Challenges in Protecting Data for Modern Enterprises

Mohammad M. Nur and Houssain Kettani

***Abstract*—Advances in data analytics have made data valuable in getting business intelligence about sales, marketing and customer service. Enterprises are increasingly collecting and storing more and more data about customers and their behavior. Consequently, data breaches have evolved drastically in recent years and have become one of the key challenges faced by the organizations. A data breach typically indicates other defense mechanisms and security practices have failed and attackers have been successful in stealing data by exploiting vulnerabilities. Software bugs, misconfigurations, unpatched security bugs or employees not following security practices properly can cause these failures. Having an in-depth understanding of the points of failures and identifying the effectiveness of the defense methods are crucial factors to fight against data breaches and to minimize the aftermath impacts. This paper studies the characteristics of the exploits from recent major data breaches, evaluates the available mitigations, their effectiveness, then explores data centric security strategy and the challenges in implementing them in enterprises.**

***Index Terms*—Data security, data breach, defense-in-depth, data encryption, security at rest, challenges in data security.**

I. Introduction

Data breach generally refers to an unauthorized exposure, disclosure or loss of an organization's sensitive information, which may include its financial state, its future plans and products, its customers, and its partners. This can have long term direct and indirect impacts, depending on the nature of the stolen data. Organizations in both public and private sectors have been the victim of data breaches. Health care sector, small businesses and multinational companies have become the top targets of data breaches in 2017 [1]. Heath care industries have been a lucrative target as they have a lot of sensitive data about a large number of the population [2]. Small businesses lack the budget and resources in hardening security controls and hence they have been an easy target. Finally, multinational companies have data about users across the world and attracted hackers around the world due to its high return on investment. Protecting customer data is a legal responsibility for organizations. US Privacy Act of 1974, California Online Privacy Protection Act (CalOPPA), Customer Proprietary Network Information (CPNI), Payment Card Industry Data Security Standard (PCI DSS), Health Insurance Portability & Accountability Act (HIPAA), Federal Information Security Management Act (FISMA) are some examples of laws and regulations that mandate organizations to protect customer data. Some of these regulations are industry specific, for example, CPNI applies to telecom industry and focuses on information related to user activities on phone (for example, call detail records, web viewing activities, etc.).

Data security refers to the processes of protecting data from unauthorized access and use throughout its lifecycle by securing the tools and platforms used for processing, storing and visualizing data. Since data breach can occur in many ways, through cyberattacks or through insider attacks, it is crucial to maximize end-to-end security, and a defense-in-depth strategy with many security controls applied in a layered approach – firewalls, intrusion detection, access control, secured channel of communication, etc. can help. Data centric security, often referred to as data security, can give an organization a good level of protection if executed correctly. This approach involves classification of data and then redaction, tokenization and/or encryption of data classified as sensitive. While data centric security cannot stop actual data breaches, it can definitely minimize the aftermath impact.

Encryption is the core component of data centric security. The success of data level encryption depends on two factors; first, efficient data modeling, without that encryption may introduce performance overhead; second, support from the tools and data platforms. Not all tools and platforms come with the right level of security – some of them have strong focus on performance and scalability but a limited focus on security. Platforms also have their own way of implementing data security. The lack of a standard way of protecting data can make the development and maintenance of the encryption framework complex and expensive. Protecting the encryption keys and managing the access to those keys by users and applications can become another challenge. Unstructured and large size data like transcripts, messages and photos can add additional challenges to encryption framework. Evolution of cloud-based services has introduced another dimension of challenges - being managed service and providing limited access to some desired security controls, they are adding more complexities in achieving data centric security strategy.

The purpose of this study is to identify the attack vectors from recent data breaches, understand the effectiveness of different defense mechanisms against those attack vectors, explore data centric security and challenges in implementing that to protect data. In the next section we discuss major data breaches that occurred recently and various attack vectors. In Section III, we present an overview of traditional defenses. In Section IV, we discuss human factor in cyber defense, which is often the weakest link in cyber security. In Section V, we discuss data centric security approach to prevent the data from being exposed to the attacker. In Section VI, we present some challenges that need to be dealt with when implementing data level encryption. Finally, concluding

Manuscript received November 28, 2019; revised February 22, 2020.

