] In this context, a nonlinear support vector machine model would not be an acceptable choice. One clear benefit that the GDPR has influenced is more emphasis on model interpretability in algorithmic design.[12] In 2016, the International Conference on Machine Learning (ICML) began a yearly workshop focusing on model interpretability, aptly called the "Workshop on Human Interpretability (WHI) in Machine Learning."[13]

> "AN EXTREMELY IMPORTANT MACHINE LEARNING AUDIT CHECKLIST ITEM SHOULD BE EXAMINING IF THE DATA WERE BIFURCATED INTO TRAINING AND TEST SETS."

An extremely important machine learning audit checklist item should be examining if the data were bifurcated into training and test sets. Splitting apart the data helps to prevent against model overfitting, which means the algorithm matches the characteristics of the individual data set too closely, causing it not to generalize well to new data. Traditionally, data are split in an 80/20 split, 80 percent of the data as training data and 20 percent as testing data. Modern best practices take this a step further, using a cross-validation process to split the training data into smaller pieces and testing the model on random subsets of the data. A common approach used is called k-fold cross-validation. This divides the data set into k-1 folds (the k is specified by the machine learning engineer), where the data are iteratively split and tested against the held-out fold to reduce the risk of over-fitting the model to a particular data set. **Figure 2** illustrates when a model is underfit to the data, correctly fit to the data and overfit to the data, which has the lowest error, but will not predict new data well.

Enjoying this article?

- Learn more about, discuss and collaborate on IT audit tools and techniques in the Knowledge Center. *www.isaca.org/it-audit-tools-and-techniques*

Evaluation

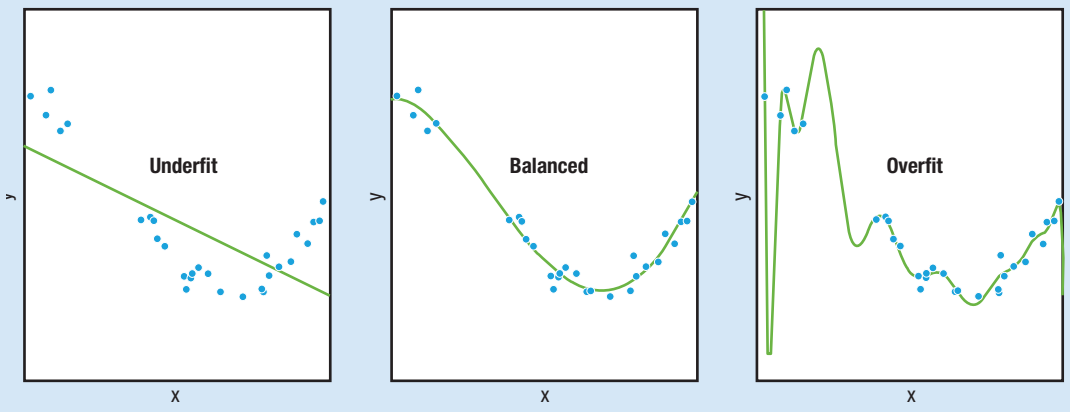
The evaluation section is arguably the most important section from an audit perspective. It is in this area of the machine learning process, or pipeline, that the model is validated for accuracy and the individual weights can be seen for some models. Traditionally, models are evaluated on their prediction accuracy and generalizability to the production data. However, from an audit perspective, an evaluation of the outcome is a key concern. If the model has an extremely high prediction accuracy (90 percent) and appears to be generalizing well, it still may not be meeting the goals of the model and/or be violating the principles of the business, such as inadvertently performing racial discrimination. In addition to examining all the steps outlined so far, the auditor should create a sample set of data to feed into the algorithm and evaluate the outcome to look for any unintended effects the model may produce. For example, for a loan approval model, the auditor could create a data set with zip codes from affluent, middle class and poor neighborhoods; 90 percent income, median household and poverty level; and each major ethnicity in the region, Caucasian, Hispanic and African American, with each combination of the above, producing a total of 84 data points.

More variables may be in the model, such as credit score, employment information, etc. Test variables would need to be created for these as well. The auditor may discover a culture bias in the model, which may increase the accuracy of the model, but enforce a self-perpetuating bias. This, in turn, could lead to bad publicity and the decrease in accuracy of taking out the race variable; for instance, it may increase the revenue and profit generated from the model. Of course, this is a simplified example and many models may not have any social issues involved, but the process for identifying potential trouble spots and testing for them remains.

Deployment

Specifically, how and where the algorithm in question is deployed is less of a concern to the auditor if the service level and desired capabilities are being met. However, there is one area from which auditor awareness and examination could provide value: technical debt.[15] Whenever a developer is building a system, certain decisions will be made about language, application programming interface (API), open-source libraries, how much documentation to create, how many unit tests to create, etc. Essentially, technical debt is less-than-ideal factors integrated into a system. Technical debt is not inherently bad. It is the result of decisions made to get projects done on time and within budget. However, it is not without consequences. In machine learning, technical debt is harder to spot and remediate than in traditional software engineering projects because of the learning aspect. For the purposes of this example, the focus is on correction cascades, an insidious variety of technical debt. A correction cascade occurs when the algorithm is not producing the desired result and rule-based "fixes" are applied on top of the model to correct for its deficiencies.

Figure 2—Overfitting



Source: F. Pedregosa, *et al*. Reprinted with permission.[14]

These deficiencies might be outlier cases or have occurred because of a poorly trained model or inadequate training/validation data. The problem is that if too many fixes are applied when the model is being trained and tweaked, it becomes increasingly difficult to ascertain what changes to the model will produce improvements, since filters are on top of the results and essentially create an upper bound on the learning capability of the model. Technical debt can be spotted by an experienced, knowledgeable data scientist working on the model. However, the knowledge gained from an audit report may solidify the need to retool a model that data scientists already knew had technical debt.

Conclusion

The CRISP-DM framework has been introduced to instruct auditors on how to perform a high-level machine learning audit. For a particularly deep dive, a machine learning specialist will be required, but by following the given framework, machine learning auditing can be accessible to more audit departments.

Endnotes

1 Clark, A.; "Focusing IT Audit on Machine Learning Algorithms," MISTI Training Institute, 3 November 2016, *http://misti.com/internal-audit-insights/focusing-it-audit-on-machine-learning-algorithms*

2 Clark, A.; "Machine Learning Audits in the 'Big Data Age'," *CIO Insights*, 19 April 2017, *www.cioinsight.com/it-management/innovation/machine-learning-audits-in-the-big-data-age.html*

3 O'Neil, C.; *Weapons of Math Destruction*, Crown Publishers, USA, 2016

4 The Editors of Encyclopedia Britannica, "Vilfredo Pareto," *Encyclopedia Britannica*, *https://www.britannica.com/biography/Vilfredo-Pareto*

5 Kaur, S.; "A Review of Software Development Life Cycle Models," *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 5, iss. 11, 2015, p. 354–60, *https://www.ijarcsse.com/docs/papers/Volume_5/11_November2015/V5I11-0234.pdf*

6 Marbán, G. M.; J. Segovia.; "A Data Mining and Knowledge Discovery Process Model, Data Mining and Knowledge Discovery in Real Life Applications," *Intech.com*, 2009, *http://cdn.intechopen.com/pdfs/5937/InTech-A_data_mining_amp_knowledge_discovery_process_model.pdf*

