Data Security and Privacy

Concepts, Approaches, and Research Directions

Elisa Bertino
Computer Science Department and CERIAS
Purdue University
West Lafayette, Indiana (USA)
bertino@purdue.edu

Abstract—**Data are today an asset more critical than ever for all organizations we may think of. Recent advances and trends, such as sensor systems, IoT, cloud computing, and data analytics, are making possible to pervasively, efficiently, and effectively collect data. However for data to be used to their full power, data security and privacy are critical. Even though data security and privacy have been widely investigated over the past thirty years, today we face new difficult data security and privacy challenges. Some of those challenges arise from increasing privacy concerns with respect to the use of data and from the need of reconciling privacy with the use of data for security in applications such as homeland protection, counterterrorism, and health, food and water security. Other challenges arise because the deployments of new data collection and processing devices, such as those used in IoT systems, increase the data attack surface. In this paper, we discuss relevant concepts and approaches for data security and privacy, and identify research challenges that must be addressed by comprehensive solutions to data security and privacy.**

Big Data; Internet of Things; Sensor Networks; Data Confidentiality; Data Trustworthiness; Application Security

I. INTRODUCTION

Data are today more critical and relevant than ever. Technological advances and novel applications, such as sensors, cyber-physical systems, smart mobile devices, cloud systems, data analytics, social networks, Internet of Things (IoT), smart and connected healthcare, are making possible to collect, store, and process huge amounts of data, referred to as *big data*, about everything from everywhere and at any time [1]. However not only today we have technology, such as cloud and high-performance computing systems, for storing and processing huge data sets, we also have sophisticated data analytics capabilities that allow one to extract useful knowledge from data and predict trends and events [2].

Recent advances toward the widespread deployment of sensors, actuators, and embedded computing devices in the physical environment and into physical objects – referred to as Internet of Things (IoT) – will further multiply our ability to collect data and also act on the physical environment. Forecasts by McKinsey&Company estimate that the economic impact of IoT technology by year 2025 will range from 2.7 to 6.2 trillion dollars [3]. Gartner forecasts predict that by the year 2020 20.8 billions of IoT devices will be installed. Such staggering numbers show that IoT will have a major impact, especially when combined with powerful data analytics and knowledge extraction techniques.

The combination of big data and IoT technologies - that we refer to as *pervasive big data (PBD) technologies* - will push a novel generation of data-intensive applications and move automation in a large number of domains, ranging from manufacturing and energy management (e.g. SmartGrid), to healthcare management and urban life (e.g. SmartCities). Applications range from monitoring the moisture in a field of crops, to tracking the flow of products through a factory, to remotely monitoring patients with chronic illnesses and remotely managing medical devices, such as implanted devices and infusion pumps.

However, as our reliance on PBD technologies increases, the security and privacy of data managed by PBD systems become crucial. Damage and misuse of data affect not only single individuals or organizations, but may have negative impacts on entire social sectors and critical infrastructures. As data collection and processing are pervasive, such as in sensor-based systems and recent fog computing systems [4], data protection becomes much more complex compared with when data collection and processing were very much confined within organizations. Increasing numbers of attacks have been reported that aim at stealing data through sophisticated attacks, including insider attacks [5]. Data trustworthiness is another critical issue for several applications, ranging from scientific research to industrial control systems [6]. Finally recent tensions between the use of data for security tasks and data privacy have added yet another dimension to the problem of data security [7].

The problem of data security is not a new problem; research addressing this problem dates back from the early 70's [8]. An early paper by Bertino and Sandhu [9] provides a short history of research efforts on data security focusing especially on access control techniques, such as discretionary and mandatory access control techniques, as these techniques represent a fundamental building block for data security. However early access control techniques were designed for data stored in corporate database systems and therefore today we need to complement such early techniques with other techniques in order to provide full spectrum data protection. Early research on statistical

0730-3157/16 $31.00 © 2016 IEEE
DOI 10.1109/COMPSAC.2016.89

IEEE computer society

databases [10, 11] pioneered initial data privacy techniques. However such early approaches assumed data to be stored in controlled database systems only accessible through specialized interfaces supporting pre-defined statistical queries. Today, applications may need to access data at record level – referred to as microdata access - and thus providing only access at aggregate level may be inadequate for many data applications.

In this paper we continue the discussion on data security that we initiated more than 10 years ago [9], in order to focus on recent relevant challenges and research directions. We first briefly discuss key security requirements. We then focus on big data and identify key challenges in data privacy as today privacy represents a major concern given the widespread ubiquitous data collection by a large variety of organizations. We then focus on IoT which, if one side multiplies our abilities to collect and use data, on the other side greatly expands the data attack surface. We finally outline a few concluding remarks.

