Inf Syst Front (2015) 17:983–986
DOI 10.1007/s10796-015-9593-y

Advances in secure knowledge management in the big data era

Chittaranjan Hota[1] · Shambhu Upadhyaya[2] · Jamal Nazzal Al-Karaki[3]

Published online: 9 September 2015

Information is increasingly becoming important in our daily lives. We have become information dependents of the twenty-first century, living in an on-command, on-demand Big data world. In this era, more information is being created by individuals than by business houses. In the past, we had to stand in a queue at a railway reservation counter to book our tickets, had to visit a cash counter in a bank to do our transactions, had to arrange a get together at a physical location within our town to meet and socialize with our friends, had to visit a theater to watch a movie, and so on. We now have Information and Communication Technology (ICT) to help us do all these by sitting in front of a computer and with a few mouse clicks. Also, all these advancements are possible because we are drenched in a flood of data today. It is important to distinguish between data, information and knowledge. Data is a set of facts about events. Information is a processed set of facts that are meaningful. Knowledge is broader, deeper and richer than both data and information. Knowledge answers questions like "How".

As the world is getting technology savvy, the collection and distribution of information and knowledge need special attention (Upadhyaya et al. 2006). As the volume of data being gathered and stored around the globe is exploding (due to globalization efforts) and the cost of technologies like Analytics, Machine learning, Statistics, and Networking is falling, researchers, industry professionals and policy makers are beginning to realize the potential of Big data to identify the needs of common man and provide services that can make their life better. Days are not far when we will experience applications like, understanding the strengths and weaknesses of our students by getting access to their transcripts available on cloud, and correlating those grades with their online experiences; building an intelligent transportation system for a city through analysis and visualization of live and detailed road network data, etc.

Beyond the transactional data used by many organizations, there exists a potential treasure trove of non-traditional, less structured data (Big data) that can be mined for useful information (Rajpathak and Narsingpurkar 2013). Today, most of the young people have twitter, facebook, linked-in, google+ accounts for their online activities. Also, people are acquainted with flickr to upload their photographs, semantria.com to see sentiment analysis or opinion mining, ebay.com to buy or sell items, and crowd sourcing tasks on Amazon.com. All these are applications of Big data. Digital information available on the Internet is increasing 10 folds every five years in a scale of Zeta-bytes. Data is now available from blogs, RFIDs, sensors, cameras, social networks, telephony, e-commerce and medical records.

✉ Chittaranjan Hota
hota@hyderabad.bits-pilani.ac.in

Shambhu Upadhyaya
shambhu@buffalo.edu

Jamal Nazzal Al-Karaki
jamal.alkaraki@adpoly.ac.ae

[1] Department of Computer Science, BITS-Pilani, Hyderabad campus, Hyderabad, Telangana, India

[2] Department of Computer Science and Engineering, University at Buffalo, Buffalo, NY, USA

[3] Information Security Engineering Technology, Abu Dhabi Polytechnic, Abu Dhabi, United Arab Emirates

To gain value from Big data, researchers and industry professionals are continuously discovering alternate ways to capture, search, integrate, and process this data. Such Big data analytics now drives almost every aspect of our modern society, including telecom, retail, manufacturing, and life sciences creating new challenges for Knowledge Management Systems (KMS). Knowledge management increases the value of an organization by identifying the assets and expertise available within the organization as well as efficiently managing the resources (Bertino et al. 2006). Knowledge management is about corporations sharing their resources and expertise, as well as building intellectual capital so that they can increase their competitiveness (Thuraisingham and Parikh 2008). Knowledge Management identifies, captures, and processes organizational knowledge with an aim to control knowledge resources within an organization (Wolf 2012).

