The Open Cybernetics & Systemics Journal

BENTHAM OPEN

Research and Practice of DataRBAC-based Big Data Privacy Protection

Huang Lanying, Xiong Zenggang, Zhang Xuemin,Wang guangwei, Ye Conghuan

The Open Cybernetics & Systemics Journal, Volume 9, 2015

ISSN: 1874-110X
DOI: 10.2174/1874110X20150610E002

Article Type: Research Article

Received: April 17, 2015
Revised: April 22, 2015
Accepted: April 27, 2015

Provisional PDF Publication Date: June 10, 2015

Research and Practice of DataRBAC-based Big Data Privacy Protection

Huang Lanying, Xiong Zenggang*,Zhang Xuemin,Wang guangwei, Ye Conghuan

School of Computer and Information Science, Hubei Engineering University, Hubei, 432000 China

Abstract: With the presentation of the era of big data, big data not only brings our life and production positive help and bus also exposes us to uncertain risk. How to ensure a secure and private big data against current technical conditions is an urgent issue subject to answer. Facing the challenges of subscriber privacy protection and access control and others, the paper mainly studies the Role-Based Access Control (RBAC) and analyzes the demand for data item control from the point of view of data security and privacy protection and then raises an access control model based on data item and achieves the multiple control to the data item in the interface based on role when various subscriber operates same function. The practice application further proves its value.

Keywords: Big data, Privacy protection, Role-Based Access Control, Information security

1 INTRODUCTION

The increasingly developed informationalized and networked society sees an explosive data growth. The statistic data discloses that about two million users in average per second use Google search engine, more than 4 billion pieces of information each day are shared by Facebook user and more than 340 million pieces of information land Twitter each day. Meanwhile, massive data continuously surge through various fields from scientific computation, medicine and health service, finance sector to retail department, etc. A global information total in 2014 has risen to 5.6ZB. However, a more surprising total of 8.2ZB would be available in 2015. Such has caused wide concern [1]. In the era of big data, the analysis and study of information and data would be more complicated and miscellaneous and difficulty for management. The investigation and statistic result tells us that the global data inventory formed during last three years exceeds that generated during last four hundreds of years. Unending supply of information would require more strict data protection in terms of security and privacy. The security and privacy of big data would be increasingly highlighted. How to face and resolve such a challenge would be a global issue [2].

The paper briefs the current situation about big data security and mainly discuss the role-based access control (RBAC) and put forward a data-based item control model which could make various user who operates same function achieve multiple control to the data item in the interface in accordance with their role right for the purpose to assure subscriber privacy secure, data content credible and verifiable and access controllable. The application practice proves that a little modification in terms of normatively to existing system is enough to realize fine management based on data item, which would be of certain reference value for building and modification of large and complicated multiple-user system.

*Address correspondence to this author at the Department of School of Computer and Information Science, Hubei Engineering University,hubei,432000,China; Tel:13117008668;E-mail: xzg@hbeu.edu.cn

2 SECURITY CHALLENGE FROM BIG DATA

Technique could benefit us but also expose us under risk. Same concern comes from data. When we enjoy the value created by the data, we also must face the possible safety impact. The PRISM event moreover aggravates the concern about big data. Compared with traditional information security, the challenge from big data security mainly focuses on the following aspects.

2.1 Subscriber Privacy Protection in Big Data

As disclosed by the police stations across Beijing, the phone fraud accounts for 42% of the annual total case reports in 2014. Among phone fraud case reports, about 62% comes from information leakage via QQ, mailbox, network shopping, network recruitment and network dating. It verifies that many enterprises, to some extent, leak the personal information of user. The fact shows that improper big data treatment would extremely threaten user privacy. In terms of protected object, privacy protection has three types, i.e. position, relation and identifier. At the era of big data, the threat which the user privacy is not limited to personal information leakage, also including the analysis and prediction of the user status and behavior with big data. Now, many businesses think the privacy protection simply means dealing with information via anonymous way and publicizing the information which doesn't carry user identifier. However, the experience indicates such insufficient [3, 4, 5]. In general, there is no applicable standard and criterion to supervise the acquisition, storage, use and management of the user data up to now. In addition, the user would not be informed where their privacy would be applied.

