*Send Orders for Reprints to reprints@benthamscience.ae*

*The Open Cybernetics & Systemics Journal*, 2015, *9*, 669-673

Open Access

# Research and Practice of DataRBAC-based Big Data Privacy Protection

Huang Lanying, Xiong Zenggang*, Zhang Xuemin, Wang Guangwei and Ye Conghuan

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

**Abstract:** With the emergence of the era of big data, big data not only helps production, it also exposes us to uncertain risks. 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 studied the Role-Based Access Control (RBAC) and analyzed the demand for data item control from the point of view of data security and privacy protection and then established an access control model based on data item to achieve multiple control of the data item in the interface based on roles when various subscribers operate the same function. The practical application further proved its value.

**Keywords:** Big data, Information security, Privacy protection, Role-based access control, Security challenge, RBAC.

## 1. INTRODUCTION

The increasingly developed informationalized and networked society is experiencing significant growth in data. The statistical data discloses that about two million users on average use the Google search engine, more than 4 billion pieces of information are shared by Facebook users each day and more than 340 million pieces of information through Twitter each day. Meanwhile, massive data continuously surge through various fields from scientific computation, medicine and health service, finance sector to retail department, etc. The total global information in 2014, reached up to 5.6ZB. However, a more surprising total of 8.2ZB was expected in 2015. This has caused wide concern [1]. In the era of big data, the analysis and study of information and data would be more complicated and difficult for the management. The investigation and statistical results show that the global data inventory formed during the last three years exceeded that generated during the last four hundred years. Unending supply of information would require more strict data protection in terms of security and privacy. The security and privacy of big data are being increasingly highlighted. How to face and resolve such a challenge is a global issue [2].

The paper explained the current situation about big data security and mainly discussed the role-based access control (RBAC) and put forward a data-based item control model which could make various users, who operate the same function, achieve multiple control of the data item in the interface in accordance with their roles meant to assure the subscriber's privacy, credible and verifiable data content and controllable access. The application proved that a little modification in the existing system is enough to realize effective management based on data item, which would be of certain reference value for building and modifying a large and complicated multiple-user system.

## 2. SECURITY CHALLENGE FROM BIG DATA

The above discussed technique could benefit us but also expose us to the risk. The same concern arises in data. When the value created by the data is appreciated, the possible safety impact must also be considered. The PRISM event, moreover, aggravates the concern about big data. Compared with the traditional information security, the challenge from big data security mainly involves 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 the phone fraud case reports, about 62% cases reported information leakage *via* QQ, mailbox, network shopping, network recruitment and network dating. It verifies that many enterprises leak the personal information of the user to some extent. The fact shows that improper big data treatment would extremely threaten user privacy. In terms of the protected object, privacy protection has three types, *i.e.* position, relation and identifier. In the era of big data, the threat is that the user privacy is not limited to personal information leakage, rather it also includes the analysis and prediction of the user status and behavior with big data. Now, many businesses enterprises are of the view that the privacy protection simply means dealing with information *via* anonymous way and publicizing the information which does not carry user identifier. However, the experience indicates that this is insufficient [3-5]. In general, there is no applicable standard and criterion to supervise the acquisition, storage, use and management of the user data up till now. In addition, the users would not be informed where their privacy would be applied.

### 2.2. Dependability of Big Data

The current popular opinion about data includes facts. However, data in fact has certain fraudulence. A person is easy to be deceived by data which has not been accessed. The fraudulence of big data is mainly embodied in two as-

*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

1874-110X/15

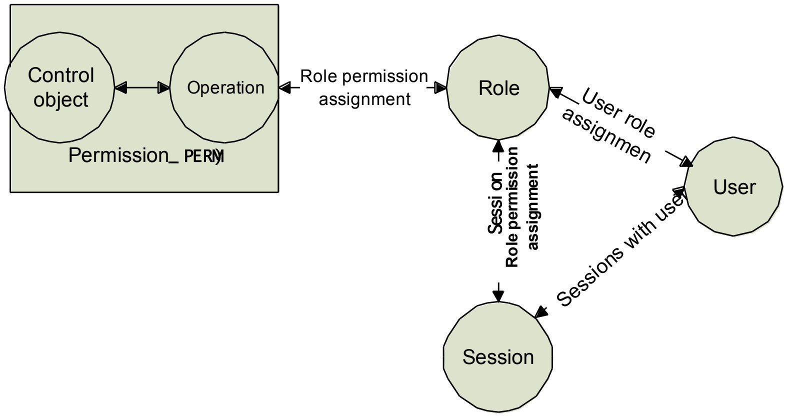

**Fig. (1).** RBAC0 model.

pects: forged data and distorted data. For certain effect, it is possible to use counterfeit data to be used by data analyst. The size and diversity of data make it difficult to judge its reality; the difficulty in judgment may lead to wrong conclusion. Furthermore, the error present during data acquisition and storage is prone to data distortion which would in turn adversely impact the analysis results [4].