II. Basic data security requirements

Our previous paper [9] had identified three basic data security requirements: *confidentiality*, referring to data protection from unauthorized accesses; *integrity*, referring to data protection from unauthorized modifications; and *availability*, referring to assuring that data be available to authorized users. These three requirements are still very critical today. However meeting those requirements is today much more challenging because data attacks are more sophisticated and the data attack surface has expanded, due to increasing data collection activities from many different sources and to data sharing. In addition to these three requirements, *privacy* has emerged as a new critical requirement. Very often data privacy is seen as the same requirement as data confidentiality. There are however some differences between the two requirements. Data privacy requires ensuring data confidentiality because if data are not well protected against unauthorized accesses, privacy cannot be ensured. However privacy has additional issues deriving from the need of taking into account requirements from legal privacy regulations as well as individual privacy preferences. For example, an individual may be fine with sharing his/her own data for research purposes, whereas another individual may not be. Therefore, systems managing privacy-sensitive data may have to collect and record the privacy preferences concerning the individuals to whom the data refer to, referred to as *data subjects.* In other cases, privacy decisions about certain data must be taken by subjects other than the data subjects, as in the case of minors. Also data subjects may change their privacy preferences over time. Addressing privacy thus requires, among other things, systems able to enforce not only the access control policies that an organization may have in place to govern accesses to the data, but also data subject preferences and legal regulations. An example of an access control system able to take into account all these three different sources of data restrictions is by Ni et al. [12]. Finally, it is important to mention that the integrity requirement has been generalized into the *data trustworthiness* requirement. Data trustworthiness refers to making sure not only that data are not modified by unauthorized subjects, but also that data are free from errors, up to date, and originating from reputable sources. Assuring data trustworthiness is thus a difficult problem which often depends on the application domain [6]. Its solution requires combining different techniques, ranging from cryptographic techniques, for digitally signing the data, and access control, for checking that only authorized parties modify the data, to data quality techniques, for automatically detecting and fixing data errors [13], provenance techniques [14], for determining from which sources data originate, and reputation techniques, for assessing the reputation of data sources.

III. A CHARACTERIZATION of big data

In order to discuss about privacy and security issues for big data management, it is crucial to better understand all dimensions related to big data. In this respect four characteristics define big data [2]:

- Volume – data sizes range from terabytes to zettabytes (that is, 1021 bytes).
- Variety – data come in many different formats from structured data, organized according to some structures like the data record, to unstructured data, like images, sounds, and videos which are much more difficult to search and analyze.
- Velocity – in many novel applications, like smart cities and smart planet, data continuously arrive at possibly very high frequencies, resulting in continuous high-speed data streams. It is critical that the time required to act on these data be very small.
- Huge number of data sources – the real value of data sets is when these data sets are integrated and cross-correlated. Integration and cross-correlation among data sets from different sources allow one to uncover information and trends that often cannot be uncovered by looking at a data set in isolation. It is clear that such massive data integration can pose major privacy risks.

Our short characterization emphasizes that volume alone is perhaps the least difficult problem to address when dealing with big data privacy and security. The real challenge arises when we have big volumes of unstructured and structured data continuously arriving from a large number of sources. Addressing such challenge requires a new generation of privacy-enhancing and security techniques designed to ensure data privacy and security by efficiently filtering and transforming large volumes of a wide variety of data.

IV. BIG DATA CONFIDENTIALITY AND PRIVACY

Many privacy enhancing techniques have been proposed over the last fifteen years, ranging from cryptographic techniques, such as oblivious data structures [15] that hide data access patterns, to data anonymization techniques that transform the data to make more difficult to link specific data records to specific individuals [16]. The problem of location privacy has also been the focus of extensive research both in the past and presently [17, 18, 19]. More recently, research efforts have been devoted to investigate privacy-preserving techniques for data on the cloud [20, 21], on smart phones [22], and on social networks [23]. However it is important to note that most proposed privacy-enhancing techniques only focus on privacy and do not address the key problem of reconciling data privacy with an effective use of data, especially when the use is for security applications, including cyber security, homeland protection, health security. The problem of how to reconcile privacy and security is today a major challenge [24]. However to date very few approaches have been proposed that are suitable for large scale datasets. An example of an initial approach along such direction is the scalable protocol for privacy-preserving data matching by Cao et al. [25] which combines secure multiparty computation (SMC) techniques and differential privacy [26] to address scalability issues.

However, just addressing scalability is not sufficient for big data privacy. Comprehensive solutions for big data privacy require addressing many other research challenges. In what follows, we outline relevant research directions.