2.2 Dependability of Big Data

The current popular opinion about data is that data means fact and may fully prove all. However, data in fact has certain fraudulence. A person is easy to be deceived by data which has not been picked. The fraudulence of big data is mainly

embodied at two aspects: forged data and distorted data. For certain effect, it is possible to use counterfeit data to make false form which would induce data analyst. The size and diversity of data makes it difficult to judge its reality; the difficulty in judgment may lead to wrong conclusion. Further, the error present during data acquisition and storage would is prone to data distortion which would in turn impact adversely on analysis result [4].

2.3 Access Control of Big Data

Access control is effective measure to make data controlled and shared. Because big data could be applied against various scenarios, access control demand is very urgent. The characteristic and difficult of big data access control lies in [4]:

(1)Difficulty role preset and classification

Big data have wide application and are usually visited by such user as comes from different organizations or departments or carries different identities and purposes, such means access control is basic demand. However, at the era of big data, many users are subject to right management, and the actual right requirement of the user is unknown. Facing massive unknown data and user, it would be very difficult to preset the role.

(2)It is hard to foresee actual right of each role

Because big data scenario contains massive data, safety administrator may lack sufficient professional knowledge, which would make the safety administrator impossible to accurately define data access range for user. On the other hand, from the angle of efficiency, to define all the authorization rules for certain user is also not best choice.

In addition, various big data presents respective access control demand. For example, WEB individual user data corresponds to an access control demand based on history record, geographical data to an access control demand based on scale and data accuracy, stream data processing to an access control demand based on data interval, etc. How to untidily describe and express access control demand is also a challenge.

3 BIG DATA PRIVACY PROTECTION BASED ON DATARBAC

Access control technique was born at the end of 1960s and came from the demand to manage the authorized access to the shared data in the mainframe. Its thought and method has been applied to each field of the computer system. Its basic function is preventing system access by illegal user and illegal system resource use by legal user. In a modern information system, the access control technique is mainly role-based access control (RBAC).

3.1 Role-Based Access Control (RBAC)

The role-based access control was raised by NIST in 2000 on the base of RBAC96 model. RBAC supports three principles of safety: principle of least privilege, principle of separation of duties and principle of data abstraction. It mainly features that a role is defined based on data security policy and respective operating permission; Then, the user would be designated with an specific role by which the user could indirectly visit information resources. The reference model includes four model components which belong to different levels. The four model components are respectively basic model RBAC0(Core RBAC), RBAC1(Hierarchal RBAC), RBAC2(Constraint RBAC) and RBAC3 (Combines RBAC). Refer to Fig.1 for RBAC0 model.