7 Ribeiro, M. T.; S. Singh; C. Guestrin; "Why Should I Trust You?," *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '16*, 13 August 2016

8 Adebayo, J. A.; "FairML: ToolBox for Diagnosing Bias in Predictive Modeling," *DSpace@MIT*, 2016, *http://hdl.handle.net/1721.1/108212*

9 Goyvaerts, J.; "How to Find or Validate an IP Address," *Regular-Expressionns.info*, *www.regular-expressions.info/ip.html*

10 Feurer, M.; A. Klein; K. Eggensperger; J. Springenberg; M. Blum; F. Hutter; "Efficient and Robust Automated Machine Learning," *Advances in Neural Information Processing Systems 28*, vol. 1, 2015, p. 2962–70, *http://papers.nips.cc/paper/5872-efficient-and-robust-automated-machine-learning.pdf*

11 Goodman, B.; S. Flaxman; "EU Regulations on Algorithmic Decision-Making and a 'Right to Explanation,'" *2016 Icml Workshop on Human Interpretability in Machine Learning (WHI 2016)*, New York, NY, USA, 2016, *https://arxiv.org/pdf/1606.08813v1.pdf*

12 *Op cit,* Goodman

13 Second Annual Workshop on Human Interpretability in Machine Learning, *WHI 2017*, 10 August 2017, *https://sites.google.com/view/whi2017/home*

14 Pedregosa, F., *et al.*; "Scikit-Learn: Machine Learning in Python," *Journal of Machine Learning Research*, vol. 12, October 2011, p. 2825-2830

15 Sculley, D.; G. Holt; D. Golovin; E. Davydov; T. Phillips; D. Ebner; V. Chaudhary; M. Young; "Machine Learning: The High Interest Credit Card of Technical Debt," *SE4ML: Software Engineering 4 Machine Learning (NIPS 2014 Workshop)*, November 2014, *www.eecs.tufts.edu/~dsculley/papers/technical-debt.pdf*

Implementation of Big Data in Commercial Banks

Do you have something to say about this article?

Visit the *Journal* pages of the ISACA® website *(www.isaca.org/journal)*, find the article and click on the Comments link to share your thoughts.

http://bit.ly/2jtuWtg

Adeniyi Akanni, Ph.D., CISA, CRISC, ITIL
Worked as an experienced systems auditor and an information security expert in commercial banks in Nigeria for more than two decades, holding various positions as head of e-fraud and investigation and head of reconciliation and compliance. He currently works with the First Bank of Nigeria Ltd.

Big data can be described as a huge volume of data that cannot be treated by traditional data-handling techniques.[1] Considering the enormity of data generated in various forms at various times via various devices, it is clear that such data would not only be unstructured, but also complex. Without proper collation, coordination and harnessing of those data, meaningful decisions may not be reached by relevant users, such as chief information technology officers, systems auditors and chief executive officers (CEOs).

Proper implementation of big data can be an indicator of effective usage of big data because data continue to grow exponentially.[2] Big data is big because of a high level of volume, velocity and variety. This high level is due to the way data are generated and continuously increasing in quantity (volume) at a very fast rate (velocity) and in various forms (variety).

The cost associated with storing petabytes of data is not the major problem for organizations such as commercial banks, because cloud service providers (CSPs) can offer such services at reasonable prices. The big challenge lies in the form in which the data are generated, which does not follow a specific pattern (relating to the variety of data generation), and the rate at which the large quantity of data is generated (velocity).[3] Thus, a careful selection of strategy is necessary so that such data can facilitate informed decision making that would, in turn, affect the security of the data and positively leverage information obtained from big data for a competitive edge.[4] This article describes a six-stage cycle of implementing big data in commercial banks, points out the major challenges in implementation and provides a suggested solution. It also assists CEOs with properly analyzing their data for optimal marketing drives.

Implementation of Big Data at the Bank's End

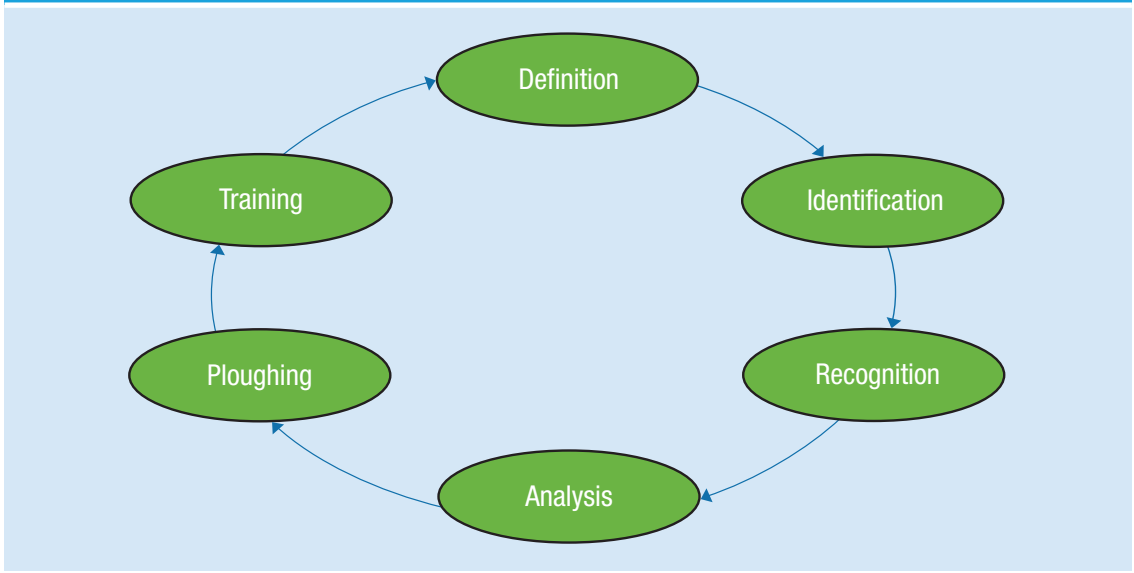
Implementation of big data involves a well-planned strategy for organizations to get the best out of it and make informed decisions that will guide their marketing drives. It can be seen as a six-stage cycle that involves definition, identification, recognition, analysis, ploughing back past experiences and training. These are critical stages to consider in big data implementation because each of them will help implementers to be focused on the expected result of the stage and the ultimate goal of the organization in implementing big data. This is illustrated by the acronym DIRAPT (**figure 1**). It should be seen as a cycle that every organization needs to repeat.

Definition of Scope

A major perception is that big data is seen as boundless. It is true that data generated by banks are enormous, ranging from daily interbranch transactions to messages sent or received via various media. The amount of data is determined by the number of branches, staff members and customers.

Sometimes, commercial banks, based on marketing focus and strategy, may decide to streamline the market in which they play. For instance, one commercial bank may be public-sector-focused while another may choose to focus on retail banking. Scope also invariably determines how large their customer base will be. For instance, banks whose strategy focuses on the retail sector will likely have more customers than those dealing with public or corporate sectors. Besides transactions emanating from various branches, there are interbank activities involving the movement of funds

Figure 1—DIRAPT Cycle



through alternative channels, such as automated teller machines (ATMs) and mobile banking, where large chunks of data are exchanged. It will not make any sense to want to treat all data from all fronts the same way and at the same time. It is necessary for banks to define the scope of big data implementation to be covered in order to get meaningful information from the data.

