Design of Data Security Model in Fog Computing

IOP Conf. Series: Materials Science and Engineering **490** (2019) 042044 doi:10.1088/1757-899X/490/4/042044

Design of Data Security Model in Fog Computing

Xuelian Xiao, Shuqing He, Haifeng Wang*

Information Science and Engineering School, LinYi University, LinYi, China

*Corresponding author e-mail: gadfly7@126.com

Abstract. edge computing system includes a large amount of storage devices, which are applied to access the large-scale data from the fog computing equipments to reduce the communication cost incurred by the large-scale data movement. However the security of storage system in the edge of network is significant. To improve the data security in fog computing system, a novel security storage model is proposed to compose the architecture design and data security management scheme. We discussed the cooperative mechanism and important data security management algorithms. Additionally, the multi-level trusting domain design scheme is provided to enhance the defense capacity to prevent the attacks from the interior networks. The experimental result indicates that this model can hold the performance loss under the acceptable range. The proposed model has better data security and scalability. So it will be used into the fog computing system or edge computing system.

1 Introduction

The fog computing or Edge computing is novel computing scheme, which improve the computing capacity in network edge to handle the large-scale data collected by the large-scale sense devices [1]. So this computing paradigm is to reduce the large-scale data communication from the network edge to the data center. To the fog computing systems, they contain a large number of fog nodes, which are in distributed geographically locations and have a variety computing resources [2,3]. In order to enhance the computing capacity, the fog node should be improved the storage capacity. The fog computing system enables the users to query the global data, which are in the distributed computing nodes. However the distributed computing severs in the edge of networks are in vulnerable network environment. Accessing data security of fog computing system is significant issue [4,5]. But the data security of fog computing system is hard to deal with due to the fact each fog node has limited computing and communication capacity. So we focus on the specific architecture of fog computing system and build a novel data security storage model. This model is composed by the trusted domain design and the architecture design. Our aim is to reduce the performance loss to adapt to the computing and communication capacity of fog node and to defense the interior and external attacks from the fog computing environment. This paper is organized as follows: Section 2 introduces the aim and issue; Section 3 provides the detailed scheme of data security model; Section 4 provides the experimental result and the analysis of performance loss. Finally, summary our research work.

2 Research Issues

The fog computing system should provide the data querying service for a lager amount of users. The data stored in distributed nodes within network edge are in vulnerable environment. This issue is very significant to guarantee data security of fog computing system. On the other hand, the storage architecture of fog computing is very complex and it may dynamically change the form to adapt to the

requirements. To deal with the particularity of fog computing architecture and adapt to the variation of storage system, the data security model should separate the security authentication and the data management. The multi-level authentication mechanism can reduce the performance bottleneck of center node. Additionally, we assume that the storage network is untrusted and it includes the variety of storage devices and the authentication severs. The data transformation channel is not credible. There exists many risks, such as data spying, network data sniffing [6]. In brief, the data attack model is as abovementioned. The following model is designed based on the attack model.

3 Storage Model in Fog Computing

a) Architecture Design Scheme

This section proposed the detailed architecture design scheme for fog computing or edge computing. This data security model for fog computing or edge computing is denoted as FEDSM. From the computing architecture perspective, this model separates the data storage and user authentication. The architecture of this model includes data access and user authentication process. The data accessing module has a large amount of fog storage nodes (FS), which are responsible for hold the large-scale data from large edge devices. In the data access management module contains some fog storage tracker nodes (FST), which can manage the FS nodes by heartbeat mechanism. FST can deal with the workloads balance of data access and the data management cooperative issues. On the other hand, FEDSM has other module called data authentication model, which has authentication nodes denoted as AS. The AS node control the user login the fog storage system by verifying the user legitimacy [7]. To reduce the performance bottleneck of the central authentication node, we design multi-level tree authentication architecture as shown in Fig. 1. The root authentication node is AS_0, which control the global authentication and manage the lower authentication nodes, such as (AS_1, AS_2, ..., AS_m). The authentication design is easy to improve the system expansive.

