ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 3, March 2015

Security and Privacy Issues in Private Cloud Storage

Girija Rajendra

PG Researcher, Department of Computer Science and Engineering, Davangere, Karnataka, India

ABSTRACT: With the expansion of business, associate degree enterprise would like to create its PSC (private storage cloud) approach associate degree infrastructure service in a very Partner/Public Cloud. In such PSCs, there are a unit some new security problems, First, a way to isolate the data keep within the PSC from internal and external attackers; Second, a way to create secure intra-cloud information migration inside an enterprise, Third, a way to secure inter-cloud information migration between the PSC and therefore the Partner/Public Cloud. during this paper, the proposed associate degree design of imposing security services on the layer of HDFS, together with information Isolation Service, Secure Intra-Cloud information Migration Service, and Secure Inter-Cloud Data Migration Service Finally, a prototype has been implemented based on HDFS by our security policies, and the time cost is given and evaluated..

KEYWORDS*:* private cloud storage; partner cloud; public cloud; isolation; intra-cloud data migration; inter-cloud data migration

I. INTRODUCTION

Storage Cloud is an rising technology that leverages commodity hardware tied along to seem as one storage device by the package, like a cluster application, distributed file system (DFS) and grid computing [1]. HDFS (Hadoop DFS) [2] is AN open supply project and was designed to dependably store terribly giant files across machines in a large cluster nowadays, there's growing interest for several enterprises to create a non-public Storage Cloud (PSC) supported DFS like HDFS, which can be closely-held and managed internally. However, with the growth of business, AN enterprise would love to create its PSC approach the infrastructure service (e.g. storage service) in a very Partner/Public Cloud [3], that might greatly relieve them from management and maintenance of storage infrastructure. But in such a PSC, some new security problems could arise. Data Isolation issue seems once a corporation has place some of its knowledge on the PSC on Partner/Public cloud.

Data Isolation- As associate enterprise continually partitions its own non-public Storage Cloud (PSC) into multiple sub-clouds by departments or regions for management; it demands that the information of various sub-clouds be isolated from one another. An enterprise hopes that its knowledge hold on within the PSC may well be isolated from those within the Partner/Public Cloud. When a number of its knowledge square measure placed within the Partner/Public Cloud, the enterprise needs them to be isolated from different enterprises' knowledge rest within the same Partner/Public cloud [3]. Data Isolation issue seems once an organization has place some of its knowledge on the PSC et al on Partner/Public cloud.1.First, internal attackers might access unauthorized knowledge of different sub-clouds (or storage clusters) within the cloud; 2.Second, external attackers within the Partner/Public Cloud might intercept or tamper with the enterprise's knowledge hold on within the 2 clouds (private storage cloud, and partner/public cloud).

Intra-Cloud knowledge Migration- Load-balance service and Fault-tolerant service on the layer of HDFS can typically mechanically initiate large-scale knowledge migration from one location to a different at intervals the cloud. As his duty changes within the company, the employee's knowledge keep in some sub-cloud of the PSC are going to be migrated manually into a brand new sub-cloud consequently. If some knowledge don't rest within the storage node near the traveller, however he can oftentimes access these knowledge throughout successive amount of your time, then it\'s helpful to migrate (or replicate) these knowledge into those storage nodes near the traveller mechanically. Intra-Cloud knowledge Migration issue might happen between different sub-clouds for knowledge replication, load-balance, or restructuring at intervals the PSC.1.First, Associate in Nursing assaulter might create the information being migrated

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 3, March 2015

into a wrong location and create it accessible for a few unauthorized users; 2.Second, a malicious user might faux an interior migration request from some legal user to initiate Associate in Nursing sudden migration and create the information inaccessible for the legal user.

Inter-Cloud knowledge Migration- With the enlargement of business, the corporate chooses to rent some cupboard space within the Partner/Public cloud and moves a number of its business knowledge from the PSC to the new cloud. For timely market research, an organization would love to frequently get some public or shared knowledge from the Partner/Public Cloud and replicate them into the PSC for additional analysis. If Associate in Nursing worker has got to work for many days or months at some place wherever he will not approach the company's PSC however can get to the Partner/Public Cloud. Then, he will replicate his knowledge from the PSC into the Partner/Public Cloud before this trip, and the other way around. Inter-Cloud knowledge Migration issue might occur between the PSC and Partner/Public cloud for business demand. During such external migration 1.First, malicious intercept or modification risks might arise once sensitive/critical knowledge square measure migrating between the 2 clouds; 2.Second, the provision of information is crucial. as an example, a faux response of fortunate migration from the target cloud can truncate the backup copies within the supply cloud, resulting in missing knowledge.

