2016 11th International Conference on Knowledge, Information and Creativity Support Systems (KICSS), Yogyakarta, Indonesia

Privacy Modelling Of Sensitive Data in Universal Healthcare Coverage in Indonesia

Irvan Santoso
School of Computer Science
Bina Nusantara University
West Jakarta, Indonesia
Email: isantoso@binus.edu

Ida Sri Rejeki Siahaan
School of Computer Science
Bina Nusantara University
West Jakarta, Indonesia
Email: isiahaan@binus.edu

Suharjito
Master in Computer Science
Bina Nusantara University
West Jakarta, Indonesia
Email: suharjito@binus.edu

***Abstract*—In developing information system, privacy is an important component that must be guaranteed especially in government system. Indonesian government has introduced universal health care system that adopts Diagnosis-Related Group (DRG). This system is used to increase the service of healthcare providers and integrate patients' data. In Indonesia the universal health care system is named as Indonesia Case Base Group (INA-CBG). In this research, several analysis will be performed on the related regulations in Indonesia and the current system to prevent sensitive data to be disclosed to unauthorized parties. This work utilizes encryption, information flow, data integration, and partial order set approaches. The objective of this research is to ensure data privacy from unauthorized parties and to prevent fraud of integrated patients' data.**

***Index Terms*—Diagnosis Related Group; privacy; access control; BPJS Kesehatan**

I. Introduction

Privacy is defined in [16] as

> "The ability of an individual (or organization) to decide whether, when, and to whom personal (or organizational) information is released."

Privacy is something that must be protected. However, in developing a system sometimes privacy is not taken care properly. In Information System development, usually developers are only concerned about the functionality of a system, but forgetting their obligations to maintain users' privacy. Indonesian government has introduced a universal health care system named Badan Penyelenggara Jaminan Sosial (BPJS). BPJS was regulated in "UNDANG-UNDANG REPUBLIK INDONESIA NOMOR 24 TAHUN 2011 TENTANG BADAN PENYELENGGARA JAMINAN SOSIAL" [14]. The goal of BPJS is to create prosperity for Indonesian residence. In addition, this system adopts Diagnosis-Related Group (DRG). DRG is used to improve healthcare provider service by changing the payment system [2]. DRG has been adopted by developed countries since 1980s. DRG groups patients based on similar diagnoses and procedures. Furthermore, Indonesia begins to adopt DRG to increase the quality of service in the payment system. Indonesia has its own name, namely Indonesia Case Base Group (INA-CBG) [8]. Furthermore, patients who want to use BPJS while receiving healthcare provider services must give their information to healthcare providers and submit it into INA-CBG system. In addition, BPJS provides a system which will help healthcare provider that does not have management system, namely Sistem Informasi Manajemen Rumah Sakit (SIMRS) that regulated in "UNDANG-UNDANG REPUBLIK INDONESIA NOMOR 11 TAHUN 2008 TENTANG INFORMASI DAN TRANSAKSI ELEKTRONIK" [12] and can be seen in "PERATURAN MENTERI KESEHATAN REPUBLIK INDONESIA NOMOR 82 TAHUN 2013 TENTANG SISTEM MANAJEMEN RUMAH SAKIT" [11]. However, some information has sensitive data, namely data privacy and it cannot be retrieved directly without any policies. it was regulated in "UNDANG-UNDANG REPUBLIK INDONESIA NOMOR 29 TAHUN 2004 TENTANG PRAKTIK KEDOKTERAN", namely medical records of patient are owned by healthcare provider and the contents are the properties of patients themselves [15]. BPJS integrate all of patient data into one database and make it more difficult to be maintained. Because of that, there is a possibility of fraud occurred when patient data is integrated. For instance, patient data can be leaked to other parties, such as pharmacies. They know which area or city with diseases that often occur. Then they can increase the price of several medicines accordance to that information. This will be detrimental to people with the disease who want to buy medicine without insurance. Although healthcare providers are aware about that issue. Inevitably, healthcare providers must submit data of patient to claim their cost of given treatment.