Data Confidentiality: Data confidentiality is a critical requirement for data privacy. Several data confidentiality techniques and mechanisms exist – the most notable being access control and encryption. Both have been widely investigated. However with respect to access control systems for big data we need approaches for:

- *Merging large numbers of access control policies*. In many cases, big data entails integrating data sets originating from multiple sources; these data sets may be associated with their own access control policies, referred to as "sticky policies", and these policies must be enforced even when a data set is integrated with other data sets. Therefore policies need to be integrated and conflicts solved possibly by using some automated or semi-automated policy integration system [27]. Policy integration and conflict resolution are, however, much more complex when dealing with privacy-aware access control models, such as PRBAC [28], as these models allow one to specify policies that include the purpose for which the access to a protected data item is allowed, obligations arising from the use of data, and special privacy-related conditions that must be meet in order to access the data. Automatically integrating such type of policies and solving conflicts is a major challenge.
- *Automatically administering authorizations for big data and in particular for granting permissions*. If fine-grained access control is required, manual administration on large data sets is not feasible. We need techniques by which authorizations can be automatically granted, possibly based on the user digital identity, profile, and context, and on the data contents and metadata. A first step towards the development of machine learning techniques to support automatic permission assignments to users is by Ni et al. [29]. However more advanced approaches are needed to deal with dynamically changing contexts and situations.
- *Enforcing access control policies on heterogeneous multi-media data*. Content-based access control is an important type of access control by which authorizations are granted or denied based on the content of data. Content-based access control is critical when dealing with video surveillance applications which are important for security. Supporting content-based access control requires understanding the contents of the protected data and this is very challenging when dealing with large multimedia data sets.
- *Enforcing access control policies in big data stores*. Some of the recent big data systems allow their users to submit arbitrary jobs encoded in general programming languages. For example, in Hadoop, users can submit arbitrary MapReduce jobs written in Java. This creates significant challenges in order to efficiently enforce fine grained access control for different users. Although there is some initial work [30] that tries to inject access control policies into submitted jobs, more research is needed on how to efficiently enforce such policies in recently developed big data stores, especially if access control policies are enforced though the use of fine-grained encryption.

Data Privacy: a major issue arising from big data is that by correlating many (big) data sets one can extract unanticipated information. Relevant issues and research directions that need to be investigated include:

- *Techniques to control what is extracted and to check that data are used for the intended purpose*. Content-based access control is one such technique in that it allows one to return certain data to a given user based on the contents of the data [31]. Content-based access control is typically supported in DBMS through the use of view mechanisms [32] or query modifications. Supporting content-based access control stored by systems other than DBMS is much more difficult because of the difficulty of characterizing the conditions that the data contents must verify in order to be returned to a user. In relational DBMS such conditions are easily expressed as SQL queries. Research is needed to design techniques able to support content-based access control for a variety of data management systems. Another more difficult question to address is how to verify that data, returned to a user, are used for the intended purpose. An initial pioneering approach was proposed that associates with each data item a set of possible purposes, from an ontology of purposes, for which the data can be used [33]. When a user accesses some data items, the user indicates in the access request the purpose(s) for which the data items are being accessed. The query purposes are then matched against the purposes associated with the data items to verify that the query purposes comply with the intended use associated with the requested data items. Such an approach needs to be complemented with techniques for automatically and securely identifying the data access purposes, instead of relying on indications given by users as part of their access requests.

- *Support for both personal privacy and population privacy*. In the case of population privacy, it is important to understand what is extracted from the data as this may lead to discrimination. Also when dealing with security with privacy, it is important to understand the tradeoff of personal privacy and collective security.
- *Usability of data privacy policies*. Policies must be easily understood by users. We need tools for the average users and we need to understand user expectations in terms of privacy.
- *Privacy implications on data quality*. Recent studies have shown that people lie especially in social networks because they are not sure that their privacy is preserved. This results in a decrease in data quality that then affects decisions and strategies based on these data.
- *Risk models.* Different types of relationship of risks with big data can be identified: (a) big data can increase privacy risks; (b) big data can reduce risks in many domains (e.g. national security). The development of models for these two types of risk is critical in order to identify suitable tradeoff and privacy-enhancing techniques to be used.
- *Data ownership.* The question about who is the owner of a piece of data is often a difficult question. It is perhaps better to replace this concept with the concept of stakeholder. Multiple stakeholders can be associated with each data item. The concept of stakeholder ties well with risks. Each stakeholder would have different (possibly conflicting) objectives and this can be modeled according to multi-objective optimization. In some cases, a stakeholder may not be aware of the others. For example a user to whom a data item pertains (and thus a stakeholder for the data item) may not be aware that a law enforcement agency is using this data item. Technological solutions need to be investigated to eliminate conflicts.
- *Data lifecycle framework.* A comprehensive approach to privacy for big data needs to be based on a systematic data lifecycle approach. Phases in the lifecycle need to be identified and their privacy requirements and implications need to be identified. Relevant phases include:
  - ➢ Data acquisition. We need mechanisms and tools to prevent devices from acquiring data about other individuals when devices like Google glasses are used. For example we need mechanisms able to automatically prevent devices from recording/ acquiring data when in certain locations [22] or notify a user that recording devices are around. We also need techniques by which each recorded subject may be able to express his/her preferences about the use of the data.
  - ➢ Data sharing. Users need to be informed about data sharing/transfer to other parties. Always informing users is, however, not always possible as sometimes information about data transfer and use is confidential to the organization's missions. It is thus critical to devise legal guidelines on such issue based on which technical mechanisms can be designed.