II. RELATED WORK

A. HDFS Architecture: HDFS is meant to dependably store terribly massive files as a sequence of information blocks across machines in an exceedingly massive cluster (see Figure 1). In HDFS, the Name Node executes filing system namespace operations like gap, closing, and renaming files and directories, and determines the mapping of blocks to Data Nodes [2]. The Data Nodes area unit accountable for serving read and write requests from the file system's shoppers. The Data Nodes conjointly perform block creation, deletion, and replication upon instruction from the Name Node. HDFS's default duplicate placement policy is to place simple fraction of replicas on completely different nodes among identical rack to enhance performance, and place the opposite third of replicas on willy-nilly chosen nodes on different completely different racks to enhance availableness.

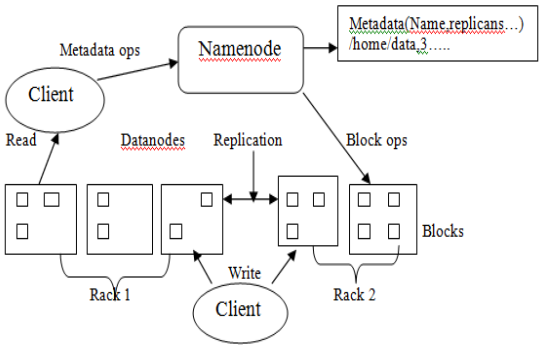

Figure1: Architecture of HDFS

B.PSC Based on HDFS: Nowadays several enterprises would love to make its own PSC (private storage cloud) supported HDFS [3]. As depicted in Figure two, this sort of PSC consists of sub-clouds separated by completely different attributes (e.g. departments, regions or security levels) and will reach a Partner/Public Cloud when necessary. In such a PSC, it includes 3 layers: HDFS, the lowest layer for managing an oversized quantity of storage resource pools; SLA (Service Level Agreement) , the middle layer for outlining service norms and charge; and Cloud Service Interface, the higher layer for users to use cloud services, like net portals, network disk, custom interfaces(e.g. REST, SOAP), and normal access protocols(e.g. FTP, BT). Unremarkably distributed file systems like HDFS, there are in the main following 5 services.

1) Fault-Tolerant Service: is chargeable for the dependableness of the cloud, together with failure nodes detection, data replication policy, knowledge exercise, etc.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 3, March 2015

2) Storage Service: includes name and knowledge storage service. within the distributed classification system for storage cloud, metadata are keep in an exceedingly name server, the big file being withdraw blocks of same size knowledge, and saved in knowledge nodes. Mapping from files to blocks and locating blocks in knowledge nodes is that the task of name server.
3) Node Service: is accountable of the configuration of the system and management of the cloud's nodes, including the low-level formatting of the system, adding and deleting knowledge nodes mechanically.
4) Knowledge Transmission Service: makes use of coding and parallel transmission techniques supported little knowledge blocks with multi-replications to fulfill the wants of security and high QoS of the info transmission.
5) Load Balance Service: is to unfold knowledge between 2 or a lot of hardware resources dynamically, so as to induce optimal resource utilization and minimize latency.

