

General Identity Management Model for Big Data Analysis

Feng GAO*, Feng ZHANG**, Junjie XIA*, Zheng MA*

*Network Technology Research Institute, China Unicom, Beijing China

** Academy of Information Science Innovation, CETC, Beijing China

gaofeng149@chinaunicom.cn, feng3982315@163.com, xiajj2@chinaunicom.cn, mazheng@chinaunicom.cn

Abstract—Nowadays, the big data analysis is more and more widely used in information and telecommunication domain. Benefits from the big data analysis, both the telecommunication operators and Internet service providers could analyze user's interests more effective and predict user's expectation more accurate thus significant improve and add value of their services. Identity management in big data analysis and application has its own requirements with considering the characteristics of big data. In this paper, we analyzed the requirements of identity management for big data analysis, and based on that, present a general identity management model with the intention that compatible advanced and diversity identity management technologies. Refer to the ITU-T Recommendation "Baseline capabilities for enhanced global identity management and interoperability", we analyze the identity management capabilities of this general identity management model. According to the analysis, the general identity model can satisfy the identity management capabilities required in ITU-T standard. Since the design of this model based on the specific requirements and characteristics of big data analysis and application, this prototype general identity management model can be used in the big data analysis and also can guide the system design for big data security.

Keywords—big data; analysis; identity management model; security

I. INTRODUCTION

Nowadays, the big data analysis is more and more widely used in ICT domain. Big data analysis has special characteristics including big volume, variety, velocity, and value while its applications cover multiple/cross application domains, for example, apply big data analysis from telecommunication domain to customize the services of e-pay, e-finance.

Benefits from the big data analysis, both the telecommunication operators and Internet service providers could analyze user's interests more effective and predict user's expectation more accurate thus significant improve and add value of their services.

Generally, big data technologies provide two benefits, one is a cost efficient and the other one is new insight from the large data volumes. As the new technology, big data analysis will bring new security issues. The Cloud Security Alliance (CSA) has published "the expanded top ten big data security

and privacy challenge [1]" to guide the security research for big data. As the basic and also the important security mechanism, the tailored identity management for big data analysis and application can mitigate the relevant security threats effectively.

The rest of this paper is organized as follows. In the next section, we summarize some related work. In Section III, we analyze the requirements of identity management in big data analysis. In Section IV, we introduce the general identity management model for big data analysis. Finally, in Section V, we conclude this paper in which we also discuss some future work.

II. RELATED WORK

The big data is very new technology, in the international standardization research domain, some activities for standardization of big data are just launching.

(1) ISO

ISO/IEC JTC1 SG32 establish "SG Next Generation Analytic and Big Data", to focus on concept model of big data, requirements of big data standard, and big data technology report.

(2) ITU

ITU-T SG 13 set up a new work item "Requirements and capabilities for cloud computing based big data," to study requirements, capabilities and use cases of cloud computing based big data.

Big data security is agreed as new topic in the ITU security workshop, September 2014 (TD1682) and also in ITU-T April meeting, 2015.

(3) NIST

NIST initiatives a big data public working group (NBD-PWG), its deliverables includes:

- develop big data definitions
- develop big data taxonomies
- develop big data requirements
- develop big data security and privacy requirements,
- develop big data security and privacy reference architectures
- develop big data reference architectures
- develop big data technology roadmap

As the innovative technology, big data is a very new topic in the standardization works. The SDOs are just launching big

data related work. Currently, there is no published standardization about identity management for big data analysis and application.

Now the big data related works are under developed, the security work need to establish in the very beginning to avoid big data standardized without dedicated security research that unsuitable guide big data develop and apply in commercial scenarios.

In the academic research domain, the big data security related works cover privacy protection [2-3], key management [4], secure data distribution [5], encryption algorithm [6], etc. The identity related works focus on specific technologies, such as key management [4], access control etc. Since these specific technologies only focus on certain aspects of identity management, the general identity management model which gives the whole picture of identity management for big data analysis need to further study.

III. IDENTITY MANAGEMENT REQUIREMENTS IN BIG DATA ANALYSIS

Identity management in big data analysis and application has its own requirements with considering the characteristics of big data. In this section, the specific identity management requirements of big data will be analysed.

A. Manage Multiple Types of User

The identity management in big data analysis needs to manage multiple types of users.

(1) Individual user

This type user uses the big data sets to mine valuable information.

(2) Machine type user

The machine-to-machine (M2M) devices that highly automated process and analysis big data sets are split up into more manageable portions and processed separately across Hadoop clusters.

(3) Internal domain user

Internal domain user owns the big data sets (e.g. the analyst of Google analyze the big data sets which belongs to Google), it can be individual user or machine type user.

(4) External domain user

External user does not own the big data sets, it only uses data sets to do the analysis according to the agreement with the data owner (e.g. the third party app uses the open interface of telecommunication operators to obtain and analyze the subscriber's interests, under subscriber's consent of course).