Knowledge Management Systems often involve a number of categories such as Data Warehousing and Mining, Decision Support Systems, Content Management Systems, etc. As a result, secure knowledge management spans across multiple aspects of any organization. With the advent of the Big data era, these issues become many-fold. Intellectual capital of any organization is closely tied-up with the KMS that is in place within an organization, which helps an organization leverage competitive advantage and enhance productivity. Hence it has become increasingly important for an organization, big or small, to identify, manage and protect its intellectual capital. With the avalanche of data and information at their disposal, number of companies store and analyze petabytes of internal as well as business data including, but not limited to, website visits and logs, emails, social media interaction with the world, employee data and records, advertising, and sensor data etc. to gain better insights about their sales, revenues, collaboration, and customer satisfaction. Such Big data further brings up the issues of information ownership and information classification. Even in the context without Big data, it is not trivial for most organizations to implement these concepts. Decreases in the cost of both storage and compute power have made it feasible to collect such Big data. As a result, more and more companies are looking to include non-traditional yet potentially very valuable data (Big data) along with their traditional enterprise data in their business intelligence analysis (Rajpathak and Narsingpurkar 2013). Big data brings in more challenges with regard to identifying the owners for the outputs of Big data processes. Thus, ownership of data, information, and the subsequent knowledge generated from them may lie in different hands. Clearly identifying these aspects is important for the success of any KMS utilizing Big data.

In a connected world, the motivations to penetrate systems or to inject malicious software vary from personal satisfaction, to espionage, to financial rewards, to revenge, and many more reasons. Increasing needs of the organization for the knowledge management also increases the need for organizations to protect their knowledge assets from leakage, destruction, unavailability, and other threats (Lee et al. 2005). Security for knowledge management is essential since organizations need to protect their intellectual assets. Some information in an organization may appropriately be shared amongst multiple stakeholders and some may be protected from unauthorized modifications. New attack patterns, increasing number of tools to attack systems, rapid growth of computer networking, a wide variety of applications dealing with confidential information, etc. are forcing the security researchers to constantly discover ways to solve these problems specially when there is a deluge of data.

Integrating Big data with security in knowledge management provides unique opportunities to consolidate and analyze logs and events from multiple sources rather than evaluate them in isolation. Integrating information from various sources can generate valuable knowledge about enhancing security which was not available to traditional business environments. By integrating information from IP-enabled smart CCTVs, data from biometric systems, logs of Intrusion Detection Systems or other sources etc., organizations can tap into the Big data and significantly enhance the security in Knowledge Management Systems. By the virtue of availability of such knowledge, we can have advanced detection of insider threats, criminal activities, and frauds. However, large volume of information also creates challenges in terms of identifying the actual security threats for an organization, and not get lost in a wave of false positives.

The purpose of the Special Issue is to report state-of-the-art research in different aspects of security in Knowledge Management Systems (KMS) in the Big data era. This special issue of Information Systems Frontiers journal provides a premier forum for researchers and practitioners of KMS to present their work to other peers in the security domain. This issue presents expanded versions of five papers presented at 6th International Conference on Secure Knowledge Management in Big Data Era (SKM 2014) held at Dubai, UAE. Having started in 2004 at SUNY-Buffalo, SKM is being organized all over the world once every two years. Authors of the presented papers at SKM 2014 were invited to submit their extended work to this special issue. All the papers (SKM '14, and others) submitted to this special issue went again through a rigorous review by experts and five papers were accepted for publication in this journal. Each paper presents different security aspects for building knowledge management systems.

In the first paper by Mehresh and Upadhyaya (2015), the authors propose a novel "mission survivability" architecture against Advanced Persistent Threats (APT) in a distributed

environment. If suspicious activity is detected but any imminent danger is not perceived, their approach refrains from sending clear signals to the attackers such as raising alerts or terminating a session. Instead, the system continues to behave normally and make observations to better understand the attacker's intent, objectives and strategies. Their framework disguises the knowledge of detection from the attackers, and thus helps in the designing of targeted recovery procedures. The 'false assurance' to an attacker prevents the attacker from switching to an aggressive strategy or an alternate plan.