Identification of Skill Set

After developing a successful definition of boundaries in which to work, it is necessary to identify human resources needed with the required skill set. Careful selection of manpower with the requisite skills is very important before a successful big data implementation. Banks should note that this should not be seen as residing in only one department of the bank. Experienced staff should be picked from operations, marketing, control and other departments to contribute their input for successful implementation. A rich blend of skilled people will go a long way to determine the success of an implementation.

Recognition of Data Sources

Effective data tracking and measurement stem from identified data sources. It has been said that if it cannot be measured, then it cannot be controlled. Each data source must be listed, although not all of them can be handled at once. Then, based on the defined scope, data from the identified sources can be prepared for the next stage.

> “EFFECTIVE DATA TRACKING AND MEASUREMENT STEM FROM IDENTIFIED DATA SOURCES. IT HAS BEEN SAID THAT IF IT CANNOT BE MEASURED, THEN IT CANNOT BE CONTROLLED.”

Analysis of Output

Analysis is the stage where data within the scope are reviewed for relevant information for management use. Both structured and unstructured

data are involved. While the former is not much of an issue, the latter may require specialized analytics tools such as Hadoop, NoSQL, Jaspersoft, Pentaho, Karmasphere, Talend Studio and Skytree for analysis. A good analysis of data helps management make informed decisions to move the organization forward.

> A MAJOR CHALLENGE WITH BIG DATA IMPLEMENTATION IS COST, IN TERMS OF HUMAN RESOURCES AND INFRASTRUCTURE.

Ploughing Back Experience

Experience is a very important aspect of big data utilization. To begin, efforts should be made to get experienced personnel when implementing big data. Over time, experience gained can be expanded and reused. No two projects will be exactly the same, but experience gathered from a previous project can always be considered in subsequent projects. Therefore, it makes sense for experienced staff to be used in subsequent iterations for optimal results.

Training and Retraining

Training is a continuum. There should be regular training for bank workers involved in implementation before, during and after each cycle of implementation. Lessons learned at every stage should be well coordinated and recorded for reference purposes. Training should be encouraged at every stage of the DIRAPT cycle as well as at the end of each cycle.

Challenge With DIRAPT Cycle Approach

A major challenge with big data implementation is cost, in terms of human resources and infrastructure. Hiring people with relevant skills is always a Herculean task because knowledgeable personnel who can handle big data are scarce. On the technology side, the cost of getting appropriate tools to handle big data is also high. However, the DIRAPT cycle has brought in cloud solutions to help enterprises reduce the huge cost with the emergence of CSPs offering their services at reasonable prices.

Practical Approach to Big Data Implementation by CSPs

Commercial banks run their day-to-day activities using various software applications. Different segments of commercial banks (e.g., local operations, foreign operations, credit risk management, internal audit, internal control and information technology) run their applications in silos. Software contains useful information. Bringing together the data generated by all departments for the organization is always a challenge.

Recent technological advancements have helped bring big data into the cloud. Several CSPs now implement big data by using Software as a Service (SaaS). At the CSP end, activities are broken down into four major areas: the enablement, migration, integration and testing (EMIT) approach.[5] Most CSPs use the EMIT approach to handle big data.

Enablement

CSPs tend to enable business applications used by banks to operate on their own environment rather than the banks', which is basically on SaaS. At this

stage, processes and workflows are configured and allowed to run on the CSP environment as if they were offered to run the business applications for their clients—the commercial banks.

Migration
CSPs do a customized transfer of data to appropriately and cost-effectively move data to the cloud. This stage takes the place of the traditional data center and its infrastructures. By doing this, the cloud solution could help reduce costs.

Integration
The integration phase helps to incorporate both applications and data of the banks with the CSP's interface. Thereafter, a development area is prepared for banks to run. Banks first operate in the development area to properly assess the integration. At the bank's end, a user accesses departmental applications, through the interface provided, once a connection is established.

Testing
All business applications are tested before go-live. This includes interface testing, unit testing, stress testing, functional testing and performance testing of the various applications running on the cloud.

Conclusion

Big data is a concept that has received much publicity recently, possibly because of the intricacies involved. Efficient use of data in bank settings is very important to dictate how far and fast a commercial bank can go in the next few years.[6] Thus, there is a need for banks that hope to thrive to have a proper understanding of their data through a carefully selected big data implementation strategy.

The DIRAPT cycle spells out ways of implementing big data in commercial banks to help them enjoy the derivable benefits of big data, which include, but are not limited to, data security and competitive advantage. On one hand, proper implementation of big data will help banks to discover security risk regarding data exposure to fraudulent manipulations and then devise appropriate measures to mitigate them. Second, banks stand the chances of designing appropriate products for the appropriate environment with a view that can help them outwit their competitors.

Endnotes

1 Mukherjee, S.; R. Shaw; "Big Data—Concepts, Applications, Challenges and Future Scope," *International Journal of Advanced Research in Computer and Communication Engineering*, vol. 5, iss. 2, February 2016
2 Ramlukan, R.; "How Big Data and Analytics Can Transform the Audit," *Reporting*, iss. 9, April 2015
3 Watson, H.; "Big Data Analytics: Concepts, Technologies and Applications," *Communications of the Association for Information Systems*, vol. 34, 2014
4 Charles, V.; T. Gherman; "Achieving Competitive Advantage Through Big Data," *Middle-East Journal of Scientific Research*, vol. 16, iss. 8, 2013, p. 1069–1074
5 Orion, "Grow With a Connected Business," *www.orioninc.com/cloud/*
6 Ball, S.; "Gartner's 2017 BI Magic Quadrant: What Differentiates the Leaders?," *Better Buys*, 5 June 2017, *https://www.betterbuys.com/bi/gartners-2017-bi-magic-quadrant/*

Data Protection Tools

Do you have something to say about this article?

Visit the *Journal* pages of the ISACA® website *(www.isaca.org/journal)*, find the article and click on the Comments link to share your thoughts.

http://bit.ly/2BK5W7C

Data protection is critical, but it can sometimes be less visible than other types of security controls. This is because, very often, data protection failures are contributory rather than directly responsible for security failures. In other words, the direct cause of a high-profile breach may be something not related directly to data protection, but data protection often causes the impact to be significantly greater or the scope of compromise to be much broader because those controls are not in place.

To illustrate this point, consider a breach such as Equifax. By now, most professionals know that the direct cause of the Equifax breach was related to patch management—specifically, failure to patch Apache Struts. However, it is also true that data protection might have played a contributory role in the resultant scope and impact of that event. For example, would the breach have been as impactful if the data were encrypted? Would the actions of the attackers have been noticed in time for the security team to take action if better exfiltration alerts had been in place? We will never know the answers to these questions, but we can surmise that, had data protection measures been in place, at least some of the scope or impact might have been mitigated.

It behooves practitioners, therefore, to understand and employ data protection measures as a part of the security and assurance tasks they undertake. It is, of course, optimal when the organization's practitioners can invest in tools that directly support data protection measures. However, practitioners do not always find themselves in "optimal" situations; that is, organizations can directly and immediately invest in data protection measures only some of the time. But immediate benefit can be gained when practitioners can adapt investments in data protection goals or tools that already exist in the ecosystem. For the savvy practitioner, this represents a potential quick-win—an area where one can move an assurance or security goal forward based on investments the organization has already made.