External user also can be individual user or machine type user.

(5) Combine domain user

The users share their own big data to do the analysis in combined data sets. This type of user also can be individual user or machine type user.

B. Customized Authentication

The identity management needs to offer customized authentication way with considering the multiple types of users, for example, compatible using password/random short

message/smart card to authenticate internal users and using certificate/bio-identifier to authenticate external users.

C. Fine Grained Authorization

The authorization for big data analysis and application involves diverse factors, e.g.:

- privilege/constraint of data process during its life cycle
- scope of data analysis algorithm usage
- time period of data analysis algorithm usage
- intention of data analysis algorithm usage
- result application of data analysis algorithm usage
- usage of Hadoop clusters scale
- frequency of data aggregation and analysis
- others

Consequently, it needs to design fine grained authorization mechanism for big data analysis and application.

D. Comprehensive Auditing

The comprehensive auditing is needed, it may include:

- rationality of the analyzed content
- the compliance of analysis procedure
- the security of user
- others

IV. GENERAL IDENTITY MANAGEMENT MODEL FOR BIG DATA ANALYSIS

A. Identity Management Model

According to the requirements analysis in section three, we design the general identity management model for big data analysis as follows.

The original intention of this model design is offering the prototype of general identity management, which can compatible advanced and diversity identity management technologies, such as certification based identification, smart card based authentication etc.

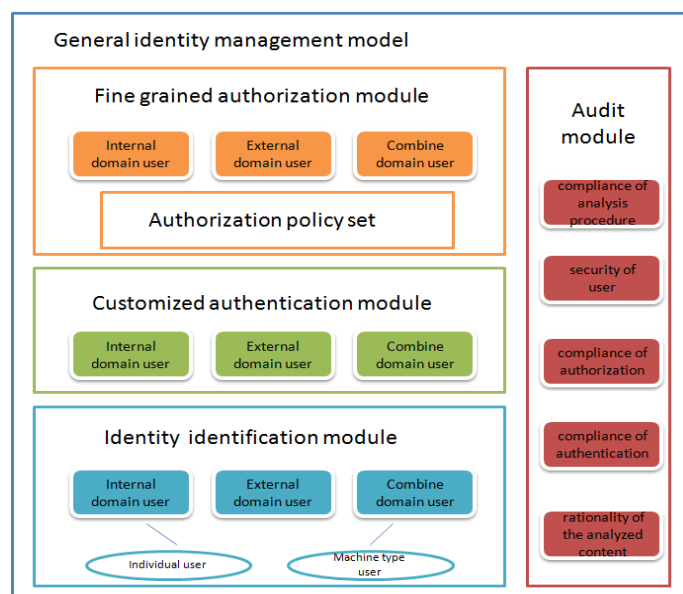


Figure 1. Architecture of identity management model

The general identity management model includes four models:

(1) Identity identification module

This module will identify the entity of big data analysis requirements, which includes internal domain user, external domain user, and also combine domain user.

(2) Customized authentication module

This module pre-defines the authentication mechanism for different user type.

For the internal domain user, the authentication method will define according to the internal domain security policy. It can use password, random short message to authenticate user.

For the external users, the authentication method will also defined according to the internal domain security policy, but generally, it will use higher security level policy than internal user to avoid un-know security risk.

For the combine domain user, since the users share their own big data to do the analysis in combined data sets, the authentication method will defined according to the agreement among combine domains. It is recommended that this authentication method based on the highest security level policy among combines domains.

(3) Fine grained authorization module

This module authorizes the bid data analysis privileges for the user. The authorization mechanism can be diversity, such as trust based authorization, certification based authorization etc., and the authorization policies stored in the authorization policy sets.

The basic factors for fine grained authorization need to consider follows:

- privilege of data process during its life cycle
- constraint of data process during its life cycle
- scope of data analysis algorithm usage
- time period of data analysis algorithm usage
- intention data analysis algorithm usage
- result application of data analysis algorithm usage
- usage of Hadoop clusters scale
- frequency of data aggregation and analysis

(4) Audit module

This module will audit the operations of identity management.

The audit includes but not limited to follows:

- rationality of the analyzed content
- compliance of authentication
- compliance of authorization
- compliance of analysis procedure

V. ANALYSIS FOR THE GENERAL IDENTITY MANAGEMENT MODEL

ITU-T has published the recommendation ITU-T X.1250 [7] to guide the identity management system design. The general identity management capabilities are defined in X.1250, based on it, analyze the utility of the proposed general identity management model as follows.

