Hardware Assisted Secure Data Aggregation and Analysis

Prime Offeror:
The University of Texas at Dallas
800 W. Campbell Rd.
Richardson, TX 75080-3021
DUNS Number: 80-018-8161
CAGE Number: 0W921 FICE 009741

Technical contact:
Dr. Murat Kantarcioglu
Computer Science Department – EC31
Voice: (972) 883-6616
Fax: (972) 883-2349
Email: muratk@utdallas.edu

Technical contact:
Dr. Zhiqiang Lin
Computer Science Department – EC31
Voice: (972) 883-4244
Fax: (972) 883-2349
Email:zhiqiang.lin@utdallas.edu

Administrative contact:
Emily Lacy
Office of Sponsored Projects – AD15
Voice: (972) 883-2041
Fax: (972) 883-2310
Email: emily.lacy@utdallas.edu

Estimated Cost: $50,000
Period of performance: 12 Months

1 Overview

Motivation. Most of the organizations today store increasing amounts of sensitive data collected from various sensors to deliver various value added services. At the same time, there are increasing concerns related to the security, privacy and accountability of such collected sensitive sensor data. In the last couple of years, many data theft incidents have been reported due to various reasons including unpatched software, insider attacks, compromise of the underlying operating system, weak passwords, improper system configuration, lost unencrypted backup etc. [3]. For example, computers attached to the Internet can easily get infected by malwares. Once infected these programs can easily monitor all types of data activity and transfer these sensitive information to the outsiders. Also, attackers could easily use some bug or badly configured software to access critical data. Once the attacker controls the servers processing the sensitive sensor data, all the sensitive data could be accessed by the attacker and confidential data could be leaked without being detected. Therefore, it is imperative to develop a practical solution that can collect and process the sensor data securely under various attacks such as insiders, etc.

State-of-the-Art. The most straightforward and also practical way is to encrypt the data sent by sensors during transit and at rest. However, this often leads to another challenge — how to perform data processing over encrypted data, especially in the cloud settings. Over the past decades, numerous techniques have been proposed to support basic computations over encrypted data. From the security perspective, the most notable of such techniques is the fully-homomorphic mechanism (FHE) that can support any computation over encrypted data [2] revealing nothing other than potentially the size of the output. FHE, however, remains extremely expensive and difficult to use as a basis for secure data outsourcing.

Most recently, there also has been a great attention from system community of using the newly introduced Intel Software Guard eXtension (SGX) [7] to achieve trusted computing including data secrecy protection and integrity. At a high level, SGX allows an application or part of an application to run inside a secure *enclave*, which is an isolated execution environment. SGX hardware, as a part of the CPU, protects the enclave against malicious software, including the operating system, hypervisor, or even low-level firmware code (e.g., SMM) from compromising its integrity. This *isolation* enabled by SGX is particularly useful in cloud computing environments, where customers cannot control the infrastructure owned by cloud providers. Consequently, initial exploration into secure execution has begun for cloud settings. Haven [1] pioneered the idea of enabling unmodified application binaries to run on SGX by utilizing an OS library [5]. VC3 [6] suggested privacy-aware data analytics on Hadoop in the cloud.

Proposed Research. Much like Intel VT-x that has made virtualization practical, we believe Intel SGX will highly likely make secure and trustworthy computing such as sensor data processing in a hostile environment practical.

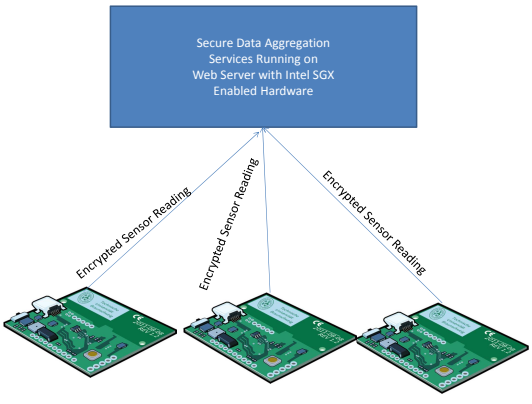

Figure 1: The architecture of our proposed system.

In this project, our goal is to protect the confidentiality and integrity of the sensor data used in mission-critical applications by leveraging recent advances in secure hardware design such as Intel software guard extension [4]. Basically, SGX [4] is an instruction set extension that makes it possible to create protected execution environments called enclaves. These enclaves are protected by the processor. Therefore, the processor will not allow any attempt to read or write the memory of a running enclave from outside (e.g., even OS cannot access the enclave memory). Using these enclaves, small amount

of code can be executed inside the enclave and can be used to to process encrypted sensor data collected securely and efficiently.

In our architecture, as shown in Figure 1, each sensor will be sending encrypted data to web service running in the enclave. Enclave based web service will be aggregating the incoming data (e.g., computing certain statistics, basic data mining models, generating alarms/actions based on predefined rules). In addition, computed results and aggregated data will be stored in encrypted format on the server storage. One of our goal in this project will be to design enclave based web services that are self-contained to prevent certain attack scenarios (e.g., disclosure due to potential side channel attacks).

Deliverables. Our deliverables include: 1) Initial design of the web service that can run in the Intel SGX enclave securely. 2) Software prototype that can do few basic data aggregation services on simulated sensor data (based on the industry partners feedback will extend it to few real sensors). 3) Documentation that explains the design and coding details.

Timeline. We plan to finish the initial design in 3 months, development of the basic prototype in 6 months after the design phase, and experimental evaluation during the last 3 months of the project.

References

[1] Andrew Baumann, Marcus Peinado, and Galen Hunt. Shielding applications from an untrusted cloud with haven. pages 267–283.

[2] C. Gentry. *A fully homomorphic encryption scheme*. PhD thesis, Stanford University, 2009. crypto.stanford.edu/craig.

[3] Tom Zeller Jr. An ominous milestone: 100 million data leaks. *New York Times*, December 18 2006.

[4] Frank McKeen, Ilya Alexandrovich, Alex Berenzon, Carlos V. Rozas, Hisham Shafi, Vedvyas Shanbhogue, and Uday R. Savagaonkar. Innovative instructions and software model for isolated execution. In *Proceedings of the 2nd International Workshop on Hardware and Architectural Support for Security and Privacy (HASP)*, pages 1–8, Tel-Aviv, Israel, 2013.

[5] Donald E. Porter, Silas Boyd-Wickizer, Jon Howell, Reuben Olinsky, and Galen C. Hunt. Rethinking the library os from the top down. pages 291–304.

[6] Felix Schuster, Manuel Costa, Cedric Fournet, Christos Gkantsidis, Marcus Peinado, Gloria Mainar-Ruiz, and Mark Russinovich. VC3: Trustworthy Data Analytics in the Cloud using SGX. 2015.

[7] Intel software guard extensions (intel sgx) sdk support.