There are, literally, hundreds (if not thousands) of tools that can be purchased, adapted or applied to forwarding data protection. The tools discussed here are a starting point—some that are useful to practitioners across a broad swath of industries, areas where one or more tool investments are likely to already exist in the ecosystem, and those that are likely to be useful regardless of whether the practitioner is an audit, risk or security professional.

Utimately, data protection should be thought through from the perspective of the goals the organization wants to accomplish. As practitioners do so, they may find opportunities such as those described here for adding value through the use of tools the organization is already using.

Ed Moyle
Is director of thought leadership and research at ISACA®. Prior to joining ISACA, Moyle was senior security strategist with Savvis and a founding partner of the analyst firm Security Curve. In his nearly 20 years in information security, he has held numerous positions including senior manager with CTG's global security practice, vice president and information security officer for Merrill Lynch Investment Managers, and senior security analyst with Trintech. Moyle is coauthor of Cryptographic Libraries for Developers and a frequent contributor to the information security industry as an author, public speaker and analyst.

1 Data Discovery and Inventory cont.

Tools can play a beneficial role in discovering and inventorying what data are in an organization and mapping out where the data are stored, processed and transmitted.

It should be stated explicitly that the specifics of the data a given organization might wish to locate will vary based on the organization itself. While discovery and inventory of sensitive data are important regardless, the specifics of the data (and, therefore, the specific tools an organization might employ to find the data) are different based on the type of organizations are and the specific considerations they have.

This means that tools that support discovery and inventorying of data can directly assist practitioners in a few ways:

1. By helping them verify that other controls are performing as expected
2. By building out a "map" of where sensitive data live throughout the organization

Once this is complete, the tools can also be run periodically in *ad hoc* fashion to find and flag situations where data have been stored or transmitted to an unexpected location.

It should be noted that there are a few different categories of tools that can help in this regard:

- Commercial data discovery tools, which assist organizations in finding, collecting and consolidating data stores for business intelligence or advanced analytics purposes
- Data leak prevention (DLP) tools, which can be used in an ongoing way to find and flag data that should not be stored or transmitted through certain channels based on business rules, and can help to prevent data exfiltration
- For practitioners on a budget, special purpose tools can be useful, (e.g., tools such as ccsrch (PANs), hashfind/passhunt (locates passwords and related artifacts), and grep/egrep when used in combination with specially-crafted regular expressions

2 Data Encryption

There are also tools that help practitioners encrypt data where the data are stored or transmitted. It should be noted that there are absolutely any number of special-purpose encryption tools out there that can be directly employed or adapted to encrypt data at any level of the Open Systems Interconnection (OSI) stack—at the application layer, at the file system layer, for data in transit, etc. It is always the better option to systematically and holistically address encryption use, applying it in combination with something like a formalized threat modeling exercise to protect against known, analyzed and thought-through threat scenario. Such systematic analysis is the ideal case.

However, because the ideal case is not always the actual case in every organization, it is worth noting that practitioners have data encryption options even in the absence of that broader investment. First and foremost, most modern operating systems have file system encryption options built into them. In combination with data discovery and a reliable inventory, tools are often available natively on the operating system platform used within the enterprise. This includes tools such as BitLocker (Windows), eCryptfs or LUKS (Linux), and others on a platform-by-platform basis. Likewise, database and middleware software can sometimes support encryption natively within it.

Many cloud service provider (CSP) storage and computing implementations make encryption of data at rest and in transit directly available to the customer in such a way that minimal additional overhead (other than checking the box) is required. The mechanics of enabling this depend on the CSP and the platform the organization employs, but almost all serious providers offer this for storage, Infrastructure as a Service (IaaS), and in application programming interfaces (APIs) or other services for Platform as a Service (PaaS) implementations.

3 Exfiltration

There are many tools that the organization may already have in place that can be used to detect and alert on potential exfiltration activity. Any network monitoring device (e.g., firewall or intrusion detection systems [IDS]) can potentially be adapted to help provide value for an exfiltration scenario. Firewalls or HTTP forward proxies can be employed to look for suspicious outbound connections (e.g., entities that are on IP black lists). IDS devices can do this and also potentially be adapted to trigger on custom regular expressions that might correspond to sensitive internal information.

One important thing to note is that, for the purposes of exfiltration, many attackers will employ encrypted channels such as Transport Layer Security (TLS), Secure Shell (SSH) or even nonstandard encrypted communications techniques.

Therefore, while potentially a valuable addition, it cannot be assumed that an IDS (monitoring, as it does, plaintext traffic) will necessarily always be able to detect this activity. As such, keeping an eye out for suspicious connections is a useful step, whether or not the organization also employs an IDS to detect exfiltration.

Do you have something to say about this article?

Visit the *Journal* pages of the ISACA® website *(www.isaca.org/journal)*, find the article and click on the Comments link to share your thoughts.

http://bit.ly/2B0Ktus

Q The European Union (EU) General Data Protection Regulation (GDPR) will take effect in May 2018. My organization is not doing any business in Europe currently, but we have plans to expand. How will GDPR affect us? Will it affect us if we do not have an office in Europe?

A GDPR (Regulation [EU] 2016/679) is a regulation by which the European Parliament, the Council of the European Union and the European Commission intend to strengthen and unify data protection for all individuals within the European Union. It also addresses the export of personal data outside the European Union for processing. The primary objective of the GDPR is to give assurance to EU citizens that their personal data are processed in a secure environment and have adequate legal protection. The regulation was adopted on 27 April 2016. It will be enforceable beginning on 25 May 2018 after a two-year transition period. Unlike a directive, it does not require any enabling legislation to be passed by national governments; thus, it is directly binding and applicable.

An important aspect of GDPR is that it applies to organizations that are not part of the European Union but collect and/or process personal data of EU residents outside the EU's geographical boundaries. Although the regulation does not apply to the processing of personal data for national security activities or law enforcement, it has a separate Data Protection Directive for the police and criminal justice sector that provides rules on personal data exchanges at the national, European and international levels.

The term "personal data" refers to any information relating to an individual's private, professional or public life. It includes the individual's name, home address, photo, email address, bank details, posts on social networking websites, medical information, or IP address of a personal device.[1]

The new EU data protection regime extends the scope of the EU data protection law to all foreign organizations processing the data of EU residents. It provides for a harmonization of the data protection regulations throughout the European Union, thereby making it easier for non-European companies to comply with these regulations. However, this comes at the cost of a strict data protection compliance regime with severe penalties of up to 4 percent of worldwide turnover or upper limit of £20 million, whichever is higher.[2]

Some facts about GDPR include:[3]

- If a business is not in the European Union, it still must to comply with the regulation if it is processing the personal data of EU citizens outside the EU's geographical boundaries.
- The definition of "personal data" is broader, bringing more data into the regulated perimeter.
- Consent will be necessary for processing children's data.
- The rules for obtaining valid consent have been changed.
- The appointment of a data protection officer (DPO) will be mandatory for certain organizations.
- Mandatory data protection impact assessments (technical as well as privacy assessments) have been introduced.
- There are new requirements for data breach notifications.
- Data subjects have the following rights:
  - Right to be forgotten/erasure
  - Right to access
  - Right to rectification
  - Right to object
- There are new restrictions on international data transfers.
- Data processors share responsibility for protecting personal data, along with the data controller.
- There are new requirements for data portability.
- Processes must be built on the principle of privacy by design.
- The GDPR is a one-stop shop.

