Research on the Tracking Algorithm of Program Level Fine-grained Data based on Cloud Virtual Environment

Zhigang ZHANG[1,2, a], Wei ZHANG[1, b], Juncheng GENG[1, c], Hongyi ZHANG[1,2, d]

[1]Haepc Electrlc Power Research Institute Zhengzhou 450000, China

[2]Henan Epri Gaoke Group Co. Ltd Zhengzhou 450000, China

[a]email: rh_liang@126.com, [b]email: 957007380@qq.com, [c]email: 278870280@qq.com,

[d]email: elwin8670@126.com

Keywords: DIFC model; fine-grained information flows; cloud computing

Abstract. The virtual machine in the fine-grained information flow tracking is the basis for realization of transparent cloud platform program level control. The information flow control access to sensitive information in the process, because the authority transfer security level and cannot read or write the non sensitive data, the coarse granularity information flow control is difficult to meet the actual demand of diversification, this paper proposes extended DIFC (Distributed Information Flow Control) model, this model avoids component of cloud platform virtual machine because of the higher level of security sensitive data through reading, it sends or modifies the defects of non sensitive data by transfering the authority, and effectively overcomes the defect that the existing information flow control method for the coarse granularity, and the shortcomings which unable to meet the actual demand, this model guarantees the tracking and control of fine-grained information flow within the virtual machine application, and it does not affect the original cloud service operation.

Introduction

Aiming at the problem of data security in cloud platform, the extensive research have been carried out at home and abroad. Most of the existing research is mainly using encryption and authentication algorithm for static data storage protection. For the read and write, calculate and process data in the cloud platform, the data file is read as a privacy of users in the process, it is difficult to use cryptographic techniques for data protection, because the operation is a non encrypted reside in memory[1]. Information flow control is a kind of access control technology in the cloud computing environment, the introduction of information flow control method, separately from the three levels of program level, system level and network level of data security control, all control units with each other, the implementation of the cloud platform data flow of multi granularity whole process tracking and control. This method effective implementation of fine-grained data, it prevents malicious behavior, uses the illegal data, avoids information leakage, ensures the data confidentiality and integrity.

Process control monitors virtual machine system API interface behavior and montiors fine-grained information flow control, once the sensitive data operation was found, the process control corresponds control strategies by marking the implementation, So as to realizes the program level information flow control, and program control. System management level control monitors virtual machine manager different virtual machine interface behavior, the flow of information between the control of virtual machine[2], and virtual machine escape detection of malicious behavior, the system management controls the flow of information between virtual machines by marking the manipulated data file, and real-time monitoring of the process, according to the monitoring results selection strategy level, and informs the strategy system for real-time status updates, finaly it generates log of system level control. Network management monitors network interface behavior, controls flows from different security domains of virtual or physical isolation gateway between the data flow of network protocol, through the analysis of data packets, according

to the corresponding packet marking binding strategy, management control behavior by the network information flow control device executes corresponding, at the same time, generating a network level control log.

Analysis of the Traditional Access Control Model

Access control is a difficult problem in cloud security, cloud computing system is open, dynamic and heterogeneous, the protection of data should be taken into account when the different participant, security strategy and the use of model. In the SaaS cloud application platform, access control model is the most commonly used is RBAC (Role Based Access Control) model. In order to solve the defects of the traditional access control model applied in the open adn dynamic environment application, researchers have extended the RBAC model adapt to the cloud services, including IRBAC2000[3], A-IRBAC2000[4], DRBAC[5], X-RBAC[6], ABAC [7] [8], TBAC[9] [10], TRBAC[11], BTRBAC[12] and UCON [13] [14]. These models implement the access control based the role. However, for the cloud platform, these models are mainly reflected in the management of authorization of users, and lack of effective access control of the user oriented data.

Different from the traditional access control model, some scholars try to solve the problem of access control from the angle of information flow. Denning puts forward a kind of information flow control model in IFC[15] [16], it definites control structure of information flow by using the lattice model, the flow of information is abstracted as the partial ordering relation between safety level. IFC model based tracking system data in the spreading process, it allows untrusted code to access the confidential data, and it will prevent the confidential data spread to unauthorized subjects. The BLP model[17] is a security model based on lattice earlier, in order to improve the availability, BLP model introduces the concept of trusted entities, all the behavior of trusted entities are considered to be credible, This will lead to the problem of excessive permissions. Bell proposed the model of the main network security level baded on the BLP[18], which makes the trusted subject beyond the mandatory access control policy is limited in a range, but in these methods, trusted subject authority has nothing to do with the state of the system in its life cycle, actually has not changed, it does not meet the principle of least privilege. According to the IFC model, the centralized authorization lacks of autonomy and flexibility, Myers and others proposed a distributed information flow control model DIFC in 2000[19]. DIFC is variety of traditional IFC, which more suitable for distributed computing environment for data distributed, autonomous, flexible access control. The DIFC model can effectively achieve the isolation of different security domains, but its data protection granularity is too coarse, when a subject to read sensitive data security level increases, it will not be able to read and write non sensitive data, therefore, it can not fully meet the needs of data fine-grained control in cloud platform.