The authors are with the Beacom College of Computer & Cyber Sciences, Dakota State University, Madison, South Dakota 57042, USA (e-mail: houssain.kettani@dsu.edu).

doi: 10.18178/joebm.2020.8.2.615

remarks are presented in Section VII.

II. Data Breaches and Attack Vectors

Data breaches were one of the top fifteen cybersecurity threats identified in the ENISA Threat Landscape Report 2017 published by the European Union Agency for Network and Information Security (ENISA) [1]. The report identified weak or stolen credentials, phishing attacks and SQL injection attacks as the most common attack vectors. Open Web Application Security Project (OWASP) foundation published a list of top ten attack vectors for web applications observed in 2017 – injection, broken authentication, sensitive data exposure, XML external entities, broken access control, security misconfiguration, cross-site scripting, insecure deserialization, using components with known vulnerabilities, insufficient logging and monitoring [3]. Phishing attacks were also a common attack vector. Re-identification attacks is another attack vector where individuals can be identified through correlation of big data or confidential and publicly available data [4]. With the evolution of social engineering, this kind of attack is getting popular among the hackers, especially to develop phishing attacks. Table I, compiled from data presented in [5]-[7], captures the impact and the attack vector of some major data breaches occurred in last five years in the United States. The common attack vectors here are phishing attacks, unpatched software, stolen login credentials, credentials stored in source code, and malware. Note that some of them may be interrelated, for example, a phishing attack might have been used to steal credentials or to install a malware. This gives us a good idea to design a defense strategy that can prevent data breaches or minimize the aftermath.

TABLE I: Major Data Breaches in Last Five Years

Company	*Timeline (Approx.)*	*Affected Users*	*Type of Stolen Data*	*Attack Vectors (Published/ Suspected)*
Yahoo	2012-2014	3 Billion	Customer data and logon credentials	Phishing, Malware
Adult Friend Finder	Oct 2016	412 Million	Customer data and logon credentials	Local File Inclusion (LFI) Vulnerability
eBay	May 2014	145 Million	Customer data and logon credentials	Stolen login credential
Equifax	Mar 2017	143 Million	Customer data and logon credentials	Unpatched software
Target	Dec 2013	110 Million	Customer and credit card data	Breach through third party vendor
JP Morgan Chase	July 2014	83 Million	Customer data	Stolen login credential
Anthem	Feb 2015	78.8 Million	Customer data	Phishing
Uber	Late 2016	57 Million	Customer data	Stolen login credential from GitHub
Home Depot	Sept 2014	56 Million	Customer data	Stolen login credential, Malware
US Office of Personnel Management (OPM)	2012 -2014	22 Million	Employee data, security clearance, background check data and fingerprint	Breach though third-party vendor

III. Overview of Traditional Defenses

Security cannot be achieved by one solution or one party. Multiple layers of defenses need to be placed throughout an organization to ensure end-to-end data security [8] and customers, employees and IT personnel have their share of responsibility. The following are some typical defense mechanisms placed in layers by organizations to protect resources.

A. Perimeter Defense Mechanisms

Perimeter defense mechanisms refer to the methods of isolating internal networks or applications from the outside world. There are primarily two classes of perimeter defense mechanisms:

1) Firewall: It is a networking device that separate internal networks from external networks. Firewalls are the very first level of defense that protects an organization from unauthorized accesses and intrusions from external attacks. However, they may not prevent malicious activities through legitimate channels, for example, through externally exposed web applications, emails, etc. [9].

2) Virtual Private Network (VPN): It is an essential component for an enterprise and provide a secure channel between two private networks of a geographically distributed organization through Internet. They also enable legitimate users to connect to corporate network remotely.