### 2.3. Access Control of Big Data

Access control is an effective measure to control and share data. Because big data could be applied in various scenarios, access control demand is very urgent. The characteristic and difficulty in big data access control are discussed below [4]:

(1) Difficulty role preset and classification

Big data have wide application and are usually accessed by users from different organizations or departments, incorporating different identities and purposes, therefore, the access control is the basic demand. However, in the era of big data, many users are subjected to effective management, and the actual right requirement of the user is unknown. Due to massive unknown data and numerous users, it would be very difficult to preset the roles.

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

Because big data scenario contains massive data, the safety administrator may lack sufficient professional knowledge, which would make it impossible for the safety administrator to accurately define data access range for the users. While, to bring efficiency, defining all the authorization rules for certain user is also not the best choice.

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

## 3. BIG DATA PRIVACY PROTECTION BASED ON DATARBAC

Access control technique was introduced in the end of 1960s as a result of the demand to manage authorized access to the shared data in the mainframe. Its ideas and method have been applied to each field of the computer system. Its basic function is preventing system's access by illegal users and the use of illegal system resource by legal users. In 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 introduced by NIST in 2000 on the basis 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 a 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 the basic model RBAC0(Core RBAC), RBAC1(Hierarchal RBAC), RBAC2(Constraint RBAC) and RBAC3 (Combines RBAC). Refer to Fig. (**1**) shows the RBAC0 model.

RBAC includes five basic data elements of users, roles, objects and operations and permissions (PRMS). The right is assigned to roles but not to the user. When one role is designated to one user, the user would own the right associated to the role. The sessions are mapped between the user to enable the set role. The difference between RBAC0 and the 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 additional separation of duties; RBAC3 comprises RBAC1 and RBAC2, which means addition of both inter-role inheritance and separation of duties.

A lot of projects have been launched domestically based on research and improvement in the access control model. For example, Xufeng, Lai Haiguang and others put forward a role access control mode due to the workflow based on the concept of service and authorization migration [6]. The popularization of web service has also led to a lot of domestic research on web-based access control model [7]. Chen Weihe and other studied double-web access control models based on task and role [8]. Long Qin and others introduced a role-based manageable access control model ERRBAC, etc.

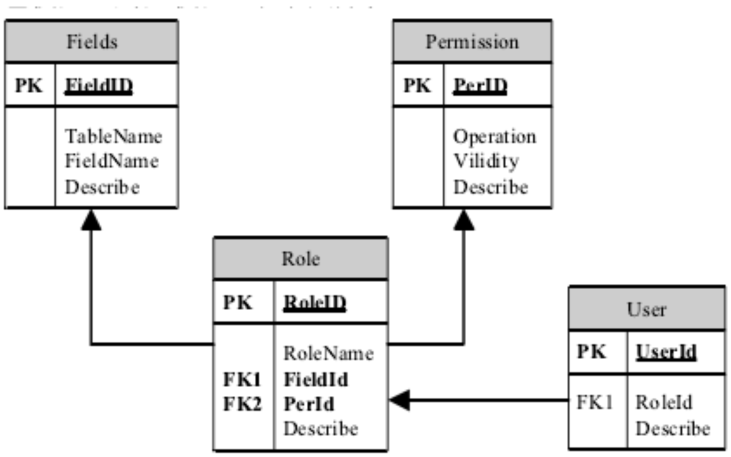

**Fig. (2).** Access control model based on data item.

[9]. The research and application launched domestically and globally indicate that RBAC model has become the mainstream access control mode in modern business operation information system. It could achieve flexibility, effective management and enhance system's augmentability and applicability. However, RBAC also has its shortcomings. For example, the access control accuracy in most information systems just extends to the system function level [10, 11]. But, the era of big data demands more strict requirement for refined data management, which demands competent information system to refine the access control to data item level., This means that the access control for data item has various options like access allowed, access prohibited, change allowed and change prohibited which are all based on the operator's role against application interface under the same function menu. A simple mode is to further refine the function menu, *i.e.* separate such a service subject to further control. However, this would enhance the service function menu multiply and go 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, and the other is based on function [12, 13]. For example, in the service system, user A (society insurance operator) gets authorization to inquire the society insurance participation 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 the service authorizes user A with 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 the function menu. Basic operation of data item includes:

1) Data read right, *i.e.* whether a certain user has the right to read certain data item.
2) Data modification right, *i.e.* whether the user has the right to modify data item in each record, including data item deletion and data value modification.
3) Constraint in the data input range: When the user has the right to modify, what is the range of modification? For example, in the society insurance payment, the constraint condition is that the payment base should remain between 60% and 300% of the social average salary. Therefore, the modification range is subjected to the above constraint. On the other hand, because some society insurance participants are beyond such constraints, 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 of time, which means modification is impossible at the time other than the specified time.

### 3.3. Construction of Data Item Access Control Model

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

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

Data item model: It includes: data item Id, sheet name, data item name, data item description. Among them, the sheet name and data item name are the key elements of the sheet.

Right model: It includes: right id, right name, aging constraint, and right description.