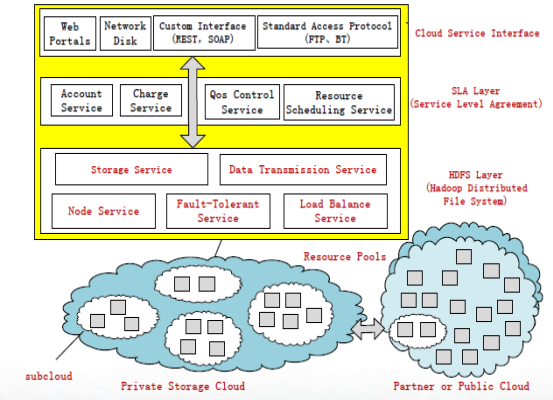

Figure2: Private Storage Cloud based on HDFS

III. SECURITY ARCHITECTURE DESIGN OF PSC BASED ON HDFS

The safety design of the PSC is based on HDFS. As shown in Figure three, the design mainly includes 3 new security services: information Isolation Service, Secure Intra-Cloud information Migration Service and Secure Inter-Cloud information Migration Service. These services mainly depend upon on top of 5 common services represented in Section II. First, once shoppers wish to read/write a file, the Storage Service would question information Isolation Service initially. Second, once Fault-Tolerant Service, Load Balance Service, or Node Service initiates information migration among the cloud, the Secure Intra-Cloud information Migration Service would confirm if the migration is permissible. Third, once a user needs to transfer his information from one cloud to a different one through information Transmission Service, it should send letter of invitation to Secure Inter-Cloud information Migration Service initially. Especially, internal or external information migration basically is to browse the info from source nodes and write them to focus on nodes. So, the latter two migration services lie on the info Isolation Service.

A. Data Isolation Service: Data Isolation Service is liable for maintaining the security of knowledge access (e.g. browse or write) and storage within the PSCs [4]. (Note, if a user tries to access a company's knowledge keep in the Public/Partner Cloud, it conjointly depends on this service in the PSC). It in the main consists of 4 separate modules: access policy management, and access call and authorization, access protocol security and personal encoding. The following square measure their purposeful descriptions severally.

Access Policy Management (APM) is responsible of the management and configuration of a versatile access management policy (see section IV) implemented within the design, including labelling subjects (e.g. roles of employees) and objects (e.g. tags of files by totally different departments or regions), defining permission rules by

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 3, March 2015

logical expressions of object tags, and assigning role-permission relationships. This job will solely be done in the master server by the administrator of the PSC. Any other users cannot modify and bypass this policy.

Access Decision and Authorization (ADA) is responsible of secure access call and authorization implemented within the master server, as well as creating call for Associate in Nursing access request in keeping with the predefined policies within the APM, and producing a token with approved rights. This token are encrypted with a key shared between the master sever and related knowledge nodes, which can be in PSC or different Clouds.

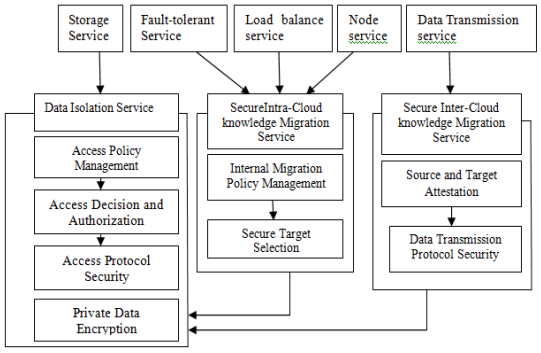

Figure3: Security Architecture of PSC based on HDFS

Access Protocol Security (APS) is responsible of securing access protocol between the master and also the consumer, and the client and also the knowledge nodes, in different words, enhancing security for knowledge transmission management within the PSC. They would negotiate some measures (e.g. token coding rule and key employed in ADA) to guard the token from malicious attackers to listen and to access the info consequently.

Private Data Encryption (PDE) is Associate in nursing no obligatory service for essential knowledge protection supported the consumer of the cloud system. to create the info keep within the cloud a lot of firmly, the data would be encrypted with the assistance of TPM embedded in the client's platform, while not requiring storage of keys within the cloud or third-party key management[8]. Only the user himself within the sure platform will decipher the info with the key. And also, the consumer will decide whether or not to migrate and share the key with users in different sure platforms.

B. Secure Intra-Cloud knowledge Migration Service: Secure Intra-Cloud knowledge Migration Service is accountable for making certain the safety of knowledge replication and shifting caused by the requests from directors or traditional users, and those from load balance service, fault-tolerant service, or node service mechanically inside the PSC. It in the main consists of 2 separate modules: internal migration policy management and secure target choice [5]. Following square measure the functional descriptions of them severally.

Internal Migration Policy Management (IMP) is in charge of the management and configuration of a versatile internal migration management policy (see section IV) implemented in our design, as well as labelling knowledge blocks by some attributes with stratified relationship (e.g. tags for different countries, regions, or departments); labelling knowledge nodes (racks or clusters) by some attributes (e.g. tags for different security levels); shaping permission rules by the mapping operate (block tags, node tags). This job will only be drained the master server by the administrator of the PSC. The other users cannot modify and bypass this policy.