V. IoT Risks

IoT represents an important emerging trend that according to various forecasts (see [3] for one such forecast) will have a major economic impact. However, as discussed in [34], while on one side, IoT will make many novel applications possible; on the other side IoT increases the risks of cyber security attacks to data. In addition, because of its fine-grained, continuous, and pervasive capabilities for data acquisition and control/actuation capabilities, IoT raises concerns about privacy and safety. A study by HP about the most popular devices in some of the most common IoT application domains show a high average number of vulnerabilities per device [35]. On average, 25 vulnerabilities were found per device. For example, 80% of devices failed to require passwords of sufficient complexity and length, 70% did not encrypt local and remote traffic communications, and 60% contained vulnerable user interfaces and/or vulnerable firmware [35].

IoT systems are at high risks for several reasons [34]. They do not have well defined perimeters, are highly dynamic, and continuously change because of mobility. IoT systems are also highly heterogeneous with respect to communication medium and protocols, platforms, and devices. IoT systems may also include "objects" not designed to be connected to the Internet. Finally, IoT systems, or portions of them, may be physically unprotected and/or controlled by different parties. Attacks, against which there are established defense techniques in the context of conventional information systems and mobile environments, are thus much more difficult to protect against in the IoT. The OWASP Internet of Things Project [36] has shown that many IoT vulnerabilities arise because of the lack of adoption of well-known security techniques, such as encryption, authentication, access control, and role-based access control. Lack of security techniques adoption may certainly be due to security unawareness by IT companies involved in the IoT space and by end-users or to cost reasons. However another reason is that existing security techniques, tools, and products may not be easily deployed to IoT devices and systems, for reasons such as the variety of hardware platforms and limited computing resources of many types of IoT devices.

Data privacy is particularly critical in the context of IoT. As medical and well-being devices are increasingly been adopted by users, and personalized medicine and health care applications are being designed and deployed that rely on continuous fine-grained data acquisition from these devices, the human body is becoming a rich source of information. Such information is typically collected from devices and then uploaded to some cloud and/or transmitted to other devices, such as mobile phones, which in turn may forward the information to other parties. The collected information is typically very rich and often includes meta-data, such as location, time, and context, thus making possible to easily infer personal habits, behaviors, and preferences of individuals. It is thus clear that on one side such information has to be carefully protected by all parties involved in its acquisition, management, and use, but also that users should be provided with suitable, easy to use tools for protecting their privacy and support anonymity depending on specific contexts [22].

VI. IOT DATA SECURITY – INITIAL EFFORTS

Addressing IoT data security requires extending or re-engineering existing security solutions as well as to develop new solutions to fit the specific requirements of IoT. Such solutions must ensure protection while data are transmitted and processed at the devices. In addition, in many cases, data availability is critical and therefore solutions minimizing data losses must be devised. In what follows, we survey some projects that cover different aspects of data security solutions and report experience from these projects.

Cryptographic Protocols. The area of encryption techniques is an active and important research area. However, just devising new encryption techniques is not sufficient to secure data. As pointed out by Schneier [37], strong security can be achieved only if cryptographic protocols are implemented and deployed correctly. The limitation of device computing resources and the differences in such resources across different devices make performance an additional critical challenge. Also, when dealing with very large IoT systems, efficient encryption key management is critical. Recently an efficient certificate-less signencryption protocol, that is, a protocol not requiring key certificates and supporting both message encryption and authentication, has been proposed and compared with other protocols on different devices, including Raspberry Pi2, and Android [38]. As this protocol does not use expensive pairing operations, it is highly efficient compared to other similar protocols.

Another interesting project is related to techniques and protocols for efficient authentication operations for networked vehicles [39]. The main requirement is that multiple concurrent authentication operations have to be supported with real-time response time. Response time is critical in that, if a vehicle has to stop suddenly, information about this event has to reach the other vehicles in a very short time so that these vehicles have enough time to break. Therefore it is crucial that authentication operations both at the sender and the receivers have minimal overhead. To address such requirement, the implementation of the authentication operations takes advantage of the GPU usually present in systems-on-chips today used in vehicles.

Finally another interesting project focuses on encryption protocols for networks consisting of small sensors and drones. In such networks, sensors are on the ground and acquire data of interest from the environment and drones fly over the sensors to collect and aggregate data from sensors [40]. The main issue here is to save energy and to make sure that drones do not have to wait too long for sensors to start generating encryption keys. To address such requirement, the approach is to use low power listening (LPL) techniques [41] at the sensors and dual radio channels at the drones. In this way, the sensors can timely start generating the cryptographic keys when drones approach.