Role model: It includes: role id, role name, data item id, right id, and description

User model: It includes: user ID, role ID, and description. Usually, the concept of user group is 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 that massive data can be generated. Moreover, the law and regulation demand statutory long data preservation. Such massive data about individuals and enterprises formed by service should be kept private and even confidential by protection. The access necessary for service and management shall be under strict control. With Qinlong Logistics Park, as the example, the 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 in addition to its service system such as performance, reconstruction and maintenance shall be of great concern. DataRBAC provides a data protection method, the addition of which would not impact the system's 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. Moreover, the reconstruction is relatively easy in the existing system. The reconstruction could be completed by using database or service logic filter. After the service logic acquires data with a role based on function space, the filtering of the data reported by the server would be carried out in accordance with the content of the data access control of the session user. The function of the application interface is initiated to control the data item at the interface on the basis of returned value *via* filtering [12, 14]. For enterprise information system, mainly the same type is considered to achieve the data item access control. Refer to Fig. (**3**) shows the detailed processing flow.

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

For such a data item that has no inquiry right, "*" is used to substitute the original data value when returning back to the user interface. For such a data item having inquiry right but no modification right or is unlimited, the control is launched at the interface end.

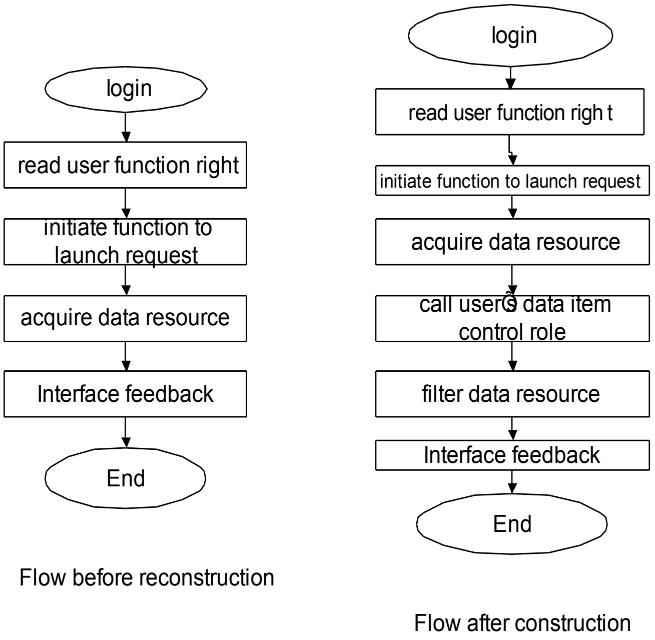

**Fig. (3).** Comparison of flow charts before and after data item access control.

## CONCLUSION

The research on big data technique provides scientific basis 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 realized effective management of the service operating system. The achievement balanced data access performance and data security and to some extent lowered the complexity and maintenance of the information system. The application of the enterprise management information system indicates that the model has relatively weak impact on the system performance, with high maintainability and it could effectively promote user privacy. The model could expand in terms of the right table on the basis of service type. For more complicated system, further expansion could be made in terms of role inheritance and constraint on the basis of RBAC model. While, for a complicated information system which features numerous user types, further research is still expected on the aspects of authentication right and filtering efficiency.

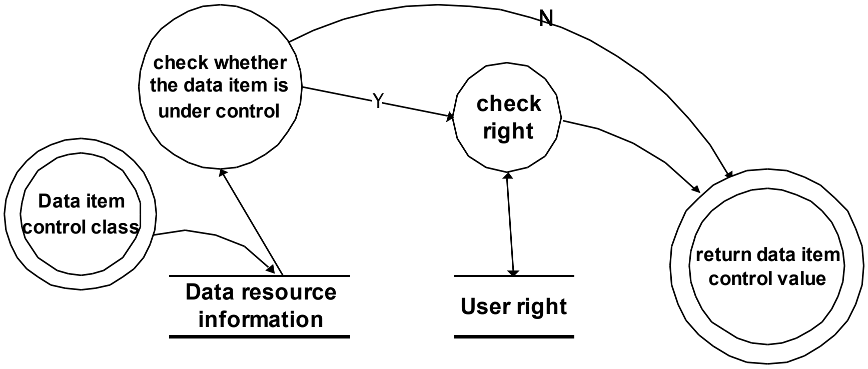
check whether the data item is under control
N
Y
check right
Data item control class
Data resource information
User right
return data item control value

**Fig. (4).** Value assignment course of data item control class.

## CONFLICT OF INTEREST

The authors confirm that this article content has no conflict of interest.

## ACKNOWLEDGEMENTS

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] V. Mayer-Schonberg, and K. 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] D. Chang, M. Song, and J. Yang, “Role based access control and data entry methods”, *Software*, vol. 35, pp. 40-43,2004

[12] L. Li, H. Jinpeng, and L. Xian, “Attribute-based access control strategy for the synthesis of algebra”, *Software*, vol. 20, pp. 403-413, 2014.

[13] A. Arasu, S. Chaudhuri, Z. Chen, K. Ganjam, R. Kaushik, and Narasayya, *“*Experience with using data cleaning technology for bing services”, *IEEE Data Engineering Bulletin*, vol. 35, pp. 14-23, 2012.

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

Received: April 02, 2015 Revised: May 23, 2015 Accepted: June 06, 2015