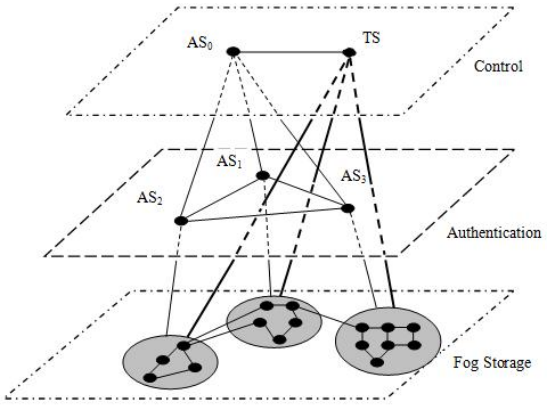


Figure 1. FEDSM system structure diagram

From the software perspective, the data security model has three layers shown in Fig. 1. The bottom layer is fog storage layer, which is responsible for containing the large-scale data and communication with FST. The middle layer is authentication layer, which are control the user authentication of different network regions. The top layer is control layer, which is in the trusted networks and manages the data accessing and user authentication. In short, FEDSM can be divided into two planes. The control plane has control layer and authentication layer. The data plane has fog storage layer. So this model realizes the separation of control plane and data plane.

The architecture design of this model has three advantages. The first one is to separate the control plane and data plane, which allows the storage system adapt the variation of requirements. The second one is to design multi-level authentication mechanism to reduce the performance bottleneck of root authentication node. The third one is integrate the storage management node and the root authentication

node in the same layer with high network channel. This makes the storage module and the data security module cooperative better.

b) Cooperative Mechanism

Here we provided the detailed process about the collaborative mechanism, which guides the data storage module with the authentication module. This process is as follows: a user through Internet starts a inquiring of fog computing data, which is in the distributed fog storage system. Then the root authentication node AS_0 checks the identification of user and assigns the according lower authentication node AS_i for this user. After that, the AS_i is responsible for controlling the data accessing operations. On the other hand, the storage management module should provide the location of data in fog storage system to guide the users find their data.

c) Data Synchronization

The data synchronization is significant for the collaborative work between the storage manage module and the authentication management module. The two different modules should communicate with other and exchange the information. The authentication module needs to know the profile of distribution about data storage system. And the data storage management module should obtain the assignment of the authentication scheme. The detailed data synchronization process can be described as : if status of fog node vary, this node gives out a status updating information to the FST. Then FST updates the global storage status information and exchanges this information to the root authentication node. Finally, AS_0 is responsible for dispatching the data location status information to the lower authentication nodes.

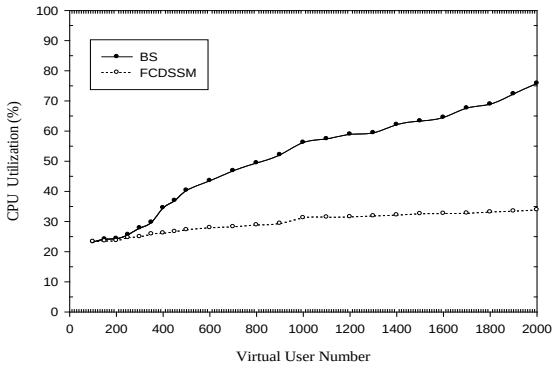
d) Authentication mechanism

In this subsection, we describe the detailed method about of the authentication mechanism in the fog storage system. The authentication mechanism is important in the multi-tenant environment due to the fact that is an vulnerable network environment. The users communicate with the authentication nodes at various levels through the client program to implement identity authentication to defense the replay attacks. The specific identity authentication method is described as follows: the user submits the request information to the authentication nodes. AS node returns a random number R_t and time *t* to the user. Then the user encrypts the random number to R_{t+1} with the private key and obtains Epu (R_{t+1}) . So it sends the information to the authentication node AS. The AS decrypts Epu (R_{t+1}) with the user's public key. If it can decrypt correctly, it proves the user's identity.