Sunil Bakshi, CISA, CRISC, CISM, CGEIT, ABCI, AMIIB, BS 25999 LI, CEH, CISSP, ISO 27001 LA, MCA, PMP
Has worked in IT, IT governance, IS audit, information security and IT risk management. He has 40 years of experience in various positions in different industries. Currently, he is a freelance consultant and visiting faculty member at the National Institute of Bank Management, India.

Therefore, if an organization needs to process EU citizens' personal data while offering goods or services to those citizens (data subjects), the organization is a data controller and is subject to GDPR regulation, regardless of whether the organization is based inside or outside the European Union. If an organization does business with data subjects in the European Union, the organization needs to comply with the EU regulation, even if the services or goods provided do not involve connecting to a payment system.

According to some experts, the wording "data subjects who are in the European Union" indicates that the GDPR covers the data of EU citizens as well as temporary residents and even those on vacation.[4] It may also be interpreted that EU citizens traveling outside the European Union whose data are collected and processed outside the European Union may not be subject to GDPR regulation.[5]

The "territorial scope" clause does not mean that every single web-based business that is accessible from within the European Union is in scope of the GDPR. The fact that someone in the European Union can visit an organization's website does not automatically bring that website into "territorial scope." The website has to be doing something to actively reach out to someone in the European Union.[6]

An organization may be exempt from GDPR compliance in the following situations:[7]

- If the processing of personal data from EU-based data subjects is occasional or not on a large scale
- If the personal data to be processed do not include special categories of personal data or relate to criminal convictions and offenses
- If the nature, context, scope and purposes of the processing are unlikely to result in a risk to the rights and freedoms of the data subject

In all other cases where the "territorial scope" extends to non-EU-based organizations that process the data of persons from the European Union (data controllers), organizations will need to nominate an EU representative within the European Union.[8] Some conditions relating to that representative include:

- The EU representative is the first point of contact for the data protection supervisory authorities and data subjects. The contact information for the EU representative of an organization and its contact information must appear on the organization's website along with terms of service between the organization and the EU representative.
- The EU representative must be located in a member state in which the organization's EU data subjects are based. If the organization targets the entire European Union, then the EU representative may be based in any member country.
- The EU representative must operate under the organization's direction and the directions must be in writing. The legislation does not specifically state "under contract," but it is fairly safe to assume it may be in the contract.
- The EU representative will be designated "without prejudice" to legal actions that may be taken against the data controller or the data processor. An organization cannot outsource accountability toward the data subjects to its EU representative.

Author's Note

The views expressed here are the author's. For more detailed understanding, consult a legal expert.

Endnotes

1 For this definition and definitions of other pertinent terms, such as "processing," "restriction of processing," "profiling" and "pseudonymization" (some experts refer to it as "tokenization"), see *https://gdpr-info.eu/art-4-gdpr/*.
2 *IT Governance, Data Protection Act (DPA)* and *EU GDPR Penalties*, *https://www.itgovernance.co.uk/dpa-and-dgpr-penalties*
3 IT Governance, The EU General Data Protection Regulation, *https://www.itgovernance.co.uk/data-protection-dpa-and-eu-data-protection-regulation*
4 Murphy, M.; "Rules of Establishment for European Data Controllers and Data Processors," Safe Data Matters, 3 August 2016, *http://safedatamatters.com/gdpr-datacontrollerspart2-establishment/*
5 *Ibid*.
6 *Ibid*.
7 *Ibid*.
8 *Ibid*.

CROSSWORD PUZZLE

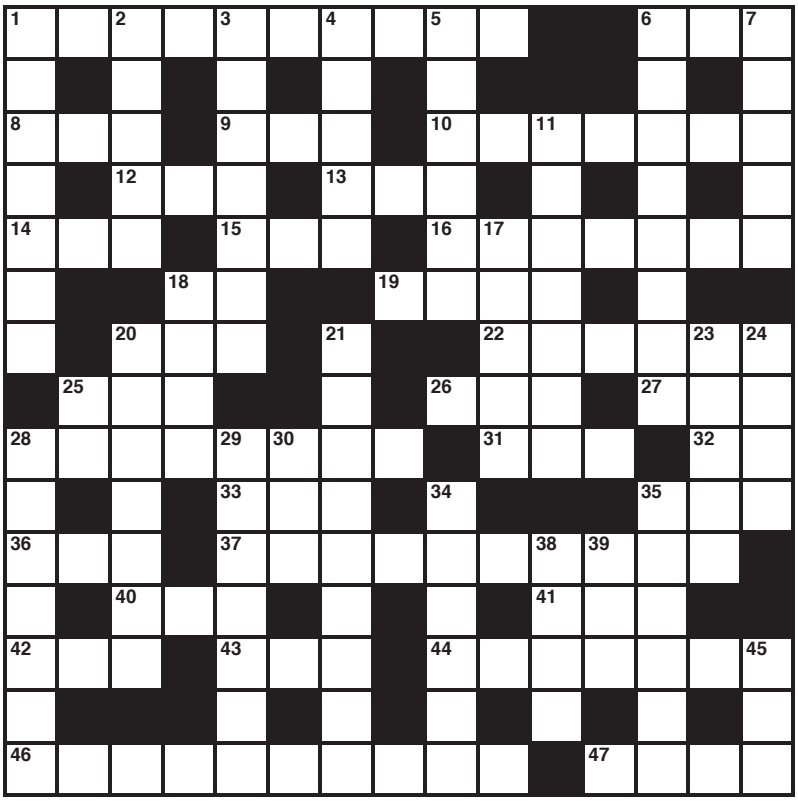
by Myles Mellor
www.themecrosswords.com

ACROSS

1. Type of shared site
6. Protocol for data transfer
8. Contract between service provider and end user
9. Berners-Lee invention
10. Sources of danger
12. Computer programming language for statistical analysis, abbr.
13. Lawyer's org.
14. Bit of binary code
15. Have legal control of
16. Incomprehensible
18. Milliliter, abbr.
19. Geometric art style
20. Spelling contest
22. Destroyed
25. Memory
26. Intersected
27. NASA term for a spacewalk, abbr.
28. Restoring a system after an attack or crash
31. Establish
32. CISAs work with this department during audits
33. Tick off
35. ___ standstill (motionless)
36. Mainframe component
37. Malicious software that demands payment with a threat
40. Placed
41. Web address ender
42. Early operating system
43. Single- prefix
44. Containing errors, as a file or system
46. Making something known
47. The heart of a Ted Talk

DOWN

1. Physical control
2. Contract for temporary use
3. Arrangement of displays to monitor a system
4. Preeminent industrialist
5. Period of time when a system is nonfunctioning
6. Stop functioning, as a system
7. Presented, as a problem
11. Create a new path for



17. Land areas
18. Office message
20. Data duplications for security
21. Science of investigation of evidence found in computer systems
23. Online party announcement
24. Information
25. Memo start
28. Diminished gradually
29. The V in VM
30. Historical chapter
34. Agree
35. Contended
38. Standalone malware computer program that replicates itself
39. Flightboard abbreviation
45. Drink that is served hot or cold

Answers on page 58

CPE QUIZ
#176

Based on Volume 5, 2017—Enabling the Speed of Business
Value—1 Hour of CISA/CRISC/CISM/CGEIT Continuing Professional Education (CPE) Credit

Take the quiz online.

http://bit.ly/2AkgMor

TRUE OR FALSE

SATHIYAMURTHY ARTICLE

1. Despite its explicit support of privacy and data protection by design as a legal obligation, the EU General Data Protection Regulation (GDPR) is not among the most commonly used frameworks for managing privacy, according to ISACA's 2014 Privacy Survey.
2. Human biases and national sensitivities have an impact on the definition of private information and associated privacy expectations.
3. When organizations are creating products and solutions and are faced with prioritizing privacy protection against business objectives (such as speed to market), they will generally place a higher priority on business objectives.