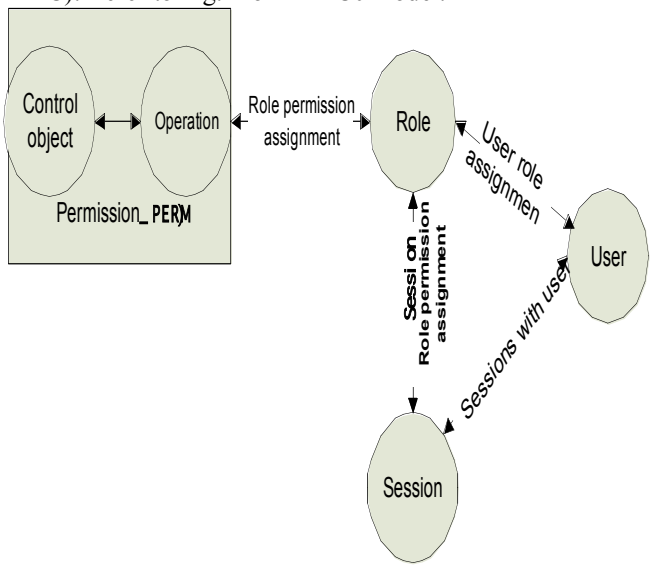

Fig.1 RBAC0 model

RBAC includes five basic data elements of users , roles , objects and operations and permissions (PRMS). The right is assigned to role but not user. When one role is designated to one user, the user would own the right attached to the role. The sessions are mapping between user and enabled role set. The difference between RBAC0 and traditional access control is the addition of a layer of indirection which brings flexibility. RBAC1, RBAC2 and RBAC3 are the expansion based on RBAC0. RBAC1 introduces inter-role inheritance relation; RBAC2 model has the addition of separation of duties relation; RBAC3 comprises RBAC1 and RBAC2, which means two additions of inter-role inheritance and separation of duties.

A lot of projects have been launched domestically about research and improvement of access control model. For example, Xufeng and Lai Haiguang and others put forward a role access control mode facing workflow and the concept of service and authorization migration [6]. The popularization of web service also leads to a lot of domestic research about web-based access control model [7]. Chen Weihe and other raised double-web access control model based on task and role [8]; Long Qin and others introduced a role-based manageable access control model ERRBAC, etc.[9]. The research and practice launched domestically and abroad indicates that RBAC model has become mainstream access control mode in modern business operation information system. It could achieve flexible right management and

enhance system augmentability and applicability. However, RBAC also has its shortcoming. For example, the access control accuracy in most information system just extends to system function level [10,11]. But, the era of big data raises more and more strict requirement for refined data management, which demands the information system competent to refine access control to data item level as required, which means the access control for data item has various selections of Access allowed, Access prohibited, Change allowed and Change prohibited which are all based on operator role against application interface under same function menu. A simple mode is further refining the function menu, i.e. separate such a service as subject to further control. However, such would make service function menu multiplied and beyond control.

3.2 Data Access Control Based on Big Data

In the practice, the access control of system has two divisions, one based on service logic lay, the other based on function [12,13]. For example, in the service system, user A (society insurance operator) usually acts as user role to get the authorization and use the authorized right to inquire the society insurance participant against function access control. But, the society insurance operator is limited to inquire the participant under his administration. Such access control needs to be achieved via service logic. If service authorize user A the right to modify some data items, the modification could be realized by data item control model. In the above model, it is necessary to analyze the operation of the data item, which differs from the access control model based on function menu. Basic operation of data item includes:

1) Data read right, i.e. whether certain user has right to read certain data item.

2) Data modification right, i.e. whether user has right to modify data item in each record, including data item deletion and data value modification.

3) Constraint to data input range: When user has right to modify, what is the modification range? For example, in the society insurance payment, the constraint condition is that the payment base shall float between 60% and 300% of the social average salary. Therefore, the modification range shall be subject to above constraint. On the other hand, because some society insurance participants are beyond such constraint, it means the operation would be executed by the operator with special authorization.

4) Modification aging for data item: i.e. the modification of some data items in certain record shall meet the requirement about time, which means modification is impossible at the time other than that specified.

3.3 Construction of Data Item Access Control Model

In order to solve the above issue discussed above, the paper introduces an access control model based on data item. The model mainly emphasizes the access control to the data item on the base of functional RBAC.

DataRBAC is composed of the following models: data item model, right model, role model and user model (see Fig.2).

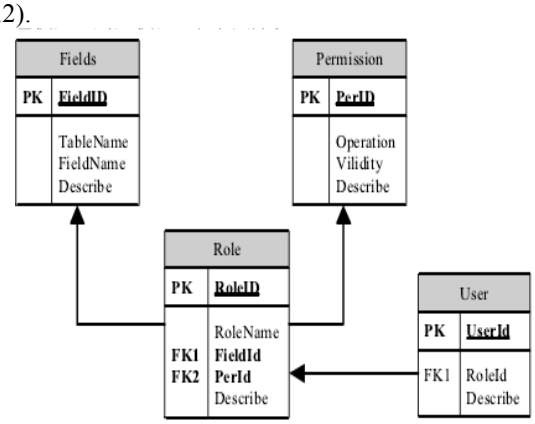

Fig.2 Access control model based on data item

Data item model: Data item Id, sheet name, data item name, data item description. Among them, the sheet name and data item name is sole in the sheet.