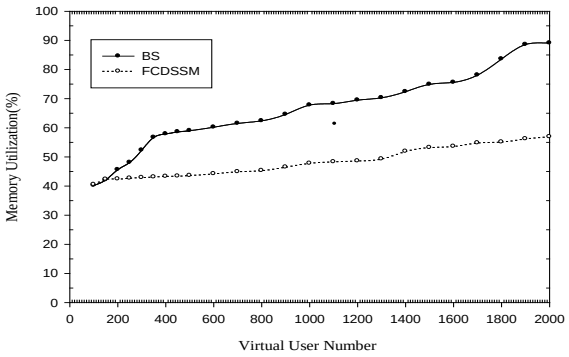
4 Experiment and Analysis

We designed a simulation environment to verify the authentication performance about this model [11]. This simulation environment is based on the cross-city university networks and each campus deploys 30 fog nodes. The hardware of fog node is as follows: CPU- Intel Core i5, memory-4GB, SSD-256GB, hard disk-1TB. The authentication system is deployed as two layers and each lower authentication node is responsible for manage 10 fog nodes. To simplify the experiment, we use the notebook computer as the authentication node, whose configuration is as follows: CPU i5-5200U, 4GB memory, 1TB hard disk.

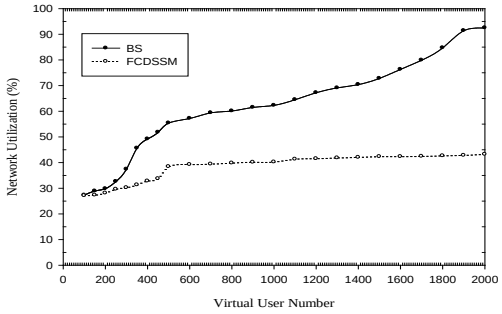
The first aim is to verify the system scalability of the fog storage model. We used the software called LoadRunner to generate the large-scale user inquiring events. This can give the fog storage system a large request pressure. The stimulation inquiring events have five sets as shown in Fig. 2. Each inquiring needs the fog storage node return a specific size data block to the user. To compare with our proposed model, we selected the central authentication scheme as the benchmark model dented as BS. The benchmark model just has a single authentication node, which needs to manage all the fog nodes in the networks.



(a) CPU utilization comparison chart



(b) Memory utilization comparison chart



(C) Network bandwidth utilization comparison chart

Figure 2. Performance comparison between FEDSM and benchmark server

As can be seen from Figure2, the turning point of the performance of the FEDSM system is 500 virtual users. As 500 users or more users making demands, the capacity of the benchmark system with the single authentication server starts drop slowly. As the number of the virtual user is nearby 2000, the capacity of the benchmark system begins to drop dramatically. That is to say, as the concurrent user number exceeds 2000, it is difficulty for the benchmark system to improve its capacity, and the benchmark system reaches its scalability limit. After analyzing, the reason is that the system with the single authentication server manages massive global user demands in only one node, and the resources of the system will suffer a great pressure and lead to the performance inefficiency. On the contrary, the capacity of our proposed architecture obtains better result and does not to lead to performance loss much, which the reason is that the architecture adopts the multi-authentication servers and multi-layers controller mechanism. As a result that all the devices of our architecture endures the pressure of the resource consumption with the users number increasing and exceeding 500 far more than. Most users should communicate with the root authentication node once. Then having determined the lower authentication nodes served for them will stop connecting with the root authentication nodes. The results can be obtained that the architecture of this model has good scalability.

Secondly, we analyze the capacity gap by comparing the distributed storage file system HDFS with the FEDSM system, mainly from the perspective of read and write performance of the big files and massive small files. The first step is to test the result of big files capacity of read and write demands, which is initiated by the remote users, and the range of the size of the tested big files is set from 100M bytes to 500M bytes, each time read and write test unit is set 64K bytes. The big files are deployed in the benchmark system in a system environment similar to the simulation fog, where the location of the nodes of the fog layer take the place of the storage units, and each storage unit are place to the two other fog nodes and as backup.