Therefore, this research proposes a new model of access control based on the authority held by each role in healthcare provider. In previous research, Role-Based Access Control (RBAC) has been proposed by Sandhu et al., [17]. However, this research will use another method to make access control. This is performed to adjust the existing situation. The model will be described by using partial order set as a reference to make constraints. This approach has been proposed by Dawson et al., which is applied to make a model for healthcare providers [6]. Furthermore, constraints that made will be used to prevent a violation transaction in healthcare providers and BPJS. There are several analysis that performed to obtain the appropriate method in the data store, as follows: hashing, fragmentation, randomize, and encryption. Those methods have advantages and disadvantages. Hashing can ensure the

978-1-5090-5130-4/16/$31.00 ©2016 IEEE

data privacy in INA-CBG database by encrypting the data. Unfortunately, hashing is one way encryption. The data could not be retrieved or decrypted. This causes the data can not be audited. Furthermore, fragmentation is a method for storing data based on existing constraints. This method requires more than one database in data storage, because data is considered as sensitive should not be merged into one database based on constraints that made. However, this method can not be applied, because it must change the structure of INA-CBG database. Moreover, it will become rigid and requires a lot of database. Furthermore, randomize is a method for scrambling data so the data are not easy to know the relationship of the data. However, it will become harder to retrieve the data after randomize. There must be a pattern for randomize. If authorized user already knows the pattern, the data will be not secure. Afterward, this research will propose an encryption technique to ensure patient data from unauthorized people. Encryption is considered as the most suitable method to ensure the data. This method no needs to change the structure of database. Moreover, the data can be retrieved or decrypted after encryption for audit. Then, secure information flow is applied to ensure the information only can be read by authorized people.

II. Literature Review and Related Works

In conducting a research, related theories and previous researches are sources that can support research. Those references also required as evidence for conducted research. In this section, there are several references that used and will be separated into two sections, literature reviews and related works. In literature reviews, the references are: Diagnosis-Related Group (DRG); privacy theory; access control theory; Role-Based Access Control (RABC); partial order set; information flow; and encryption. In related works, the references are: encryption; Platform for Privacy Policy Preferences (P3P); eXtensible Access Control Markup Language (XACML); and Sistem Informasi Manajemen Rumah Sakit (SIMRS).

A. Literature Review

1) Diagnosis-Related Group (DRG): Healthcare providers use DRG to improve their services by changing the payment system from retrospective to prospective [2]. The payment with retrospective system is performed by paying the given services. For instance, a patient having dengue fever. That patient receives treatments, such as infusion and lab test. Therefore, the patient must pay the fees based on price of infusion and lab test. Furthermore, prospective system is different from retrospective system. Prospective system changes the payment based on general case. For instance, some patients having dengue fever and they got different treatments. However, the total of payment is same for dengue fever although the treatments are different. This practice also occurs for other diseases because DRG groups patients based on similar diagnoses and procedures.

Originally, DRG was adopted by United State in 1983. Afterward, DRG has been spread to fifteen countries in Europe in 1987. In the next year, eight commonwealth countries also adopted DRG to improve healthcare provider services. Furthermore, Indonesia began to adopt DRG in 2008 to increase the quality of service in the payment system. The name is adapted to Indonesia Diagnosis-Related Group (INA-DRG). INA-DRG is used code from a private company, namely 3M. Then in 2010, Indonesia adopts Indonesia Case Base Group (INA-CBG) that used code developed by UNU [8]. INA-CBG source is open for government and can be customized.

2) Privacy: Saltzer and Schroeder have conducted a research on protection of information in Computer Systems [16]. This research was conducted because the majority of current applications involve storing information by simultaneous users. The information of users will be integrated into database of application. Those applications, such as credit bureau data banks; law enforcement information systems; time-sharing service bureaus; online medical information systems; and government social service data processing systems need protection to their stored information. However, privacy has not been considered properly by application's owner. Authorized users may access information of other users. Some issues may arise and may even harm other people. Therefore, they developed a technique of information protection in modern computer system.

3) Access Control: Haddad et al., proposed access control for data integration in presence of data dependencies [10]. Data integration provides convenience to query different data by using a unique entry point. However, some privacy issues may arise from this case. When data are integrated, several data may be accessed by unauthorized user. Therefore, they propose a method to identify possible violation in transaction. The objective is to prevent retrieval of data by unauthorized user.

Ciriani et al., conducted a research on outsourcing data [4]. There exists a benefit for individuals or organization in outsourcing their data to external servers. Data management will be transferred including the responsibility. However, some data are confidential, because they are sensitive and cannot be released to others. Therefore, they propose a method that involves data owner in storing part of data, while giving the other part to external parties. The owner will take care of privacy data that should not be known by anyone.