Right model: Right id, right name, aging constraint, right description.

Role model: Role id, role name, data item id, right id, description

User model: User ID, role ID, description. Usually, the concept of user group would be introduced in general project. Some rights would be authorized to a user group, and the user would belong to certain given user group.

3.4 Application of Data Item Access Control in Information Protection

The information system used by Hubei Qinlong Logistics Park is a typical big data service system with highlighted big data feature. The system has wide coverage and complicated service logic, which means massive data would be generated. Moreover, the law and regulation demands statutory long data preservation. Such massive data about individual and enterprise formed by service shall be private even confidential by protection. The access necessary for service and management shall be under strict control. With Qinlong Logistics Park as the example, its information system includes more than 3200 tables, hundreds of storage processes, large relational database system with a three-year data capacity of 5TB and above. Its core service system owns more than 15000 pc programs. The transaction per minute is up to 20000 pieces. Against big data, any function addition to its service system means that performance, reconstruction and maintenance shall be of great concern. DataRBAC provides a data protection method the addition of which would not impact on system efficiency. For example, in the right model, if small constraint is down to day, it would be allowed to convert data type to char (8) so that the system performance could be improved. The data item access control is a general extension of access control. It could be achieved by combination of functional right control and service logic in the course of system design and planning. And, the

reconstruction is relatively easy to the existing system. The reconstruction could be completed by the way of database or service logic filter. After a piece of service logic acquires data with a role based on function space, the filtering to the data reported by the server would be launched in accordance with the content of the data access control of the session user. The function at the application interface would be initiated to control the data item at the interface on the base of returned value via filtering [12,14]. For enterprise information system, we mainly depend on calling the same type to achieve the data item access control. Refer to Fig.3 for detailed processing flow.

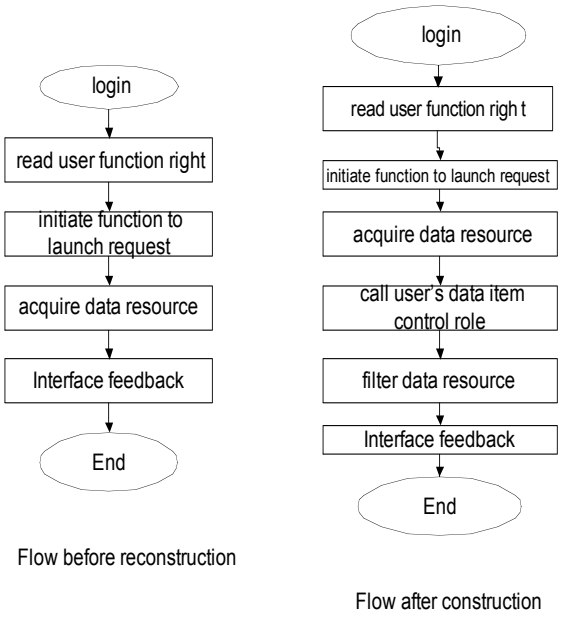

Fig.3 Comparison of flow charts before and after data item access control

Because data item access control is a flexible structure, the value assignment would be unnecessary for the data in the system which is not subject to control. That is to say, the value assignment only applies to the data item which shall be under control. Such would further improve the performance and flow (see Fig.4).

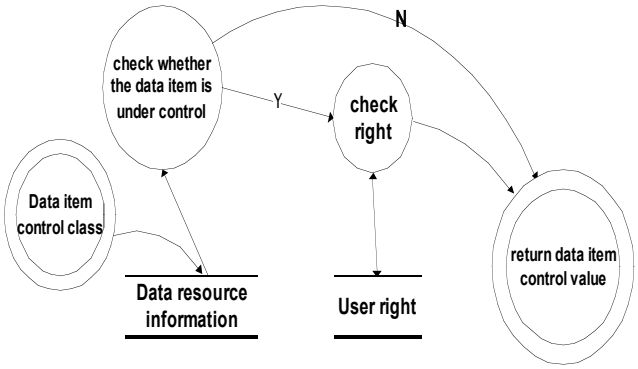

Fig.4 Value assignment course of data item control class

For such a data item as has no inquiry right, use "*" to substitute original data value when returning back to user interface. For such a data item as has inquiry right but has no modification right or is beyond aging, the control shall be launched at the interface end.