Results from those projects show that a careful engineering of cryptographic protocols is critical to the effective deployment of cryptographic protocols in IoT. In particular, it is critical to analyze in details the protocols in order to determine the expensive operations so to replace or optimize them, and to understand how to take advantage of specific hardware features of the devices in order to enhance the implementation of the different steps of the protocols.

Application Security. Protecting applications is crucial for data security as attacks to steal data often use application vulnerabilities as stepping stones. It is important to notice that even though today we have several techniques for program analysis and hardening, such techniques need substantial extensions to fit IoT devices.

A first example is represented by techniques to protect programs against code injection attacks and code reuse attacks [42]. Both those attacks aim at modifying the execution flow of applications in order to, for example, modify data acquired from the external environment [43]. An approach to protect against those attacks is to instrument the application binary code by inserting a static check statement before any instruction that modifies the program counter. Such check verifies that the target's address, to which the program execution has to move, is the correct address, that is, that the next instruction to be executed is the expected one and not an instruction to which the attacker is trying to redirect the execution. Such technique has been shown to be quite efficient as the run-time overhead introduced by these additional checks ranges between 0.51% and 12.22% based on the benchmarked applications [42]. However the application of this technique requires identifying for each platform the critical instructions, that is, the instructions that can modify the program counter. These instructions are different for different platforms; such variations thus require devising specific instrumentation techniques for specific platforms.

Another approach to application security focuses on protecting against memory vulnerabilities [44] for applications written in variant of the C language specific for

TinyOS applications. Such an approach statically analyzes an application to identify memory vulnerabilities. As in some cases it is not possible to statically determine if a certain piece of code will lead to a vulnerability at run-time, the approach adds some code to check at run-time whether a vulnerability occurs. Also in this project, the main issue is to minimize the run-time overhead as this is critical for devices with limited capabilities.

Both those projects show that significant work is required to modify existing application program security techniques for use in IoT systems.

Network Security. Security techniques at network level are critical in order to minimize data losses. Such minimization is crucial for many applications, such as monitoring applications and control systems. In order to minimize data losses, it is critical to be able to quickly diagnose the cause of data packet losses so to quickly repair the network. A recent project [45] has addressed this requirement by developing a fine-grained analysis (FGA) tool that investigates packet losses and reports their most likely cause. Such FGA tool is based on profiling the wireless links between the nodes as well as their neighborhood, by leveraging resident parameters, such as RSSI and LQI, available within every received packet. By using those profiles, the FGA tool is able to determine whether the cause of a packet loss is a link that has been jammed or a sensor that has been compromised. In the former case, the FGA tool is able to quite reliably detect the source of interference. The design of the system is fully distributed and event-driven, and its low overhead makes it suitable for resource-constrained entities such as wireless motes.

This project is however just an initial approach. Research is needed to develop more advanced FGA tools able to deal with mobile systems and heterogeneous communication technologies which may require using different profiling parameters.

VII. IoT Data Security – Research Directions

Securing IoT data requires, however, the use of other techniques [34], in addition to the techniques discussed in the previous sections. Data confidentiality requires access control to govern access to the data by taking into account information on data provenance and metadata concerning the data acquisition context, such as location and time. Therefore early work on temporal [46] and location [47] based access control is today very relevant. Data trustworthiness is particularly challenging in an IoT context as data acquired and transmitted by IoT devices may be of poor quality. Reasons for poor quality include bad device calibration, device errors, and deliberate data deception attacks. Solutions like data fusion need to be revised and extended to deal with dynamic environments and large-scale numbers of heterogeneous data sources. Understanding how to deploy and configure security tools for IoT is also very challenging as one has to optimally trade-off security risks with costs and energy consumption [48]. Finally privacy introduces new challenges, including how to prevent personal devices from acquiring and/or transmitting information concerning the user location and other context information.

VIII. Conclusions

This paper has discussed research directions in big data confidentiality and privacy, and IoT data security. Another relevant research area which has been the focus of intense research in the past ten years is the area of data security and privacy on the cloud. This area has seen significant research in different directions, such as for example approaches to support privacy-preserving fine-grained attribute-based access control on the cloud [49, 50], and provable possession of data on the cloud [51]. Also the area of data privacy in social networks has received significant focus. One of the key issues emerging from such research is that in social networks collaborative approaches are needed for access control [52, 53]. The reason is that in social networks, a given piece of data, such as a picture, may refer to multiple social network users, and it is thus crucial that all such users be able to express their privacy preferences when sharing the piece of information.