CARON ARTICLE

4. Immutability, or a single version of the truth, is a benefit to blockchain technology because it contributes to reducing the asymmetry between the networked entities engaged in a transaction.
5. Storage of the cryptographic keys that are used as digital signatures constitutes a security weakness for blockchain.
6. In the United States, blockchain solutions that are used for sharing and recording patient records are not subject to the US Health Insurance Portability and Accountability Act (HIPAA).

KELLY ARTICLE

7. Executive buy-in is critical to an enterprisewide security program, and one way to appeal to executives is by communicating the program's benefits in terms of their specific pain points.
8. Although it is important to raise security awareness among employees, it is not as important as focusing on external malicious actors, because less than half of cyberattacks are carried out by insiders, according to the IBM X-Force 2016 Cyber Security Intelligence Index.
9. An acceptable-use policy should clearly define who owns mobile devices and what access the enterprise has to the data on those devices.

BRADFORD AND HENDERSON ARTICLE

10. Research indicates that barriers to a system's use dissuade users, but when those barriers disappear, their absence encourages use.
11. According to the research presented in the article, financial auditors find generalized audit software (GAS) much easier to use than do other auditors.
12. Two factors that encourage the use of GAS are perceived ease of use and perceived usefulness.
13. The authors' research indicates that internal auditors perceive less threat from using GAS than do external auditors.

WERNEBURG ARTICLE

14. The impact that data breaches have on the individuals whose personal information is compromised is considered a secondary risk factor, rather than a proximate risk factor.
15. Regulatory requirements and service audit systems such as Service Organization Control (SOC) 2 specify that an application vulnerability scan should be performed at least annually and the report shared with the client.
16. Budget, differing priorities and conflicting objectives are just a few reasons a client may not want to deal with security fixes.

MUKUNDHAN ARTICLE

17. According to PricewaterhouseCoopers (PwC), between 30 and 50 billion devices will be connected to the Internet by the year 2020.
18. When a client requests a connection to the server via a SYN message and the server responds with ACK, that is a Transmission Control Protocol (TCP) three-way handshake.
19. The 2016 coordinated distributed denial-of-service (DDoS) attack on DYN was initiated by the Mirai botnet—a cluster of approximately 100,000 enslaved Internet of Things (IoT) devices delivering different types of DDoS attacks.
20. IoT devices are widely known to take a strong stance on security, which they demonstrate through such regular practices as unguessable passwords, secured ports and regularly updated firmware.

CPE QUIZ
#176

THE ANSWER FORM

Based on Volume 5, 2017

TRUE OR FALSE

SATHIYAMURTHY ARTICLE

1. _____
2. _____
3. _____

CARON ARTICLE

4. _____
5. _____
6. _____

KELLY ARTICLE

7. _____
8. _____
9. _____

BRADFORD AND HENDERSON ARTICLE

10. _____
11. _____
12. _____
13. _____

WERNEBURG ARTICLE

14. _____
15. _____
16. _____

MUKUNDHAN ARTICLE

17. _____
18. _____
19. _____
20. _____

Please confirm with other designation-granting professional bodies for their CPE qualification acceptance criteria. Quizzes may be submitted for grading only by current *Journal* subscribers. An electronic version of the quiz is available at *www.isaca.org/cpequiz*; it is graded online and is available to all interested parties. If choosing to submit using this print copy, please email, fax or mail your answers for grading. Return your answers and contact information by email to info@isaca.org or by fax to +1.847.253.1755. If you prefer to mail your quiz, in the US, send your CPE Quiz along with a stamped, self-addressed envelope, to ISACA International Headquarters, 3701 Algonquin Rd., #1010, Rolling Meadows, IL 60008 USA. Outside the US, ISACA will pay the postage to return your graded quiz. You need only to include an envelope with your address. You will be responsible for submitting your credit hours at year-end for CPE credits. A passing score of 75 percent will earn one hour of CISA, CRISC, CISM or CGEIT CPE credit.

Name _____

PLEASE PRINT OR TYPE

Address _____

CISA, CRISC, CISM or CGEIT # _____

Answers: Crossword by Myles Mellor
See page 56 for the puzzle.

ISACA Member and Certification Holder Compliance

The specialized nature of information systems (IS) audit and assurance and the skills necessary to perform such engagements require standards that apply specifically to IS audit and assurance. The development and dissemination of the IS audit and assurance standards are a cornerstone of the ISACA® professional contribution to the audit community.

IS audit and assurance standards define mandatory requirements for IS auditing. They report and inform:

- IS audit and assurance professionals of the minimum level of acceptable performance required to meet the professional responsibilities set out in the ISACA Code of Professional Ethics
- Management and other interested parties of the profession's expectations concerning the work of practitioners
- Holders of the Certified Information Systems Auditor® (CISA®) designation of requirements. Failure to comply with these standards may result in an investigation into the CISA holder's conduct by the ISACA Board of Directors or appropriate committee and, ultimately, in disciplinary action.

ITAF™, 3rd Edition

(www.isaca.org/itaf) provides a framework for multiple levels of guidance:

IS Audit and Assurance Standards

The standards are divided into three categories:

- **General standards (1000 series)**—Are the guiding principles under which the IS assurance profession operates. They apply to the conduct of all assignments and deal with the IS audit and assurance professional's ethics, independence, objectivity and due care as well as knowledge, competency and skill.
- **Performance standards (1200 series)**—Deal with the conduct of the assignment, such as planning and supervision, scoping, risk and materiality, resource mobilization, supervision and assignment management, audit and assurance evidence, and the exercising of professional judgment and due care.
- **Reporting standards (1400 series)**—Address the types of reports, means of communication and the information communicated.

Please note that the guidelines are effective 1 September 2014.