4 CONCLUSIONS

The research about big data technique provides scientific base for data integration, analysis and knowledge discovery. However, the feature of big data drives the people to pay more attention to data privacy protection. The data item access control model introduced in the paper realizes the refined management to the service operating system. The achievement balances data access performance and data security and to some extent lowers the complexity and maintenance of the information system. The application at the enterprise management information system indicates that the model has relatively weak impact on the system performance, high maintainability and could effectively prevent leakage of user privacy. The model could expand in the right table on the base of service type. For more complicated system, a further expansion could be made in terms of role inheritance and constraint on the base of RBAC model. On the other hand, for a complicate information system which features numerous user types, a further research is still expected at the aspects of authentication right and filtering efficiency.

CONFLICT OF INTEREST

The author confirms that this article content has no conflict of interest.

ACKNOWLEDGMENT

This work is partly supported by National Natural Science Foundation of China(No.61370092), Natural Science Foundation of Hubei Province of China(No.2013CFC005,No.2014CFB188),Hubei Provincial Department of Education Outstanding Youth Scientific Innovation Team Support Foundation (No.T201410) and Hubei Provincial Department of Education Research Foundation(No.20122604, , NO.2012367, NO.XD2014255) .

REFERENCES

[1]Viktor Mayer-Schonberg, Kenneth Cukier, *"Big Data: A Revolution that Will Transform How We Live, Work and Think"*, Houghton Mifflin Harcourt, Boston, 2013

[2] X. Meng, and X. Ci, "Big data management: Concepts, techniques and challenges", *Journal of Computer Research and Development*, vol.50, pp.146-169 , 2013.

[3] S. Wang, H. Wang, X. Qin, and X Zhou, "Big data architecture: Challenges, Present and Future", *The Computer Journal,* vol.34 ,pp.1741-1752, 2013.

[4] D. Feng, M. Zhang, and H. Li, "Big data security and privacy protection Computers", *Chinese journal of computers,* vol.37, pp. 246-258, 2014.

[5] Q. Wei, and Y. Lu, "Advances location privacy protection", *computer science*, vol. 135 , pp.21-25, 2008.

[6] F. Xu, "A service-oriented role-based access control technology", *Computer Technology*, vol. 28, no. 4, pp.687-693, 2005.

[7] X. Yanxue, Q. Wang, and H. Tai. Ma, "Web services access control model", *Magazine Computer Science*, vol. 35, pp.38-41, 2008.

[8] W. Chen, X. Yan, B. Mao, and L. Xie, "Dual access control model for Web-based tasks and roles", *Computer Research and Development*, vol. 41, no. 9, pp.1466-1473, 2004.

[9]L. Qin, P. Liu, and A. Pan, "Such as role-based access model to manage the expansion and implementation of control", *computer research and development*, vol. 42 , pp.868 ~ 876, 2005.

[10]H. Shen, and F. Hong, "Survey of Research on Access Control Model", *Computer application research*, no.6, pp.9-11, 2005.

[11]Chang dou, Song Mei-na, Yang Jun, "Role based access control and data entry methods", *software*, Vol. 35: 40-43,2004

[12]Lin Li, Huai Jinpeng, Lixian Xian, "Attribute-based access control strategy for the synthesis of algebra", *Software*, Vol. 20,PP. 403-413,2014

[13]Arasu A, Chaudhuri S, Chen Z,etal., "Experience with using data cleaning technology for bing services", *IEEE data Engineering Bulletin*, Vol.35,PP.14-23,2012

[14]Wang Q, Jin H., "Quantified risk-adaptive access control for patient privacy protection in health information systems",Proceedings of the 6th AC'M Symposium on Information, Computer and Communications Security(ASIACCS' 2011),Hong Kong, China, pp.406-410,2011