B. Deceptive Defense Mechanisms

Deceptive defense mechanisms refer to the methods of collecting intelligence about malicious activities and deflect attackers by putting a set of legitimate looking systems in an isolated and monitored environment. *Honeypot Framework* is a deceptive defense method designed to serves two purposes: early warnings and forensic analysis. These systems are placed to mislead cyber attackers, and then to detect and study the attempts to gain unauthorized access to information systems [10], [11]. Honeypot frameworks are relatively new technology and being adopted by large cloud service providers. However, it can be complex and expensive to implement and maintain a robust honeypot framework and have good coverage.

C. Detection Defense Mechanisms

Detection defense mechanisms refer to the methods of monitoring network and systems for malicious and anomalous activities. They can primarily be divided into two categories:

1) Intrusion detection system (IDS): It monitors both external facing and internal systems to detect malicious activities [12]. It scans live traffics, contents on different systems, and system logs, and perform various data mining and forensic techniques on them to detect attacks and malware in near real time from both external and internal entities. With proper instrumentation and careful planning IDS may help in detecting data leaks.

2) Endpoint protection: Is a component in IDS that protects servers, client computers, mobile devices, and Internet of Things (IoT) devices in a network by monitoring and finger printing network, browser, and file system activities for malware, phishing, and other forms of attacks [13]. This is one of the fastest growing market for security products due to its effectiveness and scope [14].

D. Application Defense Mechanisms

Application defense mechanisms refer to the security measurements built in applications to prevent vulnerabilities and only allow users to use the application in authorized manner. There are many mechanisms to secure applications and these vary based on the nature of the application. Examples of traditional defense mechanisms is provided in Table II. A few principles and methodologies that are commonly applied and are considered crucial for any applications are:

1) Implementation of security fundamentals: Implementing cybersecurity first principles in the software is the very first step of building defenses in applications. Domain separation, process isolation, resource encapsulation, least privilege, modularity, layering, abstraction, data hiding, simplicity and minimization [15] are the fundamentals blocks of a secure software and can help in building defenses against data theft.

2) Defenses against common exploitation techniques: Writing code to deal with common application exploit techniques, for example, SQL injection, Cross Site Scripting (XSS), and Cross Site Request Forgery (CSRF) and testing applications against these attacks are crucial to prevent applications being compromised.

3) Access control: It ensures that only the authorized users are allowed to access applications. Having an effective access control mechanism is essential for protecting any applications, data and prerequisite for any defense mechanisms. Role Based Access Control (RBAC) [16] and Attribute Based Access Control (ABAC) [17] can help in defining access control policies based on user's roles and responsibilities and thus can help in simplifying privilege management especially for large enterprises, resulting in better defense against data breaches.

4) Adaptive authentication: Authentication enables an application to identify its users. Adaptive authentication, also known as Risk Based Authentication, is a method for stronger authentication performed by profiling users' behavior and software and hardware usage profiling [18], and whenever a deviation is observed, step up authentication by challenging additional authentication scheme, for example, Two-Factor Authentication, and Multi-Factor Authentication (MFA).

E. Cryptographic Defense Mechanisms

Cryptographic defenses refer to the methods of encoding data using cryptographic keys and certificates such a way that only authorized users can access that without the loss of intelligibility of the content.

1) Security in transit: Using secure channel such as Transport Layer Security (TLS) encryption for transactions between client and server, between applications, between on-premise and cloud platforms prevents hackers from viewing and modifying any information.

2) Security at rest: It refers to the processes to protect data from viewing in the event of unauthorized access or theft. Encryption of data can provide such security. Encryption can be applied in different level:

a) Disk or file-system level encryption: It is done at hardware or operating system layers. This type of encryption is typically transparent from users and primarily offers protection from physical theft of storage.

b) File level encryption: It is done at file or folder level. Pretty Good Privacy (PGP) is a popular method to encrypt and protect data in files [19]. This is also helpful when handing data with external vendors who are on different networks and therefore cannot take the benefit of inhouse security infrastructure.

c) Application or database level encryption: Often applications or databases offer encryption that protect data while in storage. When the data are retrieved for use by the user, the decryption takes place. The decryption may be transparent, i.e. may not require any interaction from the users.