In the next paper, Lee et al. (2015) examine message diffusion on Twitter social networking platform during extreme events. Their work examines the impact of tweet features on the diffusion of two types of messages during 2013 Boston marathon tragedy – rumor related and non-rumor related (both in the context of the Boston tragedy). The work of the authors demonstrates the adaptation of the innovation diffusion model to the diffusion of information, specifically to the diffusion of tweet messages. The authors identify several features of tweets as important to diffusion – *(a)* reaction time of tweet, *(b)* Type of message, *(c)* number of followers, and *(d)* hashtag usage of tweet.

In the third paper of this special issue, Sun and Upadhyaya (2015) develop a rule based data sanitization scheme to detect and remove personally identifiable and other sensitive information from the data sets for user authentication using keystroke dynamics and mouse movements. The authors argue that the privacy of collection and transmission of keyboard and mouse data has not received much attention. The authors develop two new architectures for providing privacy preserving data processing support for active authentication and its adoption as a secure authentication scheme in the real world. The two contributions are – one, a lightweight rule based scripting tool for removing personally identifiable information and other sensitive information from keystroke dynamics data in behavioral biometrics based authentication systems; and two, an Extensible Messaging and Presence Protocol (XMPP) based scheme for safely transmitting keystroke dynamics data from the client to the server. Both schemes were evaluated with real-world data.

Narang and Hota (2015), in the penultimate paper, present a game-theoretic approach for deployment of Intrusion Detection Systems (IDS) in a Peer-to-Peer (P2P) network. The authors model a P2P network in the form of a graph, and consider an arbitrary attacker who could connect to any peer in the network and try to launch attacks on peers which hold higher 'value' in the network (labeled as 'target' nodes). The game-theoretic evaluation by the authors provides randomized strategies for deployment of IDS at different peers within the network. The authors present two kinds of solutions – a simpler solution, wherein the responsibility of running the IDS lies on the target nodes themselves, and an advanced solution, wherein the responsibility of running the IDS is given to nodes other than the target nodes.

Finally, in the last paper of this special issue, Nath and Mehtre (2015) present analysis of a multi-stage attack that uses video files as a covert channel. The authors argue that a naïve user, using his system to view an interesting video, could innocuously open a video file which contains a multi-stage attack. The damage will be higher if the machine is part of a mission critical system. The authors analyze that some video files contain malicious link through which an exploit gets downloaded into the host machine. This sets up the stage for a targeted attack launched in multiple stages. Finally, the authors propose a new method for detection of such attacks using API calls.

Acknowledgments We would like to thank Prof. Nasir Memon (Professor of Computer Science and Engineering at New York University) for his guidance as the Advisory Editor of this special issue.

References

Bertino, E., Khan, L.R., Sandhu, R., & Thuraisingham, B. (2006). Secure knowledge management: confidentiality, trust, and privacy. *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, *36*(3), 429–438.

Lee, J., Upadhyaya, S.J., Rao, H.R., & Sharman, R. (2005). Secure knowledge management and the semantic web. *Communications of the ACM*, *48*(12), 48–54.

Lee, J., Agrawal, M., & Rao, H. (2015). Message diffusion through social network service: The case of rumor and non-rumor related tweets during boston bombing 2013. *Information Systems Frontiers*, *17*(5). doi:10.1007/s10796-015-9568-z.

Mehresh, R., & Upadhyaya, S. (2015). Surviving advanced persistent threats in a distributed environment–architecture and analysis. *Information Systems Frontiers*, *17*(5). doi:10.1007/s10796-015-9569-y.

Narang, P., & Hota, C. (2015). Game-theoretic strategies for ids deployment in peer-to-peer networks. *Information Systems Frontiers*, *17*(5). doi:10.1007/s10796-015-9582-1.

Nath, H.V., & Mehtre, B. (2015). Analysis of a multistage attack embedded in a video file. *Information Systems Frontiers*, *17*(5), 1–9. doi:10.1007/s10796-015-9570-5.

Rajpathak, T., & Narsingpurkar, A. (2013). Knowledge from big data analytics in product development. TCS white paper.

Sun, Y., & Upadhyaya, S. (2015). Secure and privacy-preserving data processing support for active authentication. *Information Systems Frontiers*, *17*(5). doi:10.1007/s10796-015-9587-9.