In addition, to the research directions mentioned so far in the paper, there are two other additional research directions that we would like to emphasize:

- Data protection from insider threat - protection against insider threat requires combining many different techniques, including context-based access control, anomaly detection in data access and use [54], and user behavior monitoring. User behavior monitoring however may entail privacy issues and therefore it requires a careful trade-off between security risks and individual privacy.
- Privacy-aware software engineering – engineering software to provide strong privacy assurance requires, among other things, to identify the code portions that deal with sensitive data, the ability of applications to work on anonymized data and to deal with lack of permissions depending on specific spatial and temporal contexts; also as forensic tools are today able to recover memory contents after applications complete their execution, it is critical that applications scrub memory to permanently delete sensitive data. Finally tools are needed able to create profiles of expected usage of privacy-sensitive data by application programs and use these profiles at run-time to detect anomalies in the data use by the applications [55].

As final remark we would like to mention that addressing the today and tomorrow challenges in data security and privacy require multidisciplinary research drawing from many different areas, including computer science and engineering, information systems, statistics, risk models, economics, social sciences, political sciences, human factors, psychology. We believe that all these perspectives are needed to achieve effective solutions to the problem of privacy in the era of big data and pervasive data acquisition and use, and especially, to the problem of reconciling security with privacy.

Acknowledgments

The work reported in this paper has been partially funded by the Purdue Cyber Center and the National Science Foundation under grants CNS-1111512 and ACI-1547358. The contents of Section 3 are partially based on the results of the privacy session (chaired by the author of the present paper) organized as part of the NSF Big Data Security and Privacy Workshop, September 16-17, 2014, http://csi.utdallas.edu/events/NSF/NSF%20workshop%202014.htm. I would like to thank the workshop Chair, Professor Bhavani Thuraisingham, the chair of the workshop security session, Professor Murat Kantarcioglu, and the workshop participants for the discussions during the workshop which lead to an initial draft agenda for privacy research for big data.

References

[1] “Data, data everywhere”, The Economist, 25 February 2010, available at http://www.economist.com/node/15557443 (Downladed on April 30, 2012).

[2] E. Bertino, “Big Data – Opportunities and Challenges”, Panel Position Paper, Proceedings of the 37th Annual IEEE Computer Software and Applications Conference, COMPSAC 2013, Kyoto, Japan, July 22-26, 2013.

[3] J. Manyika, M. Chui, J. Bughin, R. Dobbs, P. Bisson, and A. Marrs. Disruptive technologies: Advances that will transform life, business, and the global economy. http://www.mckinsey.com/insights/business technology/disruptive_technologies, May 2013.

[4] E. Bertino, S. Nepal, R. Ranjan, “Building Sensor-Based Big Data Cyberinfrastructures”, IEEE Cloud Computing 2(5): 64-69 (2015).

[5] E. Bertino. Data Protection from Insider Threats. Synthesis Lectures on Data Management, Morgan & Claypool Publishers 2012

[6] Elisa Bertino, “Data Trustworthiness - Approaches and Research Challenges”, Data Privacy Management, Autonomous Spontaneous Security, and Security Assurance - 9th International Workshop, DPM 2014, 7th International Workshop, SETOP 2014, and 3rd International Workshop, QASA 2014, Wroclaw, Poland, September 10-11, 2014. Revised Selected Papers.

[7] Elisa Bertino, “Big Data - Security and Privacy”, Proceedings of the 2015 IEEE International Congress on Big Data, New York City, NY, USA, June 27 - July 2, 2015.

[8] D. E. Denning, P. J. Denning. “Data Security”, ACM Comput. Surv. 11(3): 227-249 (1979).

[9] E. Bertino, R. Sandhu, “Database Security – Concepts, Approaches, and Challenges”, IEEE Trans. Dependable Sec. Comput. 2(1):2-19 (2005).

[10] M. D. Schwartz, D. E. Denning, P. J. Denning, “Linear Queries in Statistical Databases” ACM Trans. Database Syst. 4(2): 156-167 (1979).

[11] D. E. Denning, P. J. Denning, M D. Schwartz. “The Tracker: A Threat to Statistical Database Security”, ACM Trans. Database Syst. 4(1): 76-96 (1979).

[12] Q. Ni, S. Xu, E. Bertino, R. S. Sandhu, W. Han, “An Access Control Language for a General Provenance Model”, Secure Data Management, Proceedings of the 6th VLDB Workshop, SDM 2009, Lyon, France, August 28, 2009.

[13] C. Batini, M. Scannapieco. Data and Information Qaulity – Dimensions, Principles and Techniques. Springer, 2016.

[14] S. Sultana, E.Bertino: A Distributed System for The Management of Fine-grained Provenance. J. Database Manag. 26(2): 32-47 (2015)

[15] H. X. Wang, K. Nayak, C. Liu, E. Shi, E. Stefanov, Y. Huang, “Oblivious Data Structures”, IACR Cryptology ePrint Archive 2014: 185.

