ICIC Express Letters
Volume **14**, Number **9**, September **2020**

ICIC International ©2020 ISSN 1881-803X
pp. **837–843**

A SECURITY LATTICE MODEL AND SIMULATION OF DATA PRIVACY IN INA-CBG APPLICATION IN *BADAN PENYELENGGARA JAMINAN SOSIAL* (BPJS)

Irvan Santoso[1] and Ayu Hidayah Aslamiah[2]

[1]Cyber Security Program, Computer Science Department
[2]Computer Science Department
School of Computer Science
Bina Nusantara University
Jl. K. H. Syahdan No. 9, Palmerah, Jakarta 11480, Indonesia
{ isantoso; aaslamia }@binus.edu

Received December 2019; accepted March 2020

Abstract. *Data privacy is something that must be stored securely especially in a government system. The Indonesian government has introduced a universal coverage system that adopts Diagnosis Related Group (DRG), namely Indonesia Case Base Group (INA-CBG). INA-CBG is used to improve the service of healthcare providers and integrate user data into one database. Hence, there is a possibility of security gaps in protecting data from unauthorized users. Therefore, an analysis was performed based on the regulation published in Indonesia and the current INA-CBG system. A security lattice model of access control was produced to set constraints to prevent data privacy from being retrieved by unauthorized users. There are five security level categorizations as follows: Public, Research, Clinical, Financial, and Provider. Each security level was assigned to 26 attributes which are components of the current INA-CBG system. Furthermore, a simulation has been performed based on the model developed to ensure that the assigned level is correct. The results of the simulation will be used as a reference in determining the security lattice of each attribute if the result shows there are inappropriate levels.*
Keywords: Security level, Security lattice, INA-CBG, DRG, Universal coverage system

1. **Introduction.** The Internet revolution has changed the world communication infrastructure with a lot of data exchange including data privacy [1]. Data privacy is something that must be kept confidential [2]. To maintain the data, we need a system and the right knowledge to store and determine which data can be accessed by people who have the authority [3]. Inevitably, this becomes a difficulty for system developers and users who use the system. The more secure a system makes the more difficult a system is used by its users. Developers must consider the extent to which they must create a system that is secure in preserving their data and is easy to use by users [4].

One system that needs serious consideration in its development is a government system that stores public data [5]. The government system must be very secure from all existing attacks from parties who do not have the authority. The Indonesian government has a *Badan Penyelenggara Jaminan Sosial* (BPJS) which provides a universal coverage system under the name Indonesia Case Base Group (INA-CBG) [6] which is regulated in [7]. INA-CBG adopted the Diagnosis Related Group (DRG) which grouped patients based on diagnoses and similar procedures to improve health service providers by changing the payment system [8]. Patients who are members and want to use the service must input their data into the system. Even though sensitive data has been regulated in [9] to ensure that data cannot be retrieved directly without complying with certain policies, the INA-CBG system integrates patient data into one database which makes it hard to ensure that

DOI: 10.24507/icicel.14.09.837

there is no violation for privacy. There is a possibility of fraud such as aggregated patient illnesses that can be leaked to third parties such as pharmacies. If they know which area with illnesses that often occur, they can increase the medicine price in a particular area. This will be detrimental to people with illnesses without insurance. Besides, BPJS also provides a hospital information system named Sistem Informasi Manajemen Rumah Sakit (SIMRS) regulated in [10] which has a bridging component to INA-CBG system.

Based on the problem, this research proposes a model of privacy assurance based on access control and partial order set which regulates confidentiality in a universal coverage system. This model will be used as a reference to make constraints. 63 constraints have been used to prevent violation transaction consisting of basic constraints, inference constraints, association constraints, and upper bound constraints. The developed constraints are made based on attributes in the healthcare provider. Then, secure information flow is applied to ensuring information only can be read by authorized parties. Moreover, to prove the correctness of each constraint, algorithms and simulation using programming language will be performed to check whether the constraints have the right security level or vice versa.

