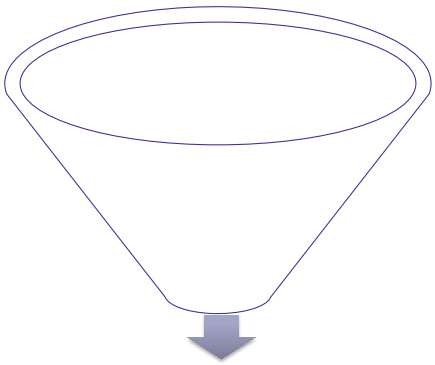
Research Statement

Xuhua Ding

School of Computing and Information Systems
Singapore Management University
Tel: (65) 6828-0683; Email: xhding@smu.edu.sg
12/12/2023

Background

Recent years have witnessed a remarkable landscape change of cyberspace attacks. Those old-fashioned attacks violently trashing critical user data or sabotaging a service by flooding the server are superseded by today's attacks which are financially driven and are targeted on stealing valuable data assets. As a result, data privacy has overtaken data availability to become the mainstream security concern. It becomes even more prominent in the new computation paradigms, such as cloud computing and Database-as-a-Service, where data are processed and managed by heterogeneous entities and platforms.



My research centers on data security and privacy protection throughout their lifetime. By and large, the data lifecycle can be divided into three stages: (1) data generation, (2) data processing and (3) data retrieval/storage. Data generation refers to the procedure whereby data are produced and transported to their designated locations. The logical essence of data processing is the computation prescribed by a function taking input data and generates output data, whereas in practice such a function is implemented by a program running on a platform. In the third stage, data are stored and organized in a storage and are accessed by user processes following certain retrieval paradigm. My work is aligned with these three stages. I undertake to investigate the security threats at each stage and to cope with them using a hybrid of algorithmic, cryptographic and system security techniques.

Research Areas

1. Security & Privacy in Data Generation

My present research focuses on designing secure systems to defeat kernel-level attacks on sensitive data during the generation procedure. One of the efforts aims at the data I/O path between the peripheral devices and the applications. In today's commodity platforms, the I/O path is barely protected. We have designed and implemented a hypervisor-based protection scheme [7], which protects I/O

security against untrusted operating systems without imposing high overhead to the system or the application. My research also addresses user-oriented data security, such as password input. One of the outcomes [6] in this area is to protect user password input during an authentication session against both the (corrupted) operating system and malicious applications in a platform.

2. Security & Privacy in Data Processing

Access pattern privacy only addresses data storage and retrieval. Privacy is also a major concern in miscellaneous applications, where data are processed among entities with divergent trustworthiness. The research problem here is how to preserve data privacy against unauthorized entities without disrupting the applications.

My study in this area covers both network and database applications. The main contributions are the privacy-preserving protocols, namely efficient cryptographic constructions whereby a desired computation function can be performed by an untrusted entity without exposing the actual data in use. My research results on network applications consist of a wireless broadcast scheduling protocol [3] where the broadcast server makes an optimal schedule without knowing each user's individual preference; and a sensor network query scheme where the untrusted sensor network gateway correctly answers a user's query without knowing the query and the actual result [4]. I also study privacy preserving in database applications. The results include a novel scheme protecting the privacy of text search queries [5] by padding queries using a cluster of keywords sharing the semantic distribution; and a privacy-preserving equi-join algorithm for outsourced data [8].

Privacy-preserving protocols are only applicable to those data processing procedures that can be modeled as a computational function. Nonetheless, many data processing do not fall into that category, for instance, a password management process or the SSL operations of a web server. In these applications, the data privacy is dependent on the intactness of software execution. My research also addresses how to establish a trustworthy environment against attacks from user space and/or kernel space [9,10,11].

3. Privacy in Data Storage/Retrieval

My research in this area is motivated by the Database-as-a-Service applications where data are stored at and retrieved from a remote untrusted storage server. My work focuses on protecting the access pattern privacy which ensures that the storage server obtains no useful information about the data being fetched. The access pattern privacy is the strongest notion of privacy protection, because it addresses the data storage and retrieval, the fundamental processes in all applications.

The known techniques for access pattern protection are Private Information Retrieval (PIR) and Oblivious RAM (ORAM). Unfortunately, neither is practical due to the overwhelming computation overhead. The contributions of my work are exactly on the computation cost reduction. The first result [1] is a new oblivious shuffle algorithm and a more recent result [2] removes the need for shuffling the entire database. By shuffling only a portion of the database using the new shuffle algorithm, the cost of access pattern protection is significantly reduced, especially when the database comprises less than tens of billions of items.

These outcomes are a giant leap towards the realization of access pattern privacy for small and medium databases in practice. They have bridged the gap between the sophisticated theoretic constructions and the performance-demanding real-world applications.

Looking into the future, my research will advance in two dimensions. One dimension is on the scalability. I will tackle new the challenges imposed by large scale data sets which clearly demand highly efficient protection mechanisms. The other dimension is on the emerging applications, such as cloud computing and mobile sensing networks. My methodology is to take a hybrid approach using both cryptographic and system security techniques (in particular, virtualization techniques) to safeguard sensitive data for endusers.

Selected Publications and Outputs

[1] Xuhua Ding and Y. Yang and R. Deng and S. Wang, *A New Hardware-assisted PIR with O(n) Shuffle Cost*. In International Journal of Information Security, Vol 9, Issue 4, August 2010.
[2] Xuhua Ding and Y. Yang and R. Deng, *Database Access Pattern Protection Without Full- shuffles*. In 2011, 6, 1, IEEE Transactions on Information Forensics and Security (TIFS).
[3] Xuhua Ding and S. Wang and B. Zheng, *Secure Real-time User Preference Collection for Broadcast Scheduling*. In Proceedings of the 2nd International Conference on Security and Privacy in Communication Networks (SecureComm), 2006.
[4] E. Cristofaro and Xuhua Ding and G. Tsudik, *Privacy-preserving Querying in Sensor Networks*. In Proceedings of the 18th International Conference on Computer Communications and Networks (ICCCN), 2009
[5] H. Pang and Xuhua Ding and X. Xiao, *Embellishing Text Search Queries to Protect User Privacy*. In Proceedings of the 36th International Conference on Very Large Data Bases (VLDB), 2010
[6] Y. Cheng and Xuhua Ding, *Virtualization Based Password Protection Against Malware in Untrusted Operating Systems*, In Proceedings of TRUST 2012.
[7] Y. Cheng, Xuhua Ding, and Huijie Robert Deng, *DriverGuard: virtualization based fine-grained protection on I/O flows*, 09/2013, 16, 2, ACM Transactions on Information and System Security (TISSEC)
[8] H. Pang and Xuhua Ding, *Privacy-Preserving Ad Hoc Equi-Join on Outsourced Data*, 09/2014, 39, 3, ACM Transactions on Database Systems (TODS)

[9] Y. Cheng, Zongwei Zhou, Yu Miao, Xuhua Ding, and Huijie, Robert Deng, *ROPecker: A Generic and Practical Approach for Defending Against ROP Attack*, The 21th Annual Network and Distributed System Security Symposium (NDSS'14)
[10] Y. Cheng, Xuhua Ding, and Robert Deng, *Efficient Virtualization-Based Application Protection Against Untrusted Operating System*, ACM AsiaCCS 2015
[11] S. Zhao, Xuhua Ding, *On the Effectiveness of Virtualization Based Isolation on Multicore Platforms*, IEEE EuroS&P 2017