TABLE II: EXAMPLES OF TRADITIONAL DEFENSE MECHANISMS

Defense Layer	***Defense Mechanisms***
Perimeter Defenses	Firewalls
	Virtual Private Networks
Deceptive Defenses	Honeypots
Detection Defenses	Intrusion Detection Systems
	Endpoint Protection
Application Defenses	Software Security
	Access Control
	Adaptive Authentication
Cryptographic Defenses	Security in Transit
	Security at Rest

IV. HUMAN FACTOR IN CYBER DEFENSE

Employees are widely acknowledged to be responsible for security breaches in organizations and hence are considered one of the biggest potential threats to their cyber security [20]. Gartner predicts that, through 2020, 95 percent of cloud security failures will be the customer's fault [21]. Regardless of how strong the defenses are, they can fail if employees do not follow common security practices. Improving security awareness and building secure infrastructure to reduce human errors are crucial to fight against data breaches. Here are some areas where trainings can help in reducing human errors:

A. Security Training for End-Users

1) Phishing attack: It is an attack targeted to steal users' credentials or login sessions in the disguise of a trustworthy entity and use that to steal sensitive data for monetary gain. Phishing attack was identified as one of the most common reasons and entry point for other attacks responsible for data breaches. Phishing uses social engineering and hence is a difficult problem to solve; however, with proper awareness training it is possible to help users avoid phishing attacks [4].

2) Weak password and password reuse: Due to the major data breaches over past few years, credential leaks have become a broader risk to the online ecosystem and enterprises due to weak password selection and re-use. A case study performed by Google found 7% to 25% of exposed passwords match a victim's Google account [22]. While enforcing strong password policy can prevent users from selecting weak passwords, reuse of old or same passwords in different places can be prevented by raising security awareness.

B. Security Training for IT Personnel

1) Password and key management: Due to the well adoption of Continuous Integration and Continuous Deployment (CICD) model, it is very common for developers to check in credentials and keys required for operation inside the code. Implementing a design to store the credentials and the keys separately from the code is important [23]. Sometimes released code contains comments with information about production systems or the logic how the codebase works, which make hacking easier [24]. Training developers and IT personnel and building proper controls can help in preventing these kinds of errors.

2) Security patching: Patching software on a timely manner can reduce risk and prevent hackers exploiting known vulnerabilities. According to former Equifax CEO Richard Smith’s congressional testimony, the company was instructed to apply a patch for known vulnerability in the Apache Struts server software within 48 hours, however, the company did not apply the patch until its online dispute portal was compromised three months later, resulting in one of the largest data breaches that affected 146.6M consumers revealing highly sensitive information, for example social security number, driving license number, passport number, and credit card numbers. [25].

V. Data Centric Security

A data breach typically indicates other defense mechanisms have failed and attackers have been successful in stealing data as illustrated in Fig. 1. Data centric security can help in this situation by preventing the data from being exposed to the attacker. Data centric security refers to the process of protecting data through redaction isolation, tokenization, and encryption. The main purpose here is to minimize the aftermath of data breaches. Data centric security is applied in multiple ways:

A. Data Minimization and Redaction

Modern applications collect a lot of data, for improving security or customer experience and for business analytics. However, if some data are not needed beyond the immediate use, or only the partial data are needed, it is best not to store the data or only store the partial data [26], for example, storing only the last four digits of Social Security Numbers (SSN) instead of the whole SSN for verification or not storing the birth date when only the year of birth is needed to identify senior citizens can reduce the risk.

B. Data Classification and Isolation of Sensitive Data

Classifying data can help in identifying customers’ and company’s sensitive data. Storing sensitive data in an isolated environment, separated from non-sensitive data, with an elevated security can minimize the risk of data exposure [27].

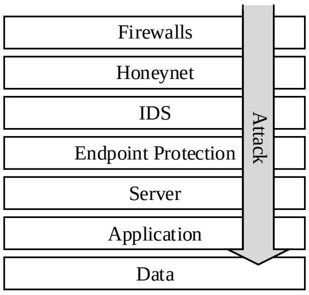

Fig. 1. Data breach - failure in defenses.