General
1001 Audit Charter
1002 Organizational Independence
1003 Professional Independence
1004 Reasonable Expectation
1005 Due Professional Care
1006 Proficiency
1007 Assertions
1008 Criteria

Performance
1201 Engagement Planning
1202 Risk Assessment in Planning
1203 Performance and Supervision
1204 Materiality
1205 Evidence
1206 Using the Work of Other Experts
1207 Irregularity and Illegal Acts

Reporting
1401 Reporting
1402 Follow-up Activities

IS Audit and Assurance Guidelines

The guidelines are designed to directly support the standards and help practitioners achieve alignment with the standards. They follow the same categorization as the standards (also divided into three categories):

- General guidelines (2000 series)
- Performance guidelines (2200 series)
- Reporting guidelines (2400 series)

General
2001 Audit Charter
2002 Organizational Independence
2003 Professional Independence
2004 Reasonable Expectation
2005 Due Professional Care
2006 Proficiency
2007 Assertions
2008 Criteria

Performance
2201 Engagement Planning
2202 Risk Assessment in Planning
2203 Performance and Supervision
2204 Materiality
2205 Evidence
2206 Using the Work of Other Experts
2207 Irregularity and Illegal Acts
2208 Sampling

Reporting
2401 Reporting
2402 Follow-up Activities

IS Audit and Assurance Tools and Techniques

These documents provide additional guidance for IS audit and assurance professionals and consist, among other things, of white papers, IS audit/assurance programs, reference books and the COBIT® 5 family of products. Tools and techniques are listed under *www.isaca.org/itaf*.

An online glossary of terms used in ITAF is provided at *www.isaca.org/glossary*.

Prior to issuing any new standard or guideline, an exposure draft is issued internationally for general public comment.

Comments may also be submitted to the attention of the Director, Thought Leadership and Research via email (standards@isaca.org); fax (+1.847.253.1755) or postal mail (ISACA International Headquarters, 3701 Algonquin Road, Suite 1010, Rolling Meadows, IL 60008-3105, USA).

Links to current and exposed ISACA Standards, Guidelines, and Tools and Techniques are posted at *www.isaca.org/standards*.

Disclaimer: ISACA has designed this guidance as the minimum level of acceptable performance required to meet the professional responsibilities set out in the ISACA Code of Professional Ethics. ISACA makes no claim that use of these products will assure a successful outcome. The guidance should not be considered inclusive of any proper procedures and tests or exclusive of other procedures and tests that are reasonably directed to obtaining the same results. In determining the propriety of any specific procedure or test, the control professionals should apply their own professional judgment to the specific control circumstances presented by the particular systems or IS environment.

ISACA® Journal, formerly *Information Systems Control Journal*, is published by the Information Systems Audit and Control Association® (ISACA®), a nonprofit organization created for the public in 1969. Membership in the association, a voluntary organization serving IT governance professionals, entitles one to receive an annual subscription to the *ISACA Journal*.

ISSN 1944-1967

Subscription Rates:

US:
one year (6 issues) $75.00

All international orders:
one year (6 issues) $90.00.

Remittance must be made in US funds.

ADVERTISERS/ WEBSITES

leaders and supporters

ISACA®

ISACA BOOKSTORE

RESOURCES FOR YOUR PROFESSIONAL DEVELOPMENT

www.isaca.org/bookstore

NEW! Online Review Courses

Get the training you need. Prepare to obtain your CISA, CRISC or CISM certification and be recognized among the world's most-qualified information systems professionals. ISACA's Online Review Courses provide internet accessible, on-demand instruction and are ideal for preparing you and fellow audit, assurance, control, security and cyber security professionals for ISACA's certification exams.

Visit: www.isaca.org/examonlinereview to learn more.

CISA® CRISC™ CISM® CGEIT®

Browse a variety of publications featuring the latest research and expert thinking on standards, best practices, emerging trends and more at ***isaca.org/bookstore***

Featured Exam Prep Materials

CISA® Review Manual, 26th Edition

The *CISA® Review Manual, 26th Edition* is a comprehensive reference guide designed to help individuals prepare for the CISA exam and understand the roles and responsibilities of an information systems (IS) auditor. The manual has been revised according to the 2016 CISA Job Practice and represents the most current, comprehensive, peer-reviewed IS audit, assurance, security and control resource available.

The 26th edition is organized to assist candidates in understanding essential concepts and studying the following job practice areas: The Process of Auditing Information Systems; Governance and Management of IT; Information Systems Acquisition, Development and Implementation; Information Systems Operations, Maintenance and Service Management; Protection of Information Assets.

The manual also serves as an effective desk reference for IS auditors.

Member: US $105.00
Non-member: US $135.00
Print Product Code: CRM26ED
eBook Product Code: EPUB_CRM26ED

BESTSELLING PRODUCT

CISA® Review Questions, Answers & Explanations Database—12-Month Subscription

The *CISA® Review Questions, Answers & Explanations Database* is a comprehensive 1,000-question pool of items that combines the questions from the *CISA® Review Questions, Answers & Explanations Manual, 11th Edition.* The database has been revised according to the recently updated 2016 CISA Job Practice.

The database is available via the web, allowing CISA Candidates to log in at home, at work or anywhere they have Internet connectivity. This database is MAC and Windows compatible.

Exam candidates can take sample exams with randomly selected questions and view the results by job practice domain, allowing for concentrated study in particular areas. Additionally, questions generated during a study session are sorted based on previous scoring history, allowing CISA candidates to identify their strengths and weaknesses and focus their study efforts accordingly.

Member: US $185.00
Non-member: US $225.00
Product Code: XMXCA15-12M

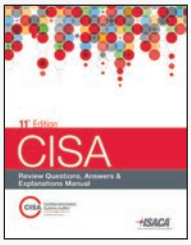
The *CISA® Review Questions, Answers & Explanations Database* is also available on CD-Rom in Spanish.

CISA® Review Questions, Answers & Explanations Manual, 11th Edition

Designed to familiarize candidates with the question types and topics featured in the CISA exam, the *CISA® Review Questions, Answers & Explanations Manual, 11th Edition* consists of 1,000 multiple-choice study questions that have previously appeared in the *CISA® Review Questions, Answers & Explanations Manual 2015* and the *CISA® Review Questions, Answers & Explanations Manual 2015 Supplement.* The manual has been updated according to the newly revised 2016 Job Practice.

Many questions have been revised or completely rewritten to be more representative of the CISA exam question format and/or to provide further clarity or explanation of the correct answer. These questions are not actual exam items but are intended to provide CISA candidates with an understanding of the type and structure of questions and content that have previously appeared on the exam. This publication is ideal to use in conjunction with the:

- *CISA® Review Manual, 26th Edition*
- *CISA® Review Questions, Answers & Explanations Database – 12 Month Subscription*



Member: US $120.00
Non-member: US $156.00
Product Code: QAE11ED

Available in: Chinese Simplified, Italian, Japanese, and Spanish

CRISC™ Review Manual, 6th Edition

The *CRISC™ Review Manual, 6th Edition* is a comprehensive reference guide designed to help individuals prepare for the CRISC exam and understand IT-related business risk management roles and responsibilities. The manual has been enhanced over the past editions and represents the most current, comprehensive, peer-reviewed IT-related business risk management resource available worldwide.

The 6th edition manual is organized to assist candidates in understanding essential concepts and studying the following job practice areas:

- IT Risk Identification
- IT Risk Assessment
- Risk Response and Mitigation
- Risk and Control Monitoring and Reporting

Member: US $85.00
Non-member: US $115.00
Print Product Code: CRR6ED
eBook Product Code: EPUB_CRR6ED

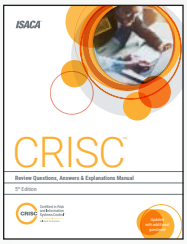
Order online at ***www.isaca.org/bookstore***

NEW!