The traditional access control mechanism can not effectively solve the problem of data security in cloud computing, the information flow control model is not suitable for cloud computing fine-grained tracking data, we need improve the information on the current control model to meet the needs of cloud computing, data security control and accountability requirements.

The Improved Extended DIFC Model

The virtual machine operating system of cloud platform provides information flow tracking and control is always the emphasis and difficulty in the research. Aiming at the difficult problem of virtual machine about fine-grained data access control and data sharing, this paper intends to study the extended DIFC model, which achieves fine-grained information flow tracking using marker. The fine-grained information flow model is established, the safety demonstrate of the model and the implement of transparent tracking technology. We study the program level fine-grained data tracking of specific technical route tracking based on three aspects.

The extended DIFC model has outstanding advantages in improving the availability of information flow model, it can capture more secure information flow characteristics of fine in non-credible system. The model introduces the lower density distributed technology, it allows lower

density of its own data, and does not require a globally trusted subject to complete this operation. The taint tracking and communication, which record levels of information flow that it sees current by marking the implicit changes marked or show change a subject, it embodies the principle of least privilege. The model introduce the right communication technology, it allows adjust their ability to transfer marker of other subjects, which makes information flow more flexible control.

However, with the reduction strategy, verification of dense distributed authorization transfer and information flow tracking characteristics of the DIFC model is a NP complete problem. In this paper, we adding some constraints on the extended DIFC model, in order to achieve the fine-grained information flow tracking in cloud platform of virtual machine. In order to carry out the strategy to test this model, we will use the MVSPA to describe the DIFC model formal. MVSPA can effectively describe the non-deterministic of concurrent systems, communication, recursive, abstract, shunt and deadlock behavior, its semantic syntax is as follows.

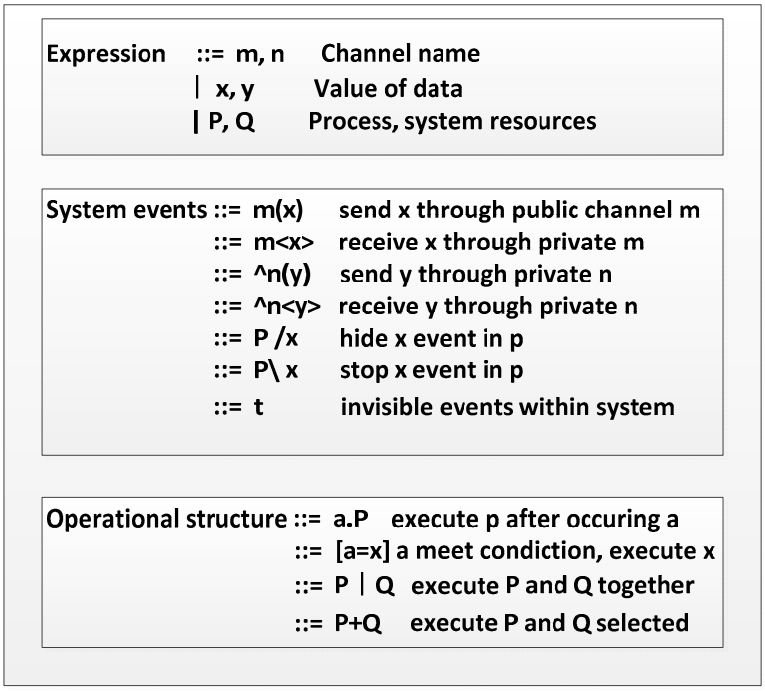

Fig.1. Semantic syntax description of MVSPA

The extended DIFC model includes a number of components, we will introduced briefly the function of each component, and describe the definition and semantics of each part of the model by the MVSPA.