2. **Partial Order Set.** A binary relation can be said as partial ordering relation if the characteristics of that relation are reflexive, antisymmetric, and transitive [11]. 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 the diagram, namely Hasse Diagram. Hasse Diagram has a pattern of tiers to the top.

Figure 1 shows an example of a Hasse Diagram. A is the lowest level and E is the highest level in the given diagram. A higher level may access the item below it, but the lower level cannot access the item above it. Furthermore, the Hasse Diagram has upper bounds and lower bounds for each node. A node is a Least Upper Bound (LUB) if it is upper bound for all nodes below it and there is no other upper bound. A node is a 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 $(L, \geq)$ that has LUB and GLB.

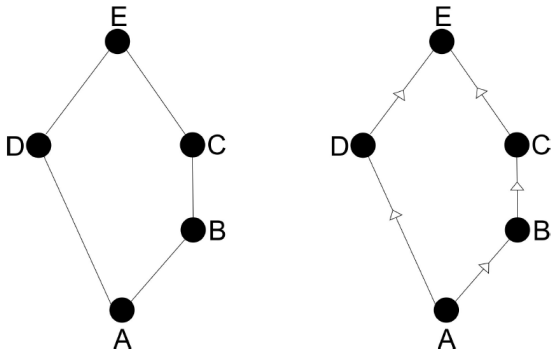

FIGURE 1. Hasse Diagram

3. **Basic Definitions.** Access class AC is made to prevent violations in access control. AC is assumed as a pair of the form a security level L and a set of attributes A. The relation between that pair is limited by a Poset, namely dominance relation $\geq$. The $\geq$ shows information which can be seen is information that has the same level as the subject or below. For instance, the expression $A \geq b$ means A dominates b. A may read the information on the level b or below. Besides, the partial order set $(AC, \geq)$ is assumed as a classification lattice LT. Therefore, LT which will be used is partial order set $(L, \geq)$ to specify A in LT may have dynamic security levels L. Furthermore, classification constraints C specify security levels on attributes based on confidentiality. A constraint

c consists of pair *lhs* and *rhs*, where $c \in C$. In *lhs* and *rhs* it can be a security level L or a mapping λ: $A \rightarrow L$, where $(L, \geq)$ is a classification lattice. Constraints will be classified as basic constraints, inference constraints, association constraints, and upper bound constraints.

Basic constraints C_{basic} are constraints that specify the security level on each attribute. In C_{basic}, AC has a mapping of attribute $\lambda(A)$ and a security level L. Furthermore, C_{basic} map an attribute to a specific level, $\lambda(A) \geq L$, when $a \in A$ and $l \in L$. It can be assumed that attribute A has a level L. Inference constraints $C_{inference}$ are made to prevent bypassing of C_{basic} through data inference. Afterward, $C_{inference}$ map attributes, $\lambda(a_x) \geq \lambda(a_y)$, when $a_x \in A$ and $a_y \in A$. It can be assumed that if the attribute in a_x is known, then the attribute in a_y can also be known. Association constraints $C_{association}$ are made to limit the information that may be known by a combination of attributes. It requires the least upper bound of the classifications to dominate the security level given by basic constraints. For instance, $\text{lub}\{\lambda(a_x), \lambda(a_y)\} \geq L$, where $a_x \in A$, $a_y \in A$, and $l \in L$. That constraint requires *lub* of the classifications between $\lambda(a_x)$ and $\lambda(a_y)$ to dominate a security level. Upper bound constraints C_{upper} are made to guarantee an attribute is always accessible to a specific level. These constraints prevent the classification of attributes being raised above certain levels. The mapping of C_{upper} is performed by setting a specific level and a specific attribute, $L \geq \lambda(A)$, where $a \in A$ and $l \in L$. In addition, upper bound constraints can indirectly affect other attributes to have the same security level. Base case constraints C_{base} will be created for induction method evaluation. In this constraint, all attributes will be assigned as the highest security level.