TABLE 1. UTILITY ANALYSIS FOR THE GENERAL IDENTITY MANAGENET MODEL

Capability in X.1250	Capability in General Identity Management Model	Module of the capability implementation
The ability of a requesting party to access/delete/modify/monitor/control its own identity information.	✓	This capability can be implemented in the identity identification module.
The ability of authorized entities (e.g., system administrators, parents, public safety, law enforcement, and other authorized third parties) to access/modify/monitor its identity information.	✓	This capability can be implemented in customized authentication module.
The import/export of identity information.	✓	This capability can be implemented in the identity identification module.
A mechanism to indicate some kind of information about the quality level of the information that they provide to relying parties.	✓	This capability can be implemented in customized authentication module.
The ability for a requesting party to delegate the management of its identity information to another entity. Lifecycle management for all identities.	✓	This capability can be implemented in the identity identification module.
A common mechanism to identify and control the dissemination of all identities	✓	This capability can be implemented in the identity identification module.

According to the analysis, the general identity model can satisfy the identity management capabilities required in ITU-T standards. Since the design of this model based on the specific requirements and characteristics of big data analysis and application, this prototype general identity management model

can be used in the big data analysis and also can guide the system design for big data security.

VI. CONCLUSIONS

Big data analysis is more and more widely used now, benefits from the big data analysis, service providers could analyze user's interests more effective and predict user's expectation more accurate thus significant improve and add value of their services. Identity management in big data analysis and application has its own requirements with considering the characteristics of big data. In this paper, we firstly analyzed the requirements of identity management for big data analysis, and based on that, present a general identity management model with the intention that compatible advanced and diversity identity management technology. This prototype general identity management model can be used in the big data analysis and also can guide the system design for big data security.

In the future, we will further refine the general identity management model through implementation and applications. We will establish experimental environment to do the simulation for this general identity management model to value its performance in larger scale concurrent requirement scenario. We will also trigger the identity management related standardization work in ITU-T SG17 in the study period 2016.

REFERENCES

- [1] The Cloud Security Alliance (CSA), "Expanded Top Ten Big Data Security and Privacy Challenge".
- [2] D.S. Raghuwanshi, and M.R. Rajagopalan "Practical data privacy and security framework for data at rest in cloud," *Computer Applications and Information Systems (WCCAIS)*, 2014, pp1-8.
- [3] L. Kaitai, W. Susilo, and J.K. Liu "Privacy-Preserving Ciphertext Multi-Sharing Control for Big Data Storage," *IEEE Transactions on Information Forensics and Security*, 2015, 10(8): 1578-1589.
- [4] S. Ningning; S. Zhiguo; S. Yan; Z. Xianwei "Multi-matrix combined public key based on big data system key management scheme," *International Conference on Cyberspace Technology (CCT 2014)*, pp1-6.
- [5] L. Xiao et al. "Secure Distribution of Big Data Based on BitTorrent," *2013 IEEE 11th International Conference on Dependable, Autonomic and Secure Computing (DASC)*, pp82-90.
- [6] A. Rahmani, A. Amine, and R.H. Mohamed "Multilayer Evolutionary Homomorphic Encryption Approach for Privacy Preserving over Big Data," *2014 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC)*, 2014, pp 19-26.
- [7] ITU-T X.1250: Baseline capabilities for enhanced global identity management and interoperability.
- [8] ITU-T X.1251: A framework for user control of digital identity.
- [9] ITU-T X.1255: Framework for discovery of identity management information.
- [10] ISO/IEC DIS 11179-1: Information technology -- Metadata registries (MDR) -- Part 1: Framework.
- [11] ISO/IEC DIS 19763-1: Information technology -- Metamodel framework for interoperability (MFI) -- Part 1: Framework.
- [12] ITU-T: Y.BigData-reqts, Requirements and capabilities for cloud computing based big data.
- [13] DRAFT NIST Big Data Interoperability Framework: Volume 6, Reference Architecture.
- [14] DRAFT NSIT Big Data Interoperability Framework: Volume 4, Security and Privacy.



Feng Gao was born in 1984, she received her Ph.D. degree in computer application technology from Beijing University of Technology in 2012, Beijing, China. She is currently working in Network Technology Research Institute, China Unicom as a senior engineer. Her main work and research interests include network security, mobile telecommunication technology and

telecommunication standardization etc.□

□
□



Feng Zhang was born in 1982, he received his Ph.D. degree in computer science and technology from Beijing University of Posts and Telecommunications in 2014, Beijing, China. He is currently working in Academy of Information Science Innovation, CETC as an engineer. His research interests include big data analysis, Internet of Things technology and applications, and network applications etc.□

□
□



Junjie Xia was born in 1975, he received his M.S. degree in telecommunication technology from Nanjing University of Posts and Telecommunications, Nanjing, China. He is currently working in Network Technology Research Institute, China Unicom as a department manager. His main research interests include network security, mobile telecommunication technology and telecommunication standardization etc.

telecommunication technology and telecommunication standardization etc.

□
□



Zheng Ma was born in 1981, he received his M.S. degree in computer science and technology from Beijing University of Posts and Telecommunications, Beijing, China. He is currently working in Network Technology Research Institute, China Unicom as a senior engineer. His research interests include telecommunication network security, information security and mobile telecommunication technology etc.