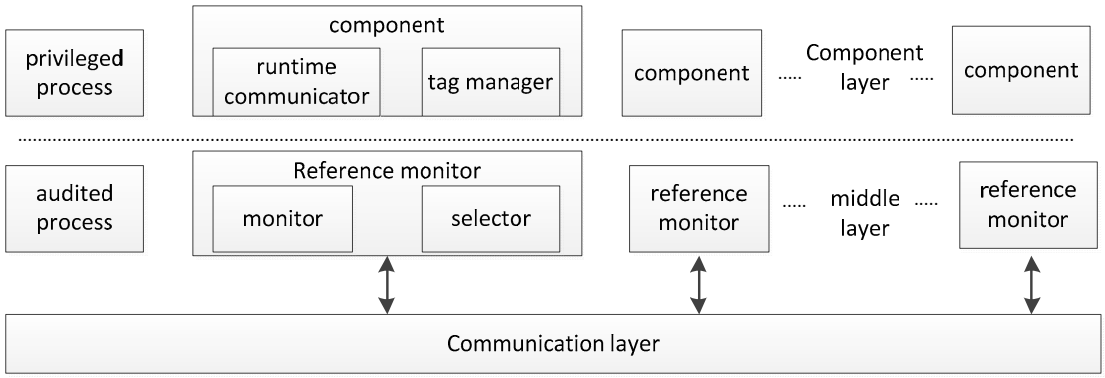

Fig.2. The composition and communication structure of extended DIFC

As shown above, in the extended DIFC model, the virtual machine system PEER componented by the runtime communication RTP and self labeling management device SLM. Reference monitor RM consists of monitor MON and selected device JUD, The information flow is decided by secure information flow rules and strategy. Privileged process *CONFIG* responsible for the ability of all the components under system initialization. Information flow logs of all the success and failure are record by the audit process *AUDIT*. The message are transfered through message channel in each

internal subsystem, and allowed multiple input and output message at the same time, the DIFC can effectively describe the information flow in the system, it reflects dynamic changes of all the system through the process of virtual mobile communication, at the same time, it can realize the dynamic configuration system strategy through the replacement name mechanism. The definition of model are described below.

Definition 1 color (Color): $C=\{color_1,color_2,\cdots,color_n\}$ a collection of basic color categories of system. When the confidentiality protection as the goal in the virtual system, a color as a kind of secret information, the color has infected represent that the secret information was obtain, the information flow legal when the color with the infection ability. When the integrity protection as the goal in the virtual system, a color can be represented as a type of malicious information sources, prohibit information flow does not have the infection ability of color components.

Definition 2 Tain: The core idea of the information flow control is the security label attached to the data, and with the data communication in the system, at the same time limit the transmission of tainted data by strategy.

$$A(\hat{}n,c_1\_a,\cdots,c_i\_a,query,config)\ \underline{\underline{\text{def}}}$$

$$n(x).\begin{Bmatrix}[x=query]\overline{n}<c_1\_a,\cdots,c_i\_a>.A\\ [x=config]n(new\_c_1\_a,\cdots,new\_c_i\_a).A(\hat{}n,new\_c_1\_a,\cdots,new\_c_i\_a,query,config)\end{Bmatrix} \quad (1)$$

Definition 3 Ability: $A=C\times(+,-)$ Defines a collection for all ability of components. $color_i^+$ represents that the components can be contaminated by $color_i$, $color_{i-}$ represents that the $color_i$ has decontamination ability. Component ability mapping function $f_b(P)$ represents tha A can obtain the ability of P components. The formal description of ability define as follows.

$$A(\hat{}n,c_1\_a,\cdots,c_i\_a,query,config)\ \underline{\underline{\text{def}}}$$

$$n(x).\begin{Bmatrix}[x=query]\overline{n}<c_1\_a,\cdots,c_i\_a>.A\\ [x=config]n(new\_c_1\_a,\cdots,new\_c_i\_a).A(\hat{}n,new\_c_1\_a,\cdots,new\_c_i\_a,query,config)\end{Bmatrix} \quad (2)$$

Definition 4 Data: $D=\{m_1,m_2,\cdots,m_n\}$ as a collection of all the data in virtual system components. Wherein m_i represents a data element, m_i identified by $taint_i$. Different data $taint_i$ can be labeled with different data m_i. Data tain mapping function $f_t(m_i)$ represents that tain $taint_i$ can be obtain from data m_i. The model supports the taint marking implicit adjustments to ensure normal taint tracking, data after pollution can develop into new pollution data. The formal description as follows.