4. **Constructed Model.** A new model is shown in Figure 2 by looking at the information flow process. The constructed model has a security lattice LT which will describe the relation of each attribute A from the lowest level Public to the highest level, Provider, where $LT = (L, \geq)$. Due to this lattice, each attribute a will not have a static security level but may have a dynamic security level l.

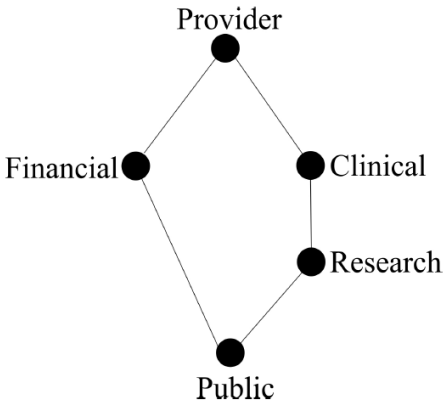

FIGURE 2. The lattice of constructed model

In the lowest level Public, some variables can be known by all of the people. Then, in the Research, some variables cannot be known by all of the people and should be given a higher level than Public. In addition, in Clinical, some variables can be revealed by combining the variables between the variable in Public and variable in Research. Then, at the highest level, Provider can read all of the variables. This level is usually owned by the government so that they can assess each patient.

5. **Evaluation Method.** Attributes that will be submitted to the system consist of 26 attributes. The model uses these attributes which must be assigned to a certain security

level. This assignment must comply with certain classification constraints. Therefore, this work solves the problem of lattice assignment formulated as follows.

Problem 5.1. (Lattice-Assignment) *Given A, LT* $= (L, \geq)$*,* C_{basic}*,* C_{base}*,* $C_{inference}$*,* $C_{association}$*, and* C_{upper} *that satisfy* λ*:* $A \geq L$*.*

To prove the correctness of our model, this work will use the induction method as an evaluation method.

Theorem 5.1. (Correctness) *The algorithms in [12] solve* **Problem 5.1 (Lattice-Assig-nment)**. *Given A, LT* $= (L, \geq)$*,* C_{basic}*,* C_{base}*,* $C_{inference}$*,* $C_{association}$*, and* C_{upper} *that satisfy C or terminate with Failure if C are inconsistent.*

Lemma 5.1. (Correctness) *It shows the correctness of algorithms by proving the consistent of the constructed model and the termination of algorithms.*

Proof: Let c be a set of classification constraints over a set of attributes and a classification lattice.

A) If the result of **computeUpperBound** is *Failure*, then the sequence of constraints is inconsistent.
B) If the result of **computeUpperBound** is *Success*, then the checked classification of mapping λ: $A \rightarrow L$ satisfies c ($\lambda \models c$), then the sequence of constraints is consistent.
C) **computeUpperBound** process always terminates.

To prove (A): for all mappings λ not dominate λ', there exists a constraint $c \in c$ such that λ' does not equal c. In addition, in the process of **computeUpperBound**, it is not possible to change the levels of the variable to higher.

To prove (B): mapping λ satisfies a sequence of constraints c by induction method. If mapping λ satisfies all immediate constraints on attributes in all constraints c, then $\lambda \geq \lambda'$, there exists a constraint $c \geq c$ such that $\lambda' \models c$.

To prove (C): **computeUpperBound** is arranged in a finite set based on the constraints. Each constraint will be called and checked by several functions. Each function will return a value in the process of checking, namely Success or Failure. If the result is Failure, then the process will be terminated. If the result is Success, then the process will be continued until the end of the program which will produce upper bounds.

6. Classified Constraints.

6.1. **Basic constraints.** Basic constraints assign security levels for each attribute. The purpose of basic constraints is to give basic security level and ascertain which attribute has a higher level in confidentiality. In our case, there are 26 basic constraints according to Table 1.

Main diagnosis and Secondary diagnosis are assigned higher than Public. This is because the diagnosis of a patient may not be known by anyone other than the patient and healthcare provider. Moreover, all attributes which are related to patient and healthcare provider are assigned as Research. This is because those attributes may not be known generally. Furthermore, SEP number and Medical record number are assigned higher than Research, namely Clinical. This is because if someone knows SEP number or Medical record number, they may know about patient information, such as who and which disease.