4) Role-Based Access Control (RBAC): Sandhu et al., proposed a new model of access control, namely Role-Based Access Control (RBAC) [17]. This access control focuses on role authority and task. Every permission in an organization structure will be given to appropriate roles. Roles are made to simplify the management in an organization. In addition, each role has their own responsibilities and qualifications. One role can give permission to another role and can revoke the permission given. A role is used as a reference to make a rule, policy, or policy-set and its more stable, because an organization rarely change the structure.

5) Partial Order Set: A binary relation can be said as partial ordering relation if the characteristics of that relation are reflexive, antisymmetric, and transitive [9]. Furthermore,

it can be said as partial ordering relation if there are at least two items and one is larger or smaller. A Partial Order Set (Poset) can be described in diagram, namely Hasse Diagram. Hasse Diagram has a pattern of tiers to the top. The bottom part is the lowest level and the top part is the highest level in Hasse Diagram.

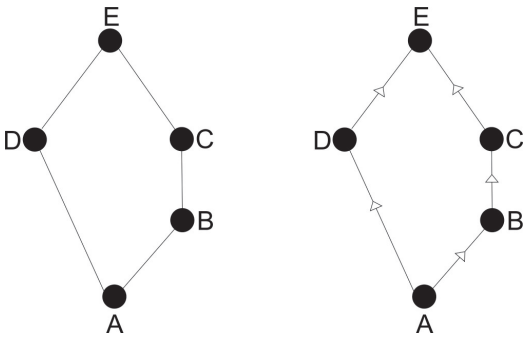

Fig. 1. Hasse Diagram

Fig. 1 shows the example of Hasse Diagram. A is the lowest level and E is the highest level in that diagram. Higher level may access the item below it, but lower level cannot access the item above it. For instance, node B may access the information of node A, but cannot access the information of node C or E. Afterward, node D may access the information of node A, but cannot access the information of node B. It is because there is no link between node B and node D. Then, node C may access the information of node B and A. Node E may access all of information.

Furthermore, Hasse Diagram has upper bounds and lower bounds for each node. For instance, node C and node E are upper bounds of node B as well as node E is upper bound of node D and node C. Moreover, a node can be said as Least Upper Bound (LUB) if it is upper bound for all nodes below it and there is no other upper bound. A node can be said as Greatest Lower Bound (GLB) if it is a lower bound of all nodes below it and there is no other lower bound. A lattice is a Poset (L,$\leq$) that has LUB and GLB.

6) Information Flow: Information flow is an important thing that must be considered in the development of system. A secure information flow must be established to ensure confidentiality [7]. The objective is to keep information from unauthorized users. For instance in government or military system, a secure information flow does not allow file or information with high level security to be transferred to user with the lower level. This is to prevent leakage without noticed. A model of information flow is defined as $FM = \langle N, P, SC, \oplus, \rightarrow \rangle$. Where $N = \{a, b, \ldots\}$ is a set of logical storage objects or information. Attributes in N may be variable of program, files, segments, or even user. Those attributes depend on the security level that are assigned. Then, $P = \{p, q, ...\}$ is a set of processes. Processes are described as active agents that have responsibility to check all of information flow. $SC = \{A, B, ...\}$ is a set of security classes. Security class is used to group objects or attributes into a specific class. Each class has their own characteristic that distinguishes one class with another class.

In information flow, class-combining operator "$\oplus$" is used to describing binary operator or function. This operator can be used to avoid semantic ambiguities that may arise when there are two functions over the same domain with overlapping ranges. In addition, a flow relation "$\rightarrow$" is needed to define the flow and relation between security class.

B. Related Works

1) Encryption: Encryption is a technique to ensure confidentiality, authentication, and integrity of communication [13]. Authentication means verifying the identity of networks that used. Therefore, it is necessary to develop a communication rule for decentralized authentication, namely protocol. It provides a standard in a communication to maintain the security. Moreover, secure communication in networks depends on the encryption of data. If the data can be retrieved without known by the others, it can be concluded that the data is secure and vice versa. Therefore, only the target that authorized can read the data by using key which is distributed along the sender and receiver [1].