$$D(\hat{}n,c_1,\cdots,c_n,query,update)\ \underline{\underline{\text{def}}}$$

$$n(x).\begin{Bmatrix}[x=query]\overline{n}<c_1,\cdots,c_n>.D\\ [x=update]n(new\_c_1,\cdots,new\_c_n).D(\hat{}n,new\_c_1,\cdots,new\_c_n,query,update)\end{Bmatrix} \quad (3)$$

Definition 5 Componet PEER: $P\in A\times D$ considers all applications, process and system resources in virtual system as the component, the data taint of components use distributed label management method. A session of system includes session initiator $SPON$ and response $RESP$. $SPON$ formal description as follows.

$$SPON(s\_rm,s\_n_i,s\_d_i,r\_n_j,r\_d_j,req,acc,ref,suc,r,s)\ \underline{\underline{\text{def}}} \quad (4)$$

$$\begin{Bmatrix}\overline{s\_rm}<req,s\_n_i,s\_d_i,r\_n_j,r\_d_j,r>.s\_rm(x).\begin{pmatrix}[x=ref]SPON\\ [x=acc]s\_rm(new\_c).new\_c(y).[y=r\_d_j]\overline{s\_rm}<suc>.SPON\end{pmatrix}\\ \overline{s\_rm}<req,s\_n_i,s\_d_i,r\_n_j,r\_d_j,s>.s\_rm(x).\begin{pmatrix}[x=ref]SPON\\ [x=acc]s\_rm(new\_c).new\_c(y).new\_c(z).[z=suc]\overline{s\_rm}<suc>.SPON\end{pmatrix}\end{Bmatrix}$$

Rule1:Secure information flow $Peer\_A\rightarrow Peer\_B \leftrightarrow \forall color_i\in f_t(m_a),color_i^+\in f_b(Peer\_B)$, it represents that message m_a can flow from the a component to $Peer\_B$ each other, any color of data

taint m_a in component $Peer\_A$, the component $Peer\_B$ must have the infection ability of color. In order to avoid it too strict, rigid control and achieve fine-grained control, the Rule 1 uses the ability and data pollution two factors rather than a single factor to control the flow of information.

Rule 2 Trusted decontamination $f_t(m_a)-\text{color}_i \leftrightarrow \text{color}_{i\_} \in f_b(Peer\_A)$, it represents that component $Peer\_A$ can delete a color from data taint m_a and the component $Peer\_A$ must have the ability of Decontamination. The formal description of trusted process C as follows.

$$C(\hat{}clear, taint, clean, query, update, c\_d, c\_nd) \underline{\underline{\text{def}}}$$

$$clear(x,y).\bar{y} < query > .y(c_1\_a,\cdots,c_i\_a).\ \bar{x} < query > .x(old\_c_1,\cdots,old\_c_i).\bar{x} < update >$$

$$\left\{ \begin{array}{l} [c_1\_a = c\_d]\bar{x} < clean > .\cdots.\left(\begin{array}{l} [c_i\_a = c\_d]\bar{x} < clean > .C \\ [c_i\_a = c\_nd]\bar{x} < old\_c_i > .C \end{array} \right) \\ [c_1\_a = c\_nd]\bar{x} < old\_c_1 > .\cdots.\left(\begin{array}{l} [c_i\_a = c\_d]\bar{x} < clean > .C \\ [c_i\_a = c\_nd]\bar{x} < old\_c_i > .C \end{array} \right) \end{array} \right\} \quad (5)$$

Rule 3: Taint Communication $(Peer\_A|m_a \to Peer\_B|M_b) \to f_t(m_b) = f_t(m_b) \cup f_t(m_a)$, it represents when the data m_a flow form component $Peer\_A$ to m_b of component $Peer\_B$, the taint update of data m_b in component $Peer\_B$ is a union collection of data m_a and data m_b; $(Peer\_A|(m_1 \to m_2)) \to f_t(m_2) = f_t(m_1) \cup f_t(m_2)$ represents when the taint flow from data m_1 to data m_2 in component $Peer\_A$, the taint of data m_2 is a union collection of data m_1 and data m_2.

