Chapter 1
Introduction

Atsuko Miyaji, Shinsaku Kiyomoto, Katsuya Tanaka, Yoshifumi Nishida, and Koji Kitamura

1.1 Purpose of Miyaji-CREST

Recently, big data analysis results are expected to be used in various situations such as medical or industrial fields for new medicine or product development. For this reason, it is important to establish a secure infrastructure of the collection, analysis, and use of big data. We need to consider mainly three entities for the infrastructure: data owner, analysis institutions, and users. This research pays attention to a balance between privacy and utilization and also realizes appropriate reduction and feedback of the data analysis results to the data owners.

To build a secure big data infrastructure that connects data owners, analysis institutions, and user institutions in a circle of trust, we construct security technologies necessary for big data utilization. Our main security technologies are oblivious RAM (ORAM), private set intersection (PSI), privacy-preserving classification,

A. Miyaji (✉)
Osaka University, 1-1 Yamadaoka, Suita, Osaka 565-0871, Japan
e-mail: miyaji@comm.eng.osaka-u.ac.jp

S. Kiyomoto
KDDI Research, Inc., 2-1-15 Ohara, Fujimino-shi, Saitama 356-8502, Japan
e-mail: kiyomoto@kddi-research.jp

K. Tanaka
National Cancer Center Japan, 5-1-1 Tsukiji, Chuo-ku, Tokyo 104-0045, Japan
e-mail: katstana@ncc.go.jp

Y. Nishida
Tokyo Institute of Technology, 2-12-1 Ookayama, Meguro-ku, Tokyo 152-8550, Japan
e-mail: nishida.y.af@m.titech.ac.jp

K. Kitamura
National Institute of Advanced Industrial Science and Technology,
2-4-7, Aomi, Koto, Tokyo 135-0064, Japan
e-mail: k.kitamura@aist.go.jp

A. Miyaji and T. Mimoto (eds.), *Security Infrastructure Technology for Integrated Utilization of Big Data*,
https://doi.org/10.1007/978-981-15-3654-0_1

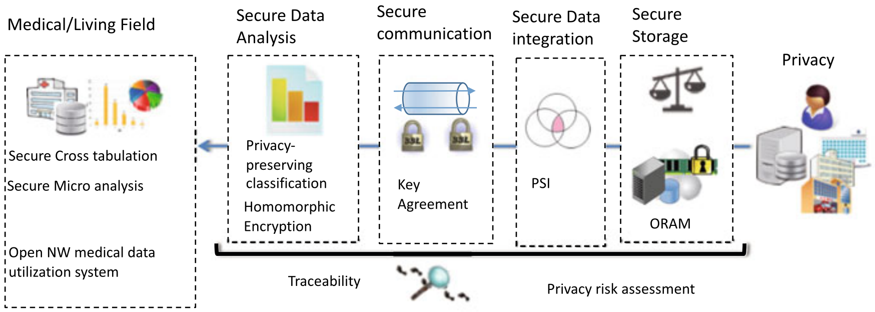

Fig. 1.1 Overview of security infrastructure from data collections to utilization

privacy configuration support, privacy risk assessment, and traceability. Furthermore, we consider the robustness against various attacks such as cyber attacks and post-quantum security.

We construct a safe and privacy-preserving big data distribution platform that realizes the collection, analysis, utilization, and return of owners of big data in a secure and fair manner.

In addition, we demonstrate our secure big data infrastructure in a medical and living safety field. Figure 1.1 shows an overview of our research.

1.2 Roles of Each Group

1.2.1 Security Core Group

We constructed security primitives in the following fields with the aim of realizing an infrastructure for big data utilization that conducts collection, analysis, and utilization of big data securely: 1. Analysis of security basis: Any security primitive which is used for an infrastructure of big data utilization is based on cryptology algorithms. That is, a security primitive becomes compromised if the underground cryptology algorithm is attacked. Therefore, security analysis on cryptographic primitives is important. In this research, we focus on elliptic curve cryptosystems, which achieve a compact public key cryptosystem, and learning with error (LWE)-based cryptosystems, which are types of post-quantum cryptosystems. 2. Privacy-preserving data integration among databases distributed in different organizations: This primitive integrates the same data among databases kept in different organizations while keeping any different data in an organization secret to other organizations. 3. A privacy-preserving classification: This primitive executes a procedure for the server's classification rule to the client's input database and outputs only a result to the client while keeping client's input database secret to the server and server's classification rule to the client.

1.2.2 Security Management Group

Our group focuses on research on data anonymization techniques. First, we analyze the existing anonymization techniques and adversary models for the techniques and clarify our research motivation. Then, we propose our adversary model applicable to several anonymization methods and propose a novel privacy risk analysis method. An implementation of our data anonymization tool based on the risk analysis method is introduced in the chapter.

1.2.3 Living Safety Testbed Group

The Living Safety Group deals with developing new technologies for injury prevention in daily environments such as school safety and home safety based on the security platform developed by the Security Core Group and the Security Management Group. This group has devoted itself to not only developing technology for handing the big data related to injury but also empowering practitioners through social implementation utilizing the developed technologies in cooperation with multiple stakeholders.

1.2.4 Health Testbed Group

Health Testbed Group is focused on implementing a secure clinical data collection and analysis infrastructure for clinical research using the cloud by applying the security primitives developed by the Security Core Group and Security Management Group. This group is working on standardization of data storage, cross-institutional collection, and analysis for electronic medical record data, management mechanism of patient consent information, and traceability for secondary use of medical data, for the development of our health testbed.