[19] have proposed Dynamic Searchable Symmetric Encryption (DSSE) which will enable client to encrypt his/her data. Their model can manage' the information that stored and processes all of possible keywords that may appear in the document. In addition, client can search and update data while the data is encrypted. This scheme has been made by considering the efficiency and confidentiality of data.

[3] have proposed an asymmetric and symmetric encryption technique over insecure communications. They make a scheme that will secure the information between one or more communications in a network. In the network, clients will use asymmetric encryption to encrypt and decrypt the data. Then, symmetric encryption will be used if there is no other party that knows the communication. Symmetric encryption only uses' one key to encrypt and decrypt the request.

This research will use asymmetric encryption. In asymmetric encryption, there are two keys, namely public and private key. Public key is used to encrypt variables and private is key used to decrypt variables. Because all variables will be integrated into INA-CBG system, healthcare providers will use public key from INA-CBG and decrypt variables by using their private key.

2) Platform for Privacy Policy Preferences (P3P): P3P has been proposed by World Wide Web Consortium (W3C) as a machine-readable online statement [5]. It works by encoding the natural language privacy statements into XML-based format which can be interpreted easily by user agents. There are eight major components that support P3P, namely: entity, access, disputes, data, purpose, recipient, retention, and consequence. A P3P policy has two important elements, namely the root element (policy) and the body element (statement). A P3P policy is a very important element in P3P. It is used to cover all information in websites and check the legality of a request. Then, a P3P statement consists of a

purpose, data, recipients, retention, and consequence elements. Those elements in statement are used to indicate the data type in every site. Each site that used P3P policy can be automatically controlled by P3P user agents. They can retrieve, interpret, and transmit from server by using standard Hypertext Transfer Protocol/Secure (HTTP/HTTPS). P3P user agents provide direct links to page on website where users can choose either to enter or exit. In addition, user will be easier to understand by using P3P. This is because users can see directly the existing policy clearly.

3) eXtensible Access Control Markup Language (XACML): XACML is a XML-based language that created by the Organization for the Advancement of Structured Information Standards (OASIS) [18]. It describes both of policy language and access control decision request language. The policy language is used to define the requirements of access control. The request language is used to give a feedback of an action. The result can be permitted, denied, indeterminate, or not applicable. Indeterminate occurs when some part of requirements are missing or error. Because of that, the request cannot be performed. Not applicable occurs when the service cannot do the request.

There are three elements in XACML, namely rule, policy, and policy-set. Rule is the most elementary unit of policy. It exists only in one major actor of the XACML domain. There are five components in rule, as follows: a target, an effect, a condition, obligation expressions, and advice expression. Policy is a set of rules and it can be a component of a policy-set. Policy is used to connect rule in each major actors. To combine rules, policy needs an algorithm, namely rule-combining algorithm. There are five components of policy, as follows: a target, a rule-combining algorithm-identifier, a set of rules, obligation expression, and advice expression. Policy-set is a set of policies and can be a component of other policy-set. To combine policies, policy-set needs an algorithm, namely policy-combining algorithm. A policy-set contains five components, as follows: a target, a policy-combining algorithm-identifier, a set of policies, obligation expression, and advice expression.

Furthermore, to perform a request, Policy Enforcement Point (PEP) is needed. PEP is used to protect the resource when a requester performs a request. Afterward, PEP will send the request to Policy Decision Point (PDP) to check the request and gi've a feedback, either permitted, denied, indeterminate, or not applicable. Then, PDP will send the answer to PEP to give it to requester. XACML process is based on attributes. A policy resolves an attribute value by using two mechanisms, namely AttributeDesignator and AttributeSelector. Both mechanisms are used by XML language.

4) Sistem Informasi Manajemen Rumah Sakit (SIMRS): SIMRS is a system proposed by Indonesian government which will integrate all processes that occurred in healthcare provider information system [11] related to BPJS system. This system is a tool that supported healthcare provider in computational process. All of medical record about patients will be submitted into SIMRS and healthcare providers compulsory to use SIMRS. If healthcare providers do not submit patient's data, they cannot claim the cost that incurred. Furthermore, the objective of SIMRS is to control and give a standard of operational. SIMRS also provide easiness and efficiency in use such that time required in administration will be reduced and the accuracy will be increased.