The formal description of the system audit and policy configuration as follows.

$$AUDIT(\hat{}a\_rm, do) \underline{\underline{\text{def}}} a\_rm(s\_nr_i, s\_d_i, r\_n_j, r\_d_j, mode, result).do.AUDIT \quad (6)$$

$$CONFIG(\hat{}n, config, c_1\_a, \cdots, c_i\_a) \underline{\underline{\text{def}}} \bar{n} < config, c_1\_a, \cdots, c_i\_a > .CONFIG \quad (7)$$

In summary, the cloud virtual machine running environment can be expressed as

$$RTP \underline{\underline{\text{def}}} EXE\ |\ (DALVIK\ +\ INVOKE)\ . \quad (8)$$

The component can be expressed as $PEER \underline{\underline{\text{def}}} (SPON + RESP)|RTP|A|D$ (9)

Reference monitor can be expressed as $RM \underline{\underline{\text{def}}} JUD|MON$ (10)

The whole system can be expressed as $SYS \underline{\underline{\text{def}}} PEER|RM|CONFIG|AUDIT$ (11)

Test results

The extended DIFC model is mainly used for the security proof theoretically verify the proposed security properties of model. In the industry has been on the interference free nature of the IFC model on the basis of the research work, this paper intends to use the MVSPA framework to achieve the above objectives.

Firstly, the extended DIFC model describes the semantics by No interference model MVSPA. The equivalent definitions of non interference model as follows, For arbitrary $P \in \varepsilon$, the collection defined as $T(P) = \{\gamma \in \zeta^* | \exists P' : P \underline{\quad\gamma\quad} P'\}$, among them, $\gamma = \alpha_1 .. \alpha_n \in \zeta^*$, P and Q equivalent as $P \approx_T^W Q \Leftrightarrow T(P) = T(Q)$. Therefore, non deterministic system (NNI, Non-deterministic Non Interference) without interference, $E \in NNI \Leftrightarrow (E \setminus_I Act_H) / Act_H \approx_T^W E \setminus Act_H$.

Secondly, this paper uses the automatic verification tools CoPS which developed by Enrico Pivato, the CoPS tool verifies the security properties of model. Because the tool using the SPA

language, so the paper will use the existing methods for converting the MVSPA expression of the model is converted into the SPA description.

To achieve the transparent tracking fine-grained information flow within the virtual machine, we need guarantee technology. This paper aims to build a variety of constraints, under the constraint conditions, we need use the key technology which by means of transparent to the application, and track of explicit or implicit information flow constraints between variables. The specific route as shown below.

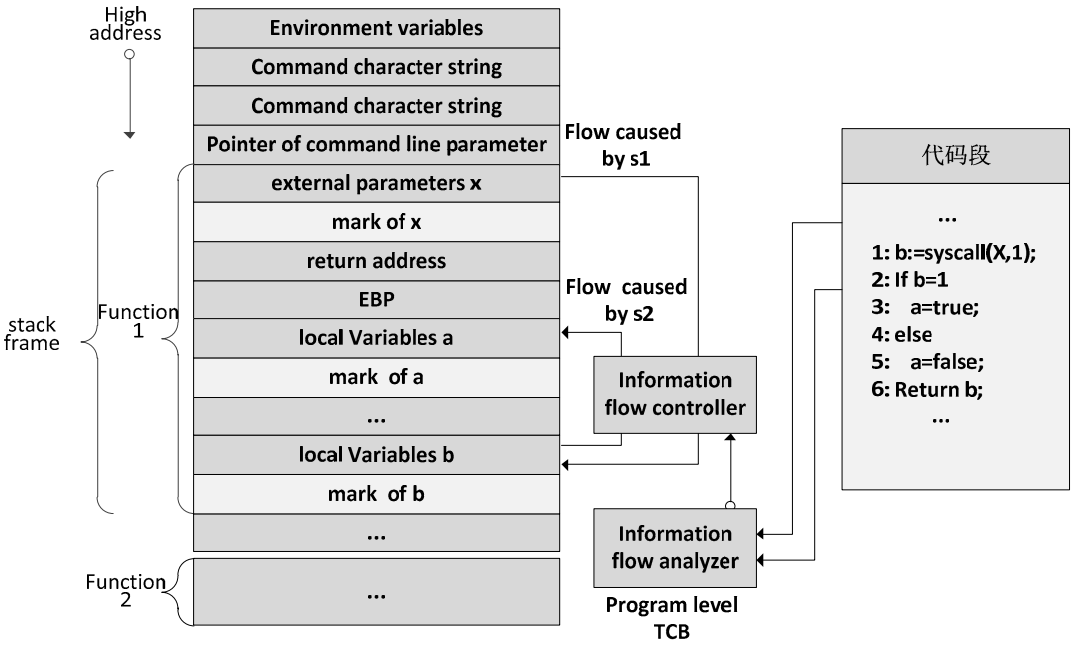

Fig.3. The experimental results

Conclusion

This paper expounds the importance of the cloud data security control, and the model of traditional access control problems, and finally puts forward the extended DIFC model, the model marks to achieve fine-grained information flow tracking, establishes the tracking model of fine-grained information flow, the model of security proof and the implemented of transparent tracking thchnology, this paper achieve the program level information flow control by tracking the fine-grained data flow of virtual machine, and doesn't affect the application of cloud.