C. Data Level Encryption

Data level encryption as illustrated in Fig. 2, is relatively new and an increasingly popular mechanism for ensuring security at rest and beyond. In this method data is secured by encrypting the data itself rather than relying on the security of networks, servers or applications. Encrypted data cannot be used by the hackers even when they are stolen.

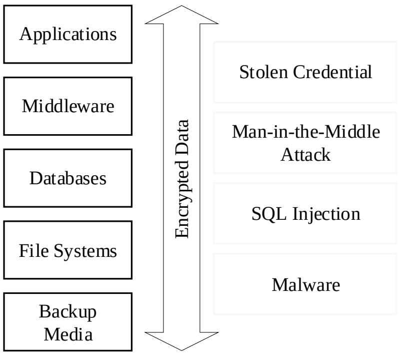

Fig. 2. Data level encryption can offer better protection against data breaches through different forms of attacks.

VI. Challenges in Implementing Data Centric Security

Here are some challenges that need to be dealt with when implementing data level encryption:

A. Performance Overhead

The first thing that comes to mind when considering encryption is performance; however, with the high performance of modern hardware, the performance overhead of encryption itself is negligible. However, the performance overhead can come from factors like how the sensitive data are laid out and what operations are being performed on the data. For example, encrypting sensitive data in a large JSON message in a real time streaming service requires parsing the messages and encrypting individual sensitive attributes, which can add overhead. Also, searching, sorting, and joins on large amount of data can see performance degrade since cyphertexts break collation rules. Collations refers to a set of rules for comparing and sorting data. Collation assigns certain characteristics to data that affects many operations in the database. Collations use case sensitivity, accent sensitivity, character sequence, type of characters and length of characters to optimize performance of sorting and comparing data. Since cyphertexts are random, they break the optimization done by databases, resulting performance degrade in sort, search, and join operations. As a result, proper data modeling with data level encryption in mind to reduce the overhead is important.

B. Encryption Capability in Platforms

Not all platforms and tools provide the ability to perform data level encryption. There are platforms that do not support encryption at all, for example, Teradata, which is a leading data warehousing platform [28]. Some platforms only support file level or database level encryption, which are typically transparent, which means the data are encrypted when storing to the file and decrypted when retrieved by the users, for example, Hadoop [29]. Even when column level encryption is supported by platforms, the key management processes vary from platform to platform. All of these make implementation of data level encryption very complex, especially in large data environments where hundreds of systems are to be dealt with.

C. Platform Agnostic Implementation

As there is no standard followed by the platform providers for data level encryption, a platform agnostic implementation seems to be a desirable solution as described in Fig. 3. In this approach, data are encrypted during ingestion and decrypted during egression. The keys are stored in a central Key Management Server and often additionally protected by Hardware Security Module, access to the keys are controlled via security groups or some other sort of authorization process, and user or service accounts get decryption privilege by going through the authorization process. Not only this simplifies the operational challenges by avoiding encryption and decryption of data in each platform with its native capability as data move from one platform to another, but also makes key management simpler due to the centralization approach.

While platform agnostic encryption appears to be a feasible solution, the challenge is how to integrate this mechanism with different platforms. Most data platforms provide User Defied Functions (UDFs) that enable writing custom code for handling data and invoking that custom code in the query or APIs. This functionality can be leveraged for encryption and decryption of data out of band. However, there are challenges in the way the UDFs are handled in different data platforms. Some platforms have limitations on how much data can be passed to an UDF, limiting the possibility of handling large data blocks.

D. Format Preserving Encryption

Cyphertext can contain control characters, which can break data handling processes in platforms and during transit. Base64 encoding can help with this problem, but that will increase the length of the data significantly. Also, the data type may need to be changed which may not be acceptable in many situations. A solution is to use Format Preserving Encryption (FPE), which can preserve the length and the format of the data, for example, integer can remain integer, alphabets can remain alphabet, and so on [30]. However, the performance of FPE can degrade after a certain size of data and hence FPE is not a solution for large unstructured data, for example, a message containing sensitive data may not be encrypted by FPE without affecting performance significantly.