CRISC™ Review Questions, Answers & Explanations Manual, 5th Edition

The *CRISC™ Review Questions, Answers & Explanations Manual, 5th Edition* has been expanded and updated to include even more practice questions. This study aid is designed to familiarize candidates with the question types and topics featured in the CRISC exam with the use of 550 questions.

Many questions have been revised or completely rewritten to be more representative of the current CRISC exam question format, and/or to provide further clarity or explanation of the correct answer. These questions are not actual exam items, but are intended to provide CRISC candidates with an understanding of the type and structure of questions and content that have previously appeared on the exam.

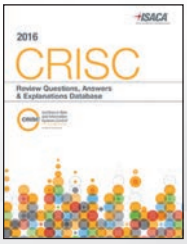


Member: US $72.00
Non-member: US $96.00
Product Code: CRQ5ED

CRISC™ Review Questions, Answers & Explanations Database—12-Month Subscription

The *CRISC™ Practice Question Database* is a comprehensive 500-question pool of items that contains the questions from the *CRISC™ Review Questions, Answers & Explanations Manual, 4th Edition.* The database is available via the web, allowing CRISC candidates to log in at home, at work or anywhere they have Internet connectivity. The database is MAC and Windows compatible.

Exam candidates can take sample exams with randomly selected questions and view the results by job practice domain, allowing for concentrated study in particular areas. Additionally, questions generated during a study session are sorted based on previous scoring history, allowing CRISC candidates to identify their strengths and weaknesses and focus their study efforts accordingly.



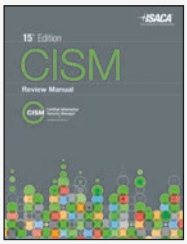
Member: US $185.00
Non-member: US $225.00
Product Code: XMXCR14-12M

CISM® Review Manual, 15th Edition

The *CISM® Review Manual, 15th Edition* is designed to help you prepare for the CISM® exam. This comprehensive, easy-to-navigate manual is organized into chapters that correspond to the four job practice areas covered in the CISM exam. The Manual is primarily designed as a tool for exam prep, but can also be useful as a reference manual for information security managers.

New to the 15th Edition:

- **In Practice Questions** help you explore the concepts in the CISM Review Manual in your own practice.
- **Knowledge Checks** are designed to help reinforce important concepts from the Review Manual to further enhance your learning.
- **Case Studies** provide real-world scenarios to help you gain a practical perspective on the Review Manual content and how it relates to the CISM's practice.
- **Comprehensive Index** has been updated to make navigating the Review Manual easier and more intuitive.



Member: US $105.00
Non-member: US $135.00
Print Product Code: CM15ED
eBook Product Code: EPUB_CM15ED

CISM® Review Questions, Answers & Explanations Manual, 9th Edition

The *CISM® Review Questions, Answers & Explanations Manual, 9th Edition* consists of 1,000 multiple-choice study questions, answers and explanations, which are organized according to the CISM job practice domains.

The questions, answers and explanations are intended to introduce the CISM candidate to the types of questions that appear on the CISM exam. This publication is ideal to use in conjunction with the *CISM Review Manual 15th Edition.*

Member: US $120.00
Non-member: US $156.00
Product Code: CQA9ED

NEW!

CISM® Review Questions, Answers & Explanations Database—12-Month Subscription

The CISM® Review Questions, Answers & Explanations Database is a comprehensive 1,000-question pool of items that contains the questions from the *CISM® Review Questions, Answers & Explanations Manual 9th Edition.*

The database is available via the web, allowing our CISM candidates to log in at home, at work or anywhere they have Internet connectivity. The database is MAC and Windows compatible.

Exam candidates can take sample exams with randomly selected questions and view the results by job practice domain, allowing for concentrated study in particular areas. Additionally, questions generated during a study session are sorted based on previous scoring history, allowing CISM candidates to identify their strengths and weaknesses and focus their study efforts accordingly.

Member: US $185.00
Non-member: US $225.00
Product Code: XMXCM15-12M

CGEIT® Review Manual, 7th Edition

The *CGEIT® Review Manual, 7th Edition* is designed to help individuals prepare for the CGEIT exam and understandthe responsibilities of those who implement or manage the governance of enterprise IT (GEIT) or have significant advisory or assurance responsibilities in regards to GEIT. It is a detailed reference guide that has been developed and reviewed by subject matter experts actively involved in governance of enterprise IT worldwide.

The manual is organized to assist candidates in understanding essential concepts and studying the following updated job practice areas:

- Framework for the governance of enterprise IT
- Strategic management
- Benefits realization
- Risk optimization
- Resource optimization

Member: US $85.00
Non-member: US $115.00
Print Product Code: CGM7ED
eBook Product Code: EPUB_CGM7ED

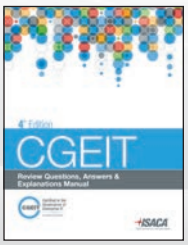
CGEIT® Review Questions, Answers & Explanations Manual, 4th Edition

The *CGEIT® Review Questions, Answers & Explanations Manual, 4th Edition* is designed to familiarize candidates with the question types and topics featured in the CGEIT exam.

The 250 questions in this manual have been consolidated from the *CGEIT® Review Questions, Answers & Explanations Manual, 2015* and the *CGEIT® Review Questions, Answers & Explanations Manual, 2015 Supplement.*

Many questions have been revised or completely rewritten to be more representative of the CGEIT exam question format and/or to provide further clarity or explanation of the correct answer. These questions are not actual exam items but are intended to provide CGEIT candidates with an understanding of the type and structure of questions and content that has previously appeared on the exam. This publication is ideal to use in conjunction with the:

- *CGEIT® Review Manual, 7th Edition*



Member: US $60.00
Non-member: US $75.00
Product Code: CGQ4ED

CSX™
CYBERSECURITY NEXUS

CYBER SECURITY TRAINING JUST GOT REAL

NOW YOUR STAFF CAN COMBAT REAL THREATS IN REAL TIME TO BUILD REAL TECHNICAL SKILLS.

Yesterday's lecture-based cyber security training won't protect your organization against tomorrow's advanced cyberthreats. That's why ISACA's new Cybersecurity Nexus™ (CSX) Enterprise Training Platform offers your security team:

- **On-demand access to 200+ hours of training for less than the cost of one typical course**
- **Practical, hands-on training labs performed in a live, dynamic network environment**
- **Continually updated content based on the latest real-world threats and scenarios**
- **Performance-based assessment of current and prospective employees' technical skills**

SCHEDULE A DEMO OF THE CSX TRAINING PLATFORM AT WWW.ISACA.ORG/CSXCYBERTRAINING

ISACA®

Does your audit software improves your productivity & efficiency or it slows you down?

R-CAP™ leverages on the latest technology & offers:

▶ Mobility ▶ Collaboration ▶ Smart Editing

R-CAP | Audit Life-Cycle and Risk Management Solution

works with Google Drive

salesforce available on AppExchange

- ✓ Observations Tracking
- ✓ Risk and Controls Matrix
- ✓ Regular Business Monitoring
- ✓ Audit Timesheet Management
- ✓ Insightful Dashboards & Reports
- ✓ Efficient Work-Paper Documentation

Built by Auditors, for Auditors

For your free 30 day trial email at **contactus@rcap.online**

www.rcap.online

TRONIXSS