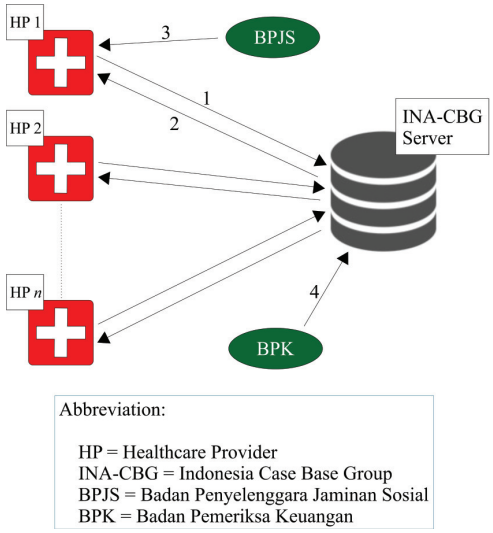

Fig. 2. General Scheme of Claim

III. Research Methodology

Ways of collecting data in this research will be described in data collection method. After data collection, an analysis will be created by making a research model and calculating the time required to finish this research. These processes will be described in research model.

A. Data Collection Method

In completing research model, there is data collection with up to BPJS and healthcare providers. In this section, the approached will be described in detail. First step begins from observation of current INA-CBG system and information that spread over internet. Those information will be used to understand the concept of current INA-CBG system. However, those information are not enough to make a new model. Because of that, a request will be asked to BPJS. A presentation will be delivered to BPJS about the objective of this research. The data that asked are process flow and the detail of current INA-CBG system from processing to data merging. Moreover, certain visit to healthcare providers will be performed to ask patient's data related to BPJS by performing interview. The objective is to know the variable that will be submitted into INA-CBG system and to analyze sensitive data.

B. Research Model

This research will propose a scheme to store patient datas into database by using asymmetric encryption. Furthermore, the general scheme will be described in Fig. 2

Healthcare provider will submit medical record variable into management system. Then, BPJS will verify the variable

before submitted into BPJS database. In addition, BPK only needs to audit the data from BPJS database. The detail of process will be described in Fig. 3.

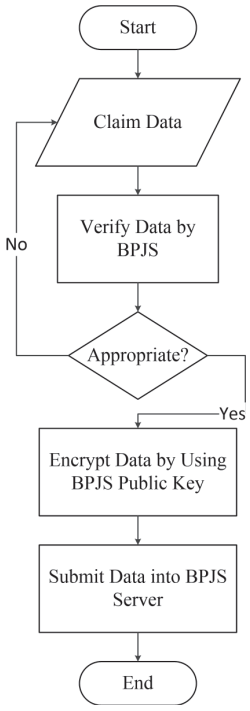

Fig. 3. Detail of Claim Process

If (3) BPJS already verify the data, (1) there is encryption process to ensure confidentiality of data privacy before submitted to BPJS database. The encryption is using public key of BPJS. This is to ensure the contents of medical records do not leak when delivered to INA-CBG database. Furthermore, (2) there is decryption process to obtain the data by using healthcare provider private key. Then, (4) BPK can obtain the data for audit by using BPJS private key. This is because the audit process will be performed in BPJS.

There are certain classification constraints c that will be applied to attributes A, such as: basic constraints, inference constraints, association constraints, and upper bound constraint. The constraints are mapping $\lambda : A \mapsto L$ that denotes to each attribute $A \in A$ and $L \in L$. Afterward, it is necessary to analyse which privacy level corresponds to the attribute based on constructed model.

Furthermore, the constructed model will be simulated by creating a program. Fig. 4 shows the detail of simulation phase. First, the constructed program will be tested to check the correctness of program. If the constructed program is not successfully compiled, the process will be back to create a program. If the constructed program is successfully compiled, then the result will be evaluated.

IV. Conclusion

In information system development, the obligation to maintain users' privacy sometimes is not taken care properly.