6.2. **Inference constraints.** Inference constraints are made to prevent bypassing basic constraints through data inference. In our case, there are 22 inference constraints according to Table 2.

Table 1. Basic constraints

C_1: λ(Healthcare provider code number) $\geq$ Public	C_{14}: λ(Procedure) $\geq$ Public
C_2: λ(Healthcare provider name) $\geq$ Public	C_{15}: λ(Length of stay) $\geq$ Public
C_3: λ(Healthcare provider class) $\geq$ Public	C_{16}: λ(Main diagnosis) $\geq$ Research
C_4: λ(Dates of discharge) $\geq$ Public	C_{17}: λ(Secondary diagnosis) $\geq$ Research
C_5: λ(Age (year)) $\geq$ Public	C_{18}: λ(Activity daily living) $\geq$ Research
C_6: λ(Age (day)) $\geq$ Public	C_{19}: λ(INA-CBG code) $\geq$ Research
C_7: λ(Dates of birth) $\geq$ Public	C_{20}: λ(Special CMG) $\geq$ Research
C_8: λ(Gender) $\geq$ Public	C_{21}: λ(INA-CBG description) $\geq$ Public
C_9: λ(Treatment class) $\geq$ Public	C_{22}: λ(SEP number) $\geq$ Clinical
C_{10}: λ(Dates of admission) $\geq$ Public	C_{23}: λ(Medical record number) $\geq$ Clinical
C_{11}: λ(Treatment type) $\geq$ Public	C_{24}: λ(Top-up rates) $\geq$ Financial
C_{12}: λ(Patient status) $\geq$ Public	C_{25}: λ(Healthcare provider rates) $\geq$ Financial
C_{13}: λ(Birth weight) $\geq$ Public	C_{26}: λ(Total fare) $\geq$ Financial

Table 2. Inference constraints

C_{27}: λ(Healthcare provider code number) $\geq$ λ(Healthcare provider name)	C_{38}: λ(Main diagnosis) $\geq$ λ(Procedure)
C_{28}: λ(Healthcare provider code number) $\geq$ λ(Healthcare provider class)	C_{39}: λ(Secondary diagnosis) $\geq$ λ(Procedure)
C_{29}: λ(Healthcare provider name) $\geq$ λ(He althcare provider code number)	C_{40}: λ(Special CMG) $\geq$ λ(Procedure)
C_{30}: λ(Healthcare provider name) $\geq$ λ(Healthcare provider class)	C_{41}: λ(INA-CBG code) $\geq$ λ(Procedure)
C_{31}: λ(Length of stay) $\geq$ λ(Treatment type)	C_{42}: λ(INA-CBG description) $\geq$ λ(Procedure)
C_{32}: λ(SEP number) $\geq$ λ(Gender)	C_{43}: λ(INA-CBG code) $\geq$ λ(Main diagnosis)
C_{33}:λ(SEP number) $\geq$ λ(Dates of birth)	C_{44}: λ(INA-CBG code) $\geq$ λ(Secondary diagnosis)
C_{34}: λ(SEP number) $\geq$ λ(Treatment class)	C_{45}: λ(INA-CBG description) $\geq$ λ(Main diagnosis)
C_{35}: λ(SEP number) $\geq$ λ(Treatment type)	C_{46}: λ(INA-CBG description) $\geq$ λ(Secondary diagnosis)
C_{36}: λ(SEP number) $\geq$ λ(Main diagnosis)	C_{47}: λ(INA-CBG description) $\geq$ λ(INA-CBG code)
C_{37}: λ(SEP number) $\geq$ λ(Secondary diagnosis)	C_{48}: λ(INA-CBG code) $\geq$ λ(INA-CBG description)

6.3. **Association constraints.** Association constraints are made to limit the information which may be exposed by a combination of attributes. It requires the least upper bound of the classifications to dominate the security level given by basic constraints. There are 10 association constraints according to Table 3.