[16] J.-W. Byun, A. Kamra, E. Bertino, N. Li, “Efficiently k-Anonymization Using Clustering Techniques”, Proceedings of the 12th International Conference on Database Systems for Advanced Applications (DASFAA 2007), Bangkok, Thailand, April 9-12, 2007. LNCS, Springer.

[17] G. Ghinita, P. Kalnis, A. Khoshgozaran, C. Shahabi, K.-L. Tan, “Private queries in location based services: anonymizers are not necessary”, Proceedings of the ACM SIGMOD International Conference on Management of Data, SIGMOD 2008, Vancouver, BC, Canada, June 10-12, 2008**.**

[18] R. Paulet, Md. G. Kaosar, X. Yi, E. Bertino, “Privacy-Preserving and Content-Protecting Location Based Queries”, IEEE Trans. Knowl. Data Eng. 26(5): 1200-1210, 2014.

[19] M. L. Damiani, E. Bertino, C. Silvestri, “The PROBE Framework for the Personalized Cloaking of Private Locations”, Transactions on Data Privacy 3(2): 123-148, 2010.

[20] M. Nabeel, E. Bertino, “Privacy Preserving Delegated Access Control in Public Clouds”, IEEE Trans. Knowl. Data Eng. 26(9): 2268-2280, 2014.

[21] S.-H. Seo, M. Nabeel, X. Ding, E. Bertino, “An Efficient Certificateless Encryption for Secure Data Sharing in Public Clouds”, IEEE Trans. Knowl. Data Eng. 26(9): 2107-2119, 2014.

[22] B. Shebaro, O. Oluwatimi, D. Midi, E. Bertino, “IdentiDroid: Android can finally Wear its Anonymous Suit”, Transactions on Data Privacy 7(1): 27-50, 2014.

[23] B. Carminati, E. Ferrari, M. Viviani, Security and Trust in Online Social Networks. Morgan&Claypool, 2013.

[24] E. Bertino, “E. Bertino, “Security with Privacy – Opportunities and Challenges”, Panel Position Paper, Proceedings of the 38th Annual IEEE Computer Software and Applications Conference, COMPSAC 2014, Vasteras, Sweden, July 21-25, 2014.

[25] J. Cao, F.-Y. Rao, E. Bertino, M. Kantarcioglu, “A Hybrid Private Record Linkage Scheme: Separating Differentially Private Synopses from Matching Records”, Proceedings of the 31st International Conference on Data Engineering (ICDE), Seoul (Korea), April 13-17, 2015.

[26] C.Dwork, A. Roth, “The Algorithmic Foundations of Differential Privacy”, Foundations and Trends in Theoretical Computer Science 9(3-4): 211-407, 2014.

[27] D. Lin, P. Rao, E. Bertino, N. Li, J. Lobo, “EXAM: a Comprehensive Environment for the Analysis of Access Control Policies”, International Journal of Information Security (IJIS), Vol.9, No.4, pp.253-273, August 2010.

[28] Q. Ni, E.Bertino, J. Lobo, C. Brodie, C.M.Karat, J. Karat, A. Trombetta, “Privacy-Aware Role-Based Access Control”, ACM Transactions on Information and System Security, Vol.13, No.3, Article 24, July 2010.

[29] Q. Ni, J. Lobo, S. B. Calo, P. Rohatgi, E. Bertino, “Automating role-based provisioning by learning from examples”, Proceedings of the 14th ACM Symposium on Access Control Models and Technologies, SACMAT 2009, Stresa, Italy, June 3-5, 2009.

[30] H. Ulusoy et al. “Vigiles: Fine-Grained Access Control for MapReduce Systems”, Proceedings of the 2014 IEEE International Congress on Big Data, Anchorage, AK, USA, June 27 - July 2, 2014.

[31] E. Bertino, G. Ghinita, A. Kamra, “Access Control for Databases: Concepts and Systems”, Foundations and Trends in Databases, 3(1-2): 1-148, 2011.

[32] E. Bertino, L.M.Haas, “Views and Security in Distributed Database Management Systems”, Proceedings of the International Conference

on Extending Database Technology (EDBT'88), Venice, Italy, March 14-18, 1988, Springer 1988 Lecture Notes in Computer Science.

[33] J.W. Byun, E. Bertino, N. Li, "Purpose based access control of complex data for privacy protection", Proceedings of the 10th ACM Symposium on Access Control Models and Technologies, SACMAT 2005, Stockholm, Sweden, June 1-3, 2005.

[34] E. Bertino, "Data Security and Privacy in the IoT", Keynote Abastract, Proceedings of Proceedings of the 19th International Conference on Extending Database Technology, EDBT 2016, Bordeaux, France, March 15-16, 2016, Bordeaux, France, March 15-16, 2016.