Thuraisingham, B., & Parikh, P. (2008). Trustworthy semantic web technologies for secure knowledge management. In *2008 IEEE/IFIP International Conference on Embedded and Ubiquitous Computing, IEEE, pp 186–193*.

Upadhyaya, S., Rao, H.R., & Padmanabhan, G. (2006). Secure knowledge management. Idea group inc.

Wolf, S. (2012). Security management for knowledge assets.

Chittaranjan Hota is a Professor and Associate Dean at Birla Institute of Technology and Science – Pilani, Hyderabad Campus, Hyderabad, India. He was the founding Head of Dept. of Computer Science at BITS, Hyderabad. Prof. Hota did his PhD in Computer Science and Engineering from Birla Institute of Technology & Science, Pilani. He has been a visiting researcher and visiting professor at University of New South Wales, Sydney; University of Cagliari, Italy; Aalto University, Finland and City University, London over the past few years. His research work has been funded by University Grants Commission (UGC), New Delhi; Department of Electronics & Information Technology (DeitY), New Delhi; and Tata Consultancy Services, India. He has guided PhD students and currently guiding several in the areas of Overlay networks, Information Security, and Distributed computing. He is recipient of Australian Vice Chancellors' Committee award, recipient of Erasmus Mundus fellowship from European commission, and recipient of Certificate of Excellence from Kris Ramachandran Faculty Excellence Award from BITS Pilani. He has published extensively in peer-reviewed journals and conferences and has also edited LNCS volumes. He is a member of IEEE, ACM, IE, and ISTE.

Shambhu J. Upadhyaya is a Professor of Computer Science and Engineering at the State University of New York at Buffalo where he also directs the Center of Excellence in Information Systems Assurance Research and Education (CEISARE), designated by the National Security Agency. His research interests are information assurance, computer security, fault diagnosis, fault tolerant computing, and VLSI Testing. He has authored or coauthored more than 250 articles in refereed journals and conferences in these areas. His research has been supported by the National Science Foundation, Rome Laboratory, the U.S. Air Force Office of Scientific Research, DARPA, National Security Agency, IBM, Intel Corporation and Harris Corporation. He has been awarded IBM Faculty Partnership Fellowship, NRC Faculty Fellowships in past. He has held visiting research faculty positions at the Center for Reliable and High Performance Computing, University of Illinois, Urbana-Champaign, Intel Corporation, Folsom, CA, Air Force Research Laboratory, Rome, NY and the Naval Research Laboratory, Washington DC. He was an associate editor of IEEE Transactions on Computers from 2001 to 2006, is a member of the editorial board of the International Journal on Reliability, Quality, and Safety Engineering published by the World Scientific Publishers and the ICST Transactions on Security and Safety. He was a guest co-editor of the book series Interfaces in OR/CS on Mobile Computing: Implementing Pervasive Information and Communication Technologies, Kluwer Academic Publishers, 2001, was a guest co-editor of a special issue on Secure Knowledge Management in IEEE Transactions on Systems, Man and Cybernetics, May 2006 and is a guest co-editor of a special issue on Emerging Security Trends for Deeply-embedded Computing Systems in IEEE Transactions on Emerging Topics in Computing, December 2015.

Jamal Al-Karaki is the Head of Information Security Engineering Technology at Abu Dhabi Polytechnic, United Arab Emirates. Jamal earned his PhD in Computer engineering from Iowa State University, USA. He has held various positions at The Hashemite University, Zarka, Jordan including the Dean, Faculty of Prince Hussein Bin Abdullah-II for Information Technology, Chair and Co-founder of Computer Engineering Department; and Director, Computer center. He has published extensively in reputed International journals and conferences in the areas of Wireless networking, Mobile computing, Network security, and Performance evaluation. He is recipient of research excellence award from Iowa State University, Ames, IA, USA; Her Majesty Queen Rania Award for entrepreneurs, Jordan; and Best Teacher Award, Delmon University, Bahrain. He is a member of Jordan Engineers Association, member of Tau Beta Pi (The Engineering Honor Society), USA, and member of IEEE.