6.4. **Upper bound constraints.** Upper bound constraints are made to ensure the attribute is always accessible to a specific level. These constraints prevent the classification of attributes from being increased. In our case, there are five upper bound constraints according to Table 4.

Table 3. Association constraints

C_{49}: lub{λ(Healthcare provider code number), λ(Main diagnosis)} $\geq$ Clinical	C_{54}: lub{λ(Healthcare provider name), λ(Main diagnosis)} $\geq$ Clinical
C_{50}: lub{λ(Healthcare provider code number), λ(Secondary diagnosis)} $\geq$ Clinical	C_{55}: lub{λ(Healthcare provider name), λ(Secondary diagnosis)} $\geq$ Clinical
C_{51}: lub{λ(Healthcare provider code number), λ(INA-CBG code)} $\geq$ Clinical	C_{56}: lub{λ(Healthcare provider name), λ(INA-CBG code)} $\geq$ Clinical
C_{52}: lub{λ(Healthcare provider code number), λ(INA-CBG description)} $\geq$ Clinical	C_{57}: lub{λ(Healthcare provider name), λ(INA-CBG description)} $\geq$ Clinical
C_{53}: lub{λ(Healthcare provider code number), λ(Special CMG)} $\geq$ Clinical	C_{58}: lub{λ(Healthcare provider name), λ(Special CMG)} $\geq$ Clinical

Table 4. Upper bound constraints

C_{59}: Clinical $\geq$ λ(Main diagnosis)	C_{62}: Clinical $\geq$ λ(INA-CBG description)
C_{60}: Clinical $\geq$ λ(Secondary diagnosis)	C_{63}: Clinical $\geq$ λ(Special CMG)
C_{61}: Clinical $\geq$ λ(INA-CBG code)	

7. **Result and Discussion.** A simulation will be performed for the proposed model by running the program which implemented the model. The results obtained are λ(Main diagnosis) = Clinical; λ(Secondary diagnosis) = Clinical; λ(INA-CBG code) = Clinical; λ(INA-CBG description) = Clinical; λ(Special CMG) = Clinical; λ(SEP number) = Clinical; λ(Healthcare provider name) = Provider; λ(Healthcare provider code number) = Provider. Furthermore, those results will be used to evaluate the consistency of results. The objective is to ensure the constructed model already provides consistent constraints.

Moreover, the induction method is applied to proving the correctness of the proposed model. As a base case, each attribute will be assigned as the highest level. Then it should be checked using the induction step. When **computeUpperBound** computes all attributes in A have already complied, then the constructed model is correct. If nothing changes, then the model is consistent. Because Provider dominates all attributes, constraints C are consistent.

In the base case, Main diagnosis, Secondary diagnosis, INA-CBG code, INA-CBG description, Special CMG are assigned as Provider. However, the level must comply with the C_{upper}. Hence, Main diagnosis, Secondary diagnosis, INA-CBG code, INA-CBG description, Special CMG will be assigned as Clinical. In addition, all related variables will be evaluated.

From the result of **computeUpperBound**, several analyses will be performed to check whether constraints are consistent. First, because these attributes dominate Procedure, then Procedure should be lowered to Clinical. Since Clinical $\geq$ Public, the statement is still true. Second, the analysis will be performed for λ(Main diagnosis) $\geq$ Research and the others. Since in C_{upper}, Main Diagnosis and the others are assigned as Clinical, then the result is Clinical $\geq$ Research. It is still true because Clinical dominates Research. Third, the analysis will be performed for λ(SEP number) $\geq$ λ(Main diagnosis) and λ(SEP number) $\geq$ λ(Secondary diagnosis). Since SEP number is assigned as Clinical still dominates Main diagnosis and Secondary diagnosis that assigned as Clinical means it is still true. Fourth, the association of the following attributes will be checked such as lub{λ(Healthcare Provider code number), λ(Main diagnosis)} $\geq$ Clinical; lub{λ(Healthcare provider name), λ(INA-CBG code)} $\geq$ Clinical. Since Healthcare provider name and Healthcare provider code number are assigned as Provider, then Main

diagnosis and INA-CBG code are assigned as Clinical, the statement is true. This is because of lub{Provider, Clinical) $\geq$ Clinical. Therefore, the constraints C of the proposed model are consistent because no modification is needed.