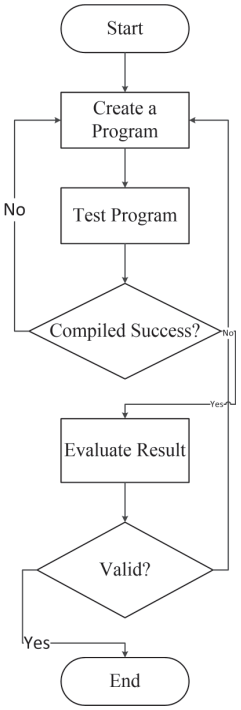

Fig. 4. Detail of Simulation Phase

This usually occurs because developers are only concerned about the functionality of the developed system. Indonesian government has introduced a system for universal healthcare, namely BPJS. BPJS can improve healthcare provider service by changing the payment system that used INA-CBG system. In this work a model had been proposed to ensure data privacy in INA-CBG system. The constructed model is created as security lattice. A simulation has also been implemented using C language which refers to constructed model. The simulation is performed to check the consistency of classification constraints and to produce complaint upper bounds.

References

[1] Dan Boneh and Matt Franklin. Identity-based encryption from the weil pairing. In *Advances in CryptologyCRYPTO 2001*, pages 213–229. Springer, 2001.

[2] Reinhard Busse, Alexander Geissler, Ain Aaviksoo, Francesc Cots, Unto Häkkinen, Conrad Kobel, Céu Mateus, Zeynep Or, Jacqueline OReilly, Lisbeth Serdén, et al. Diagnosis related groups in europe: moving towards transparency, efficiency, and quality in hospitals? *BMJ*, 346, 2013.

[3] Jay Alan Carlson. Method for secure communication using asymmetric & symmetric encryption over insecure communications, November 12 2015. US Patent 20,150,326,547.

[4] Valentina Ciriani, Sabrina De Capitani Di Vimercati, Sara Foresti, Sushil Jajodia, Stefano Paraboschi, and Pierangela Samarati. Keep a few: Outsourcing data while maintaining confidentiality. In *Computer Security–ESORICS 2009*, pages 440–455. Springer, 2009.

[5] Lorrie Faith Cranor. P3p: Making privacy policies more useful. *IEEE Security & Privacy*, (6):50–55, 2003.

[6] Steven Dawson, Sabrina De Capitani di Vimercati, Patrick Lincoln, and Pierangela Samarati. Maximizing sharing of protected information. *Journal of Computer and System Sciences*, 64(3):496–541, 2002.

[7] Dorothy E Denning. A lattice model of secure information flow. *Communications of the ACM*, 19(5):236–243, 1976.

[8] Heru Fahlevi. *The Innovation of the Role of Accounting in Public Hospitals-Lessons Learned from Hospital Financing Reforms in Indonesia and Germany*. PhD thesis, Zugl.: Speyer, Univ., Diss., 2014, 2014.
[9] George Grätzer. *General lattice theory*. Springer Science & Business Media, 2002.
[10] Mehdi Haddad, Jovan Stevovic, Annamaria Chiasera, Yannis Velegrakis, and Mohand-Saïd Hacid. Access control for data integration in presence of data dependencies. In *Database Systems for Advanced Applications*, pages 203–217. Springer, 2014.
[11] MENTERI KESEHATAN REPUBLIK INDONESIA. PERATURAN MENTERI KESEHATAN REPUBLIK INDONESIA NOMOR 82 TAHUN 2013 TENTANG SISTEM INFORMASI MANAJEMEN RUMAH SAKIT.
[12] PRESIDEN REPUBLIK INDONESIA. Undang-undang republik indonesia nomor 11 tahun 2008 tentang informasi dan transaksi elektronik.
[13] Roger M Needham and Michael D Schroeder. Using encryption for authentication in large networks of computers. *Communications of the ACM*, 21(12):993–999, 1978.
[14] Undang-Undang Republik Indonesia Nomor. Tahun 2011 tentang badan penyelenggara jaminan sosial. *Jakarta: Presiden Republik Indonesia*, 24.
[15] Undang-Undang Republik Indonesia Nomor. Tahun 2004 tentang praktik kedokteran.[homepage on internet], 29.
[16] Jerome H Saltzer and Michael D Schroeder. The protection of information in computer systems. *Proceedings of the IEEE*, 63(9):1278–1308, 1975.
[17] Ravi S Sandhu, Edward J Coyne, Hal L Feinstein, and Charles E Youman. Role-based access control models. *Computer*, (2):38–47, 1996.
[18] OASIS Standard. extensible access control markup language (xacml) version 3.0. 22 january 2013, 2013.
[19] Emil Stefanov, Charalampos Papamanthou, and Elaine Shi. Practical dynamic searchable encryption with small leakage. In *NDSS*, volume 14, pages 23–26, 2014.