[35] K. Rawlinson. HP study reveals 70 percent of internet of things devices vulnerable to attack. http://www8.hp.com/us/en/hp-news/

[36] https://www.owasp.org/index.php/OWASP_Internet_of_Things_Project

[37] B. Schneier, "Cryptography is Harder than It Looks", Computing Edge, March 2016.

[38] S.-H. Seo, J. Won, E. Bertino, "pCLSC-TKEM: a Pairing-free Certificateless Signcryption-tag Key Encapsulation Mechanism for a Privacy-Preserving IoT", submitted for publication to Transactions on Data Privacy.

[39] A. A. Mudgerikar, A. Singla, I. Papapanagiotou, A.A. Yavuz, "HAA: Hardware-Accelerated Authentication for Internet of Things in Mission Critical Vehicular Networks", Proceedings of the 34th International Conference for Military Communications (IEEE MILCOM 2015), October 2015.

[40] J. Won , S.-H. Seo, E. Bertino, "A Secure Communication Protocol for Drones and Smart Objects", Proceedings of the 10th ACM Symposium on Information, Computer and Communications Security, ASIA CCS '15, Singapore, April 14-17, 2015.

[41] http://www.tinyos.net/tinyos-2.x/doc/html/tep105.html

[42] J. Habibi, A. Panicker, A. Gupta, E. Bertino, "DisARM: Mitigating Buffer Overflow Attacks on Embedded Devices", Proceedings of the 9th International Conference on Network and System Security, NSS 2015, New York, NY, USA, November 3-5, 2015.

[43] J. Habibi, A. Gupta, S. Carlsony, A. Panicker, E. Bertino, "MAVR: Code Reuse Stealthy Attacks and Mitigation on Unmanned Aerial Vehicles", Proceedings of the 35th IEEE International Conference on Distributed Computing Systems, ICDCS 2015, Columbus, OH, USA, June 29 - July 2, 2015.

[44] D. Midi. T. Payer, E. Bertino, "nesCheck: Memory Safety for Embedded Devices", submitted for publication, 2016.

[45] D. Midi, E. Bertino, "Node or Link? Fine-Grained Analysis of Packet Loss Attacks in Wireless Sensor Networks", ACM Transactions on Sensor Networks, accepted for publication, in print, 2016.

[46] E. Bertino, P. Bonatti, E. Ferrari, "TRBAC: A temporal role-based access control model", ACM Trans. Inf. Syst. Secur. 4(3): 191-233 (2001).

[47] M. L. Damiani, E. Bertino, B. Catania, P. Perlasca, "GEO-RBAC: A spatially aware RBAC", ACM Trans. Inf. Syst. Secur. 10(1) (2007).

[48] N. Rullo, D. Midi, E. Serra, E. Bertino, "Strategic Security Resource Allocation", Poster Paper, Proceedings of the 36th IEEE International Conference on Distributed Computing Systems, ICDCS 2016, Nara, Japan, June 27 – June 30, 2016.

[49] M. Nabeel, N. Shang, E. Bertino, "Privacy Preserving Policy-Based Content Sharing in Public Clouds", IEEE Trans. Knowl. Data Eng. 25(11): 2602-2614 (2013).

[50] M. I. Sarfraz, M. Nabeel, J. Cao, E. Bertino, "DBMask: Fine-Grained Access Control on Encrypted Relational Databases", Proceedings of the 5th ACM Conference on Data and Application Security and Privacy, CODASPY 2015, San Antonio, TX, USA, March 2-4, 2015.

[51] G. Ateniese, M. T. Goodrich, V. Lekakis, C. Papamanthou, E. Paraskevas, R. Tamassia, "Accountable Storage.", IACR Cryptology ePrint Archive 2014: 886 (2014).

[52] A. C. Squicciarini, M. Shehab, F. Paci, "Collective privacy management in social networks", Proceedings of the 18th International Conference on World Wide Web, WWW 2009, Madrid, Spain, April 20-24, 2009.

[53] A. C. Squicciarini, F. Paci, S. Sundareswaran, "PriMa: an effective privacy protection mechanism for social networks", Proceedings of the 5th ACM Symposium on Information, Computer and Communications Security, ASIACCS 2010, Beijing, China, April 13-16, 2010.

[54] A. Sallam, E. Bertino, S.R. Hussain, D. Landers, R. M. Lefler, D. Steiner, "DBSAFE – An Anomaly Detection System to Protecte Databases from Exfiltration Attempts", accepted for publication in IEEE Systems Journal, 2016, in print.

S. R. Hussain, A. Sallam, E. Bertino, "DetAnom: Detecting Anomalous Database Transactions by Insiders", Proceedings of the 5th ACM Conference on Data and Application Security and Privacy, CODASPY 2015, San Antonio, TX, USA, March 2-4, 2015.