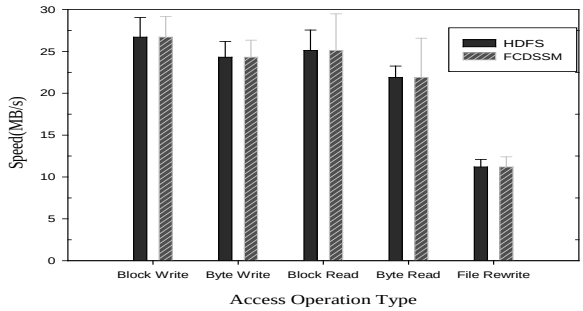
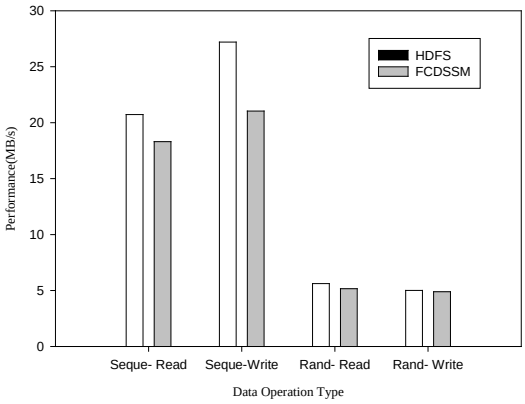


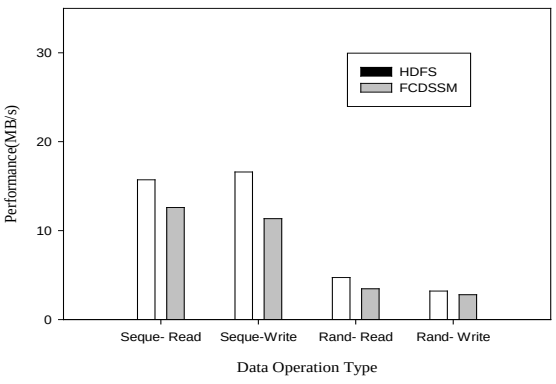
Figure 3. Performance comparison of individual fog nodes with HDFS (with copies)

As shown in Figure 3, only consider the capacity of the read and write files operations of a fog unit, from left to right are the 5 basic operations of file block write, byte-level write, file block read, byte-level read and file rewriting, testing contrastively between a single node in HDFS and FEDSM. The experimental result show that the writing operation capacity of the FEDSM system is better than the benchmark system and the reading operation capacity of the files block granularity is 26.3% lower than the benchmark system, and the reading operation capacity of the files bytes granularity is decreased by 36.8%. Read operation capacity degradation is caused by file-level content encryption and decryption overhead, especially when byte-level granularity is used to read and write operation, resulting in a significant drop due to extensive accessing to access verification, integrity verification and hash value recalculation.

Then test the reading and writing operations capacity of large-scale small files by reading and writing on the massive this type files operations. The size of each small file is set in 1024K bytes; the scale of the files is set to 1000 and 5000, using IOzone3.3 [12]. Data files are performed 4 basic operations of sequential-read, sequential-write, random-read and random-write respectively, and read and write speed are recorded. As shown in Figure 4, the capacity of the FEDSM system is worse in testing reading large-scale small files, the reason is that reading large-scale small files need to interact firstly the authentication operation and encrypted meta-data. The proportion of the process of the whole read and write is a larger one, so small files' read performance is worse than that of large ones. Through the reading and writing operations comparison in this section, we find that the random-writing's capacity is the very undesirable and worse than the other's operations and sequential-writing is very desirable and better than the other's operations. In addition, the capacity of encrypted and decrypted data manner is found to be greater than the HDFS system. In order to optimize the capacity of large-scale small files, it is necessary to increase the cache design in fog storage nodes units, and reduce the communication cost of the network and the reading and writing times of disk. Especially it is to use complex storage management optimization mechanism to improve the capacity of the fog computing system.