E. Cloud SaaS Services

Cloud based Software as a Services (SaaS) have become a common strategic move by enterprises to reduce cost, achieve scalability and mitigate risks. However, that adds a new dimension of challenges for data level encryption. The SaaS services are totally managed by the cloud vendors. Clients do not have access to the physical hosts to install any custom encryption libraries. The environments are locked down by firewalls, preventing making outside calls to get keys and authorization. Therefore, a platform agnostic implementation is not a possibility for enterprises using Cloud data platforms like Amazon Web Service, Azure, etc.

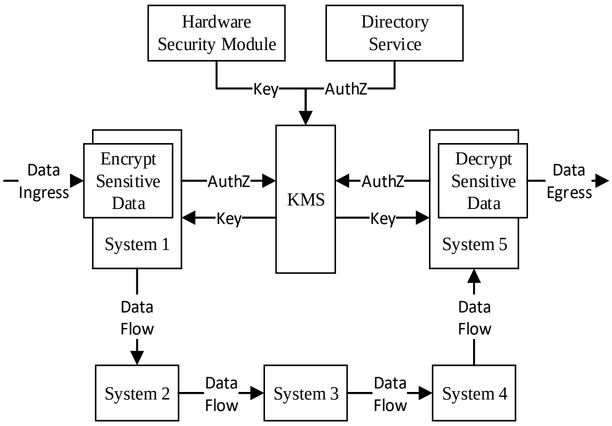

Fig. 3. A conceptual data level encryption framework.

VII. Concluding Remarks

This paper has identified the characteristics of the exploits from recent major data breaches, explored the available mitigations, discussed human factor in effectiveness of cyber defense, and finally examined data centric security strategy and the challenges in implementing that in large enterprises. The study finds a platform agnostic data level encryption strategy using FPE can offer good protection against data breaches. This approach can help in building performant scalable framework by avoiding repetitive encryption or decryption of data during movement across platforms. It can also simplify the key management and access control to sensitive data. However, the technology is yet to mature with supports from the software vendors and global standardization bodies. Data centric security can be an effective solution for protecting data against data breaches. However, as discussed in the previous sections, a successful execution has many dependencies: (a) data platform vendors need to understand the importance of data level security and provide the framework to implement that, (b) global standardization can solve many challenges by coming up with a standard way of performing data level encryption, and (c) finally cloud service providers need to add support in the managed platforms to integrate data level encryption.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Dr. Kettani supervised the work of Mr. Nur in partial fulfillment of cyber security research graduate course at Dakota State University.

References

[1] European Union Agency for Network and Information Security (ENISA). (2018). *ENISA Threat Landscape Report 2017: 15 Top Cyber-Threats and Trends*. Heraklion: ENISA. [Online]. Available: https://doi.org/10.2824/967192

[2] T. Floyd, M. Grieco, and E. F. Reid, "Mining hospital data breach records: Cyber threats to U.S. hospitals," in *Proc. IEEE Conference on Intelligence and Security Informatics (ISI 2016)*, pp. 43–48, 2016.

[3] OWASP. (2018). OWASP top 10 – 2017: The ten most critical web application security risks. *OWASP Foundation.* [Online]. Available: https://www.owasp.org/index.php/Top_10-2017_Top_10

[4] J. A. Shamsi and M. A. Khojaye, "Understanding privacy violations in big data systems," *IT Professional*, vol. 20, no. 3, pp. 73–81, 2018.

[5] T. Armerding. (January 26, 2018). The 17 biggest data breaches of the 21st century. *CSO Online.* [Online]. Available: https://www.csoonline.com/article/2130877/data-breach/the-biggest-data-breaches-of-the-21st-century.html

[6] J. Leyden. (December 23, 2014). JPMorgan chase mega-hack was a simple two-factor auth fail. [Online]. Available: https://theregister.co.uk/2014/12/23/jpmorgan_breach_probe_latest/

[7] L. Sporck. (November 22, 2017). 11 of the largest data breaches of all time. *OPSWAT.* [Online]. Available: https://www.opswat.com/blog/11-largest-data-breaches-all-time-updated

[8] T. Mavroeidakos, A. Michalas, and D. D. Vergados, "Security architecture based on defense in depth for cloud computing environment," in *Proc. IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, San Francisco, CA, 2016, pp. 334–339.