8. **Conclusion.** In the information system, the obligation to maintain users' privacy sometimes is not taken care of properly. This is because developers are only concerned about functionality. The Indonesian government has introduced BPJS system for universal coverage. BPJS can improve healthcare provider service by changing the payment system that adopts INA-CBG. In this work, a model had been proposed to ensure data privacy. The constructed model is created as a security lattice. Simulation has also been implemented using a programming language to check the consistency of classification constraints and to produce complaint upper bounds. The results of this simulation are correct and consistent because there are no changes in determining the security level.

For future development, a dynamic security lattice model will be created by using dynamic taint analysis to adapt to the condition that can be changed by government regulations, every healthcare provider's regulations, or the customer needs. Therefore, it will affect each healthcare provider to have their customized access control in competing and improving their services.

REFERENCES

[1] T.-T. Teoh, Y.-Y. Nguwi, Y. Elovici, W.-L. Ng and S.-Y. Thiang, Analyst intuition inspired neural network based cyber security anomaly detection, *International Journal of Innovative Computing, Information and Control*, vol.14, no.1, pp.379-386, 2018.

[2] D. Zhang, Big data security and privacy protection, *Proc. of the 8th International Conference on Management and Computer Science (ICMCS 2018)*, 2018.

[3] M. Jayabalan and T. O'daniel, Access control and privilege management in electronic health record: A systematic literature review, *Journal of Medical Systems*, vol.40, no.12, 2016.

[4] M. Green and M. Smith, Developers are not the enemy!: The need for usable security apis, *IEEE Security & Privacy*, vol.14, no.5, pp.40-46, 2016.

[5] J. Archenaa and E. M. Anita, A survey of big data analytics in healthcare and government, *Procedia Computer Science*, vol.50, pp.408-413, 2015.

[6] D. G. Tamtomo and B. Murti, An analysis of the difference between INA-CBG reimbursementand medical cost for patients with chronic renal disease: An evidence from Kasih Ibu Hospital, Surakarta, *Revitalizing Family Planning Program and Women's Empowerment for the Improvement of Population Well-being and Economic Development*, 2018.

[7] *Law No. 24 of 2011 Concerning Social Security Organizing Agency (Badan Penyelenggara Jaminan Sosial or "BPJS")*, https://bpjs-kesehatan.go.id/bpjs/arsip/categories/Ng/undang-undang, Accessed on 01-Dec-2019.

[8] *Regulation of the Health Minister of the Republic of Indonesia No. 27 of 2014*, http://www.jkn.kemkes.go.id/attachment/unduhan/PMK No. 27 ttg Juknis Sistem INA CBGs.pdf, Accessed on 01-Dec-2019.

[9] *Law of the Republic of Indonesia No. 11 of 2008*, https://www.kpk.go.id/images/pdf/uu pip/UU_ITE no 11 Th 2008.pdf, Accessed on 01-Dec-2019.

[10] *Regulation of the Health Minister of the Republic of Indonesia No. 82 of 2013*, https://www.kemhan.go.id/itjen/wp-content/uploads/2017/03/bn87-2014.pdf, Accessed on 01-Dec-2019.

[11] S. Bistarelli and F. Santini, A Hasse diagram for weighted sceptical semantics with a unique-status grounded semantics, in *Logic Programming and Nonmonotonic Reasoning. LPNMR 2017. Lecture Notes in Computer Science*, M. Balduccini and T. Janhunen (eds.), Cham, Springer, 2017.

[12] *ConstraintAlgorithms*, https://github.com/raphael4/contraintAlgorithms, Accessed on 01-Dec-2019.