(a) 1,000 small files



（b）5000 small files

Figure 4. FEDSM and HDFS small file performance comparison

5 Summary

In the paper, we study on the data security storage model based on the feature of the fog computing, and proposed a complete scheme and algorithm for collaborative work, authentication management and multi-layer key management. To meet requirements of the evolution of data storage system to complex systems, we designed FEDSM from logical design angle to archive the decoupling of data accessing and user authentication. This model is consist of two independent planes of data accessing and user security The data accessing layer allows dynamic adjustment of the data storage scheme without affecting the design of data security. Therefore, the FEDSM mentioned in this paper has a better flexibility from the software point of view. In order to leverage the characteristics of the geographical distribution of fog computing, we established multi-domain hierarchical authentication structure from the perspective of system design. Compared with HDFS in experiment, it was found that the reading operation capacity of large-scale medium-size files obtain very pretty efficiency. Because of improve the capacity of data security, it lead to a certain decrease of the reading and writing operation capacity, which is an accepted result compared to obtain the advantage of the data security. But when it came to massive small files the read performance was poor, needing further study of the performance optimization problem. In the future, the performance optimization of massive small files may be studied from two angles, that is, from using large-size memory as cache and from optimizing the hybrid storage system.

Acknowledgment

This project is supported by Shandong Provincial Natural Science Foundation, China (No. ZR2017MF050), Project of Shandong Province Higher Educational Science and technology program (No.J17KA049), Shandong Province Key Research and Development Program of China (No.2018GGX101005,2017CXGC0701,2016GGX109001)Shandong Province Independent Innovation and Achievement Transformation, China (No. 2014ZZCX02702).

References

[1] Ranesh K. N. Saurabh. G. et al. Fog computing:Survey of Trends, Architetures, Requirements and Research directions[J] IEEE Access, 2018,6(9):47980-48008.
[2] Sun Y. , Zhang N. A resource-sharing model based on a repeated game in fog computing[J] Saudi J. Biol. Sci., 2017,24(3):687-694.
[3] Mahmud R. , Koch F. L. Buyya R. Cloud-fog interoperability in IoT-enabled healthcare solutions[C] In ICDCN, 2018,ACM,32:1-32:10.
[4] Arwa Alrawais, A. Alhothaily, et al. Fog computing for the internet of things:security and privacy issues[J] IEEE Internet Computing,2017,21(2):34-42.
[5] Zhang Ming, Chen Wei, Distributed Security Storage Model for Large-Scale Data[J] Journal of Mathematics and Computer Science, 2017, 17(4):488-505.
[6] Xia, Q., Xu, Z., Liang, W., Zomaya, A.Y.. Collaboration- and Fairness-Aware Big Data M-anagement in Distributed Clouds[J], IEEE Transactions on Parallel and Distributed Systems, 2016, 27(7):1941-1953
[7] Junhee Ryu, Dongeun Lee,et al.File-System-Level Storage Tiering for Faster Application Launches on Logical Hybrid Disks[J] IEEE ACCESS, 2017,4:3688-3696.
[8] Hong Liu, Huansheng Ning, et al. Shared authority based privacy-preserving authentication protocol in cloud computing [J] IEEE Transaction on Parallel and Distributed System,2015,26(2):241-251
[9] Shuibing He, Yang Wang, et al. Boosting parallel file system performance via heterogeneity-aware selective data layout [J] IEEE Transaction on Parallel and Distributed Systems, 2016,27(9):674-675.
[10] Mingwei Lin, Riqing Chen,et al. Efficient Sequential Data Migration Scheme Considering Dying Data for HDD/SSD Hybrid Storage Systems[J] IEEE Access, 2017,PP(99):1. M. Etemad, M. Aazam. Using DEVS for modeling and simulating a fog computing environment[C] In Workshop on Computing,Networking and Communications, 2017,849-854
[11] IOZone. FileSystem benchmark tool[EB/OL] http://www.iozone.org/,2017-6-27