[9] S. Alsehibani and S. Almuhammadi, "Anomaly detection: Firewalls capabilities and limitations," in *Proc. 2018 International Conference on Computing Sciences and Engineering (ICCSE)*, Kuwait City, Kuwait, 2018, pp. 1-5.

[10] S. Kumar, R. Sehgal, and J. S. Bhatia, "Hybrid honeypot framework for malware collection and analysis," in *Proc. the 7th IEEE International Conference on Industrial and Information Systems (ICIIS)*, Chennai, India, 2012, pp. 1-5.

[11] B. Nagpal, N. Singh, N. Chauhan, and P. Sharma, "CATCH: Comparison and analysis of tools covering honeypots," in *Proc. the International Conference on Advances Computer Engineering and Applications (ICACEA 2015)*, Ghaziabad, India, 2015, pp. 783-786.

[12] A. Borkar, A. Donode, and A. Kumari, "A survey on intrusion detection system (IDS) and internal intrusion detection and protection system (IIDPS)," in *Proc. International Conference on Inventive Computing and Informatics (ICICI)*, Coimbatore, India, 2017, pp. 949-953.

[13] C. Onwubiko, "Exploring web analytics to enhance cyber situational awareness for the protection of online Web services," in *Proc. International Conference on Cyber Security and Protection of Digital Services (Cyber Security 2016)*, London, UK, 2016, pp. 1-8.

[14] J. Oltsik. (June 20, 2018). The new endpoint security market: Growing in size and scope. *CSO Online*. [Online]. Available: https://www.csoonline.com/article/3281023/security/the-new-endpoint-security-market-growing-in-size-and-scope.html

[15] S. Mishra, R. K. Raj, P. Tymann, J. Fagan, and S. Miller, "CyberCSP: Integrating cybersecurity into the computer science principles course," in *Proc. 2017 IEEE Frontiers in Education Conference (FIE)*, Indianapolis, IN, 2017, pp. 1-5.

[16] A. Castiglione, A. Santis, B. Masucci, F. Palmieri, A. Castiglione, J. Li, and X. Huang, "Hierarchical and shared access control," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 4, pp. 850-865. 2016.

[17] V. C. Hu, D. R. Kuhn, D. F. Ferraiolo, and J. Voas, "Attribute-based access control," *Computer*, vol. 48, no. 2, pp. 85-88, 2015.

[18] K. A. A. Bakar and G. R. Haron, "Adaptive authentication based on analysis of user behavior," in *Proc. 2014 Science and Information Conference (SAI)*, London, UK, 2014, 601-606.

[19] A. Anugurala and A. Chopra, "Securing and preventing man in middle attack in grid using open pretty good privacy (PGP)," in *Proc. 2016 Fourth International Conference on Parallel, Distributed and Grid Computing (PDGC)*, Waknaghat, India, 2016, pp. 517-521.

[20] M. Alotaibi, S. Furnell, and N. Clarke, "Information security policies: A review of challenges and influencing factors," in *Proc. the 11th International Conference for Internet Technology and Secured Transactions (ICITST)*, Barcelona, Spain, 2016, pp. 352-358.

[21] C. Pettey, (August 19, 2016). Why cloud security is everyone's business. *Smarter with Gartner.* [Online]. Available: https://www.gartner.com/smarterwithgartner/why-cloud-security-is-everyones-business

[22] K. Thomas, F. Li, A. Zand, J. Barrett, J. Ranieri, L. Invernizzi, Y. Markov, O. Comanescu, V. Eranti, A. Moscicki, D. Margolis, V. Paxson, and E. Bursztein, "Data breaches, phishing, or malware? Understanding the risks of stolen credentials," in *Proc. 2017 ACM SIGSAC Conference on Computer and Communications Security (CCS'17)*, Dallas, TX, 2017, pp. 1421-1434.