Acknowledgement

In this paper, the research is proposed by Haepc Electrlc Power Research Institute, the name of the project network information security of embedded equipment detection method and detection rule library of technical service (project number: SC-FW20140301).

References

[1] Jianxin Li, Bo Li, Tianyu Wo, Chunming Hu, etc. CyberGuarder: A Virtualization Security Assurance Architecture for Green Cloud Computing [J]. In: Future Generation Computer Systems, 28: 379-390, 2012.

[2] Yangchun Fu, Zhiqiang Lin. Space Traveling across VM: Automatically Bridging the Semantic Gap in Virtual Machine Introspection via Online Kernel Data Redirection [J]. Security and Privacy, 33: 586-600, 2012.

[3] Kapadia A, Muhtadi J A, CampbellR, et al. IRBAC2000: Secure Interoperability Using Dynamic Role Translation[R]. Technical Report UIUCDCS-R-2000-2162, University of Illinois.

Urbana, IL, USA. 2000.

[4] Muhtadi J A, Kapadia A, Campbell R, et al. The A-IRBAC2000 Model: Administrative Interoperability Role-based Access Control[R/OL], 2000.

[5] Phillips C E, Ting T C, Demurjian S A. Information Sharing and Security in Dynamic Coalitions[C]//ACM. Processingsof 7th ACM Symposium on Access Control Models and Technologies. Monterey: ACM Press: 87-96, 2002.

[6] Joshi J B D, Bhatti R, Bertino E, et al. Access-control Language for Multidomain Environments[J]. IEEE Internet Computing, 8 (6): 40-50, 2004.

[7] Han R F, Wang H X, Xiao Q, et al. A United Access Control Model for Systems in Collaborative Commerce [J]. Journal of Network, 4(4): 279-289, 2009.

[8] Zhang X W, Li Y J, Nalla D. An Attribute-based Access Matrix Model[C]//ACM. Proceedings of the 20thAnnual ACM Symposium on Applied Computing: 359-363, 2005.

[9] Thomas R K, Sandhu R S. Towards a Task-based Paradigm for Flexible and Adaptable Access Control in Distributed Applications[C]. Proceedings of 1992-1993 workshop on New Security Paradigms: 139-142, 1993.

[10] Thomas R K, Sandhu R S. Task-based Authentication Controls (TBAC): A Family of Models for Active and Enterprise-oriented Authentication Management[C]. Proceedings of IFIP WG11.3 workshop on Database Security: 166-181, 1997.

[11] Bertino E, Bonatti P A, Ferrari E. TRBAC: a Temporal Role-based Access Control Model [J]. ACM Transactions on Information and System Security, 4(3): 191-233, 2000.

[12] Joshi J B D. A Generalized Temporal Role-based Access Control Model [J]. IEEE Transactions on Knowledge and Data Engineering, 17(1): 4-23, 2005.

[13] Park J, Sandhu R.Towards Usage Control Models: Beyond Traditional Access Control [C]//ACM. Proceedings of the 7th ACM Symposium on Access Control Models and Technologies: 57-64, 2002.

[14] Park J, Sandhu R. The UCONABC Usage Control Model [J]. ACM transactions on information and system security, 7(1): 128-174, 2004.

[15] Denning D E. A Lattice Model of Secure Information Flow [J]. Communications of the ACM, 19(5): 236-243, 1976.

[16] Denning D E, Denning P J. Certification of Programs for Secure Information Flow [J]. Communications of the ACM, 20(7): 504-513, 1977.

[17] D. Bell, L. J. LaPadual. Secure computer system: Unified exposition and MULTICS interpretation. Technical Report MTR -2997 Rev.1, Bedford, MA: The MITRE Corporation, 1976.

[18] D.E. Bell. Security policy modeling for the next-generation packet switch. In: Proc. of the IEEE Symp on Security and Privacy. IEEE Computer Society Press: 212~216, 1988.

[19] Myers A C, Liskov B. Protecting Privacy Using the Decentralized Label Model [J]. ACM Transactions on Computer Systems, 9(4): 410-442, 2000.