[23] P. M. Duvall, S. Matyas, and A. Glover, *Continuous Integration: Improving Software Quality and Reducing Risk*, Boston, MA: Addison-Wesley, 2007.

[24] A. White, *Hacking: The Underground Guide to Computer Hacking, Including Wireless Networks, Security, Windows, Kali Linux and Penetration Testing*, Scotts Valley, CA: CreateSpace Independent Publishing Platform, 2018.

[25] H. Berghel, "Equifax and the latest round of identity theft roulette," *Computer*, no. 12, pp. 72-76, 2017.

[26] E. Bier, R. Chow, P. Golle, T. H. King, and J. Staddon, "The rules of redaction: Identify, protect, review (and repeat)," *IEEE Security & Privacy*, vol. 7, no. 6, 2009.

[27] M. Schmidt, S. Fahl, R. Schwarzkopf, and B. Freisleben, "TrustBox: A security architecture for preventing data breaches," in *Proc. the 19th International Euromicro Conference on Parallel, Distributed and Network-Based Processing*, Ayia Napa, Cyprus, 2011, pp. 635–639.

[28] J. Browning. (2007). Security Features in Teradata Database (Report No. EB-1895-1007). San Diego, CA: Teradata Corporation. [Online]. Available: https://www.teradataemea.com/campaign/pdf/EB1895.pdf

[29] Apache Software Foundation. (April 16, 2018). Transparent encryption in HDFS (Version 2.9.1). [Online]. Available: https://hadoop.apache.org/docs/current/hadoop-project-dist/hadoop-hdfs/TransparentEncryption.html

[30] Z. Bhatt and V. Gupta, "Advance security technique for format preserving encryption," in *Proc. the International Conference on Inventive Computation Technologies (ICICT)*, Coimbatore, India, 2016, pp. 1-4.

Mohammad M. Nur was born in Dhaka, Bangladesh, in 1979. He earned the B.S. and M.S. degrees in 2002 and 2005, respectively, both in computer science from Minnesota State University, Mankato. He is currently pursuing a Ph.D. degree in cyber operations from Dakota State University. He currently works in the Enterprise Data Services Security Team at T-Mobile USA, Inc., Bellevue, WA as a principal cybersecurity engineer. He previously worked at HP Enterprise Services, IBM Global Services, and Microsoft Corporation, mostly developing software and services in identity, credential and access management domain. His research interests include identity and access management, big data security, and software exploitation.

Houssain Kettani was born in Khobar, Saudi Arabia, in 1978. He received the B.S. degree in electrical and electronic engineering from Eastern Mediterranean University, Cyprus in 1998, and the M.S., Ph.D. degrees both in electrical engineering from the University of Wisconsin at Madison in 2000 and 2002, respectively. Dr. Kettani served as faculty member at the University of South Alabama (2002-2003), Jackson State University (2003-2007), Polytechnic University of Puerto Rico (2007-2012), Fort Hays State

University (2012-2016), Florida Polytechnic University (2016-2018) and Dakota State University since 2018. Dr. Kettani has served as staff research assistant at Los Alamos National Laboratory in summer of 2000, visiting research professor at Oak Ridge National Laboratory in summers of 2005 to 2011, visiting research professor at the Arctic Region Supercomputing Center at the University of Alaska in summer of 2008 and visiting professor at the Joint Institute for Computational Sciences at the University of Tennessee at Knoxville in summer of 2010. Dr. Kettani's research interests include computational science and engineering, high performance computing algorithms, information retrieval, network traffic characterization, number theory, robust control and optimization, and Muslim population studies. He presented his research in over one hundred refereed conference and journal publications and his work received over six hundred citations by researchers all over the world. He chaired over hundred international conferences throughout the world and successfully secured external funding in millions of dollars for research and education from US federal agencies such as NSF, DOE, DOD, and NRC.