Secure Target Selection (STS) is responsible of judgement the migration request bearing on predefined migration security policies in IMP. Normally, once an invitation is associated with some expected target nodes (e.g. a request from load balance service), this module would do choice for a right node from them, betting on whether or not it\'s in accord with the mapping operate. however once an invitation is only related to Associate in Nursing expected attribute

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 3, March 2015

regarding target nodes (e.g. an invitation from users or administers), it would initially get all allowable nodes by permission rules so more select them in keeping with operations within the storage service.

C. Secure Inter-Cloud knowledge Migration Service: Secure Inter-Cloud knowledge Migration Service is accountable for the safety of knowledge migration between the PSC and Partner/Public Cloud [5]. It will be conjointly a brand new service in current HDFS for PSC. It in the main consists of 2 separate modules: source and target attestation, and knowledge transmission protocol security. Following square measure their functions severally.

Source and Target Attestation (STA) is responsible of attesting the trustiness between 2 clouds, in the way of remote attestation mechanism projected by TCG Remote attestation is aimed to conclude whether or not to trust Associate in Nursing attesting system supported testing and corroboratory every measurement list entry severally. Because the 2 clouds square measure transparent to every different, the supply and also the target throughout attesting mean the master servers in 2 clouds. As remote attestation can value a lot of time, STA is optionally implemented here, only the migrated knowledge is essential to its owner.

Data Transmission Protocol Security (DTP): is in charge of securing knowledge transmission throughout inter-cloud migration, enhancing communication management in HDFS. Before the migrating, the supply and also the target can discuss some measures over SSL to make sure the data's confidentiality, integrity and credibility. SSL-based protocol would opt for Associate in Nursing coding rule and keys for confidentiality, a Hash algorithm for integrity, a waterproof rule for authentication, and temporary authorization tickets for connected knowledge nodes of two clouds to try and do secure transfer in parallel. Only the destination has received all knowledge properly, it returns a hit response to the supply. The supply then clears up the replicas of original blocks. If the response shows the migrating unsuccessful, the DTP ought to act with the fault tolerant service in order to revive the supply to its original secure state.

IV. IMPLEMENTATION

The implementation of three security policies based on HDFS for PSC: 1) a flexible access control policy based on RBAC(Role-Based Access Control) and CW (Chinese-Wall); 2) a label-based intra-cloud data replicating and restructuring policy, 3) an temporary-ticket based parallel inter-cloud data transmission policy. Each policy works on a set of security labels, or permission expression definition separately.

1. Flexible access control policy: Flexibility Different corporations have different internal security requirements. But it needs to provide a flexible security policy which could be easily customized to fit variant security requirements. Data Isolation Storage Cloud is a common platform shared among many corporations, even market competitors. Needs to make sure data owned by one company would not be accessed by other ones if otherwise authorized. Data Sharing Storage Cloud could be utilized as a data sharing point between different corporations. But it needs to provide data sharing mechanisms on storage cloud while still under the restriction of data isolation principle.

RBAC & CW Based Access Policy: is to ensure that 1) the data owned by a user of one sub-cloud wouldn't be accessed by any unauthorized users of other sub-clouds in the PSC, and 2) the data owned by one firm in PSC wouldn't be crossly accessed by other ones, and this policy could be easily customized to fit variant requirements in enterprises. This policy is based on RBAC (Role Based Access Control) and CW (Chinese-Wall policy). It defines a kind of organization label (CW-org), which is assigned to all users and data of the enterprise in the PSC, ensuring that any user associated with a different CW-org cannot access the data in the PSC, and vice versa. It defines a series of roles (RBAC-roles) for subjects (e.g. users) and security tags for objects (e.g. files) both with inheritance relationship. All data in a sub-cloud are assigned to the same set of tags. For any role r associated with subject s, only when all tags of object o meet the logical expression of permission defined by the role manager of the PSC, could the data be accessed by s.

2. Label-Based Intra-Cloud Data Replication & Shifting Policy: Label-Based Intra-Cloud Data Replication & Shifting Policy: is to ensure that 1) the data owned by one user would always be on the nodes with the expected attribute and accessible for legal users before and after migrating, and 2) illegal internal migration request can be

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 3, March 2015

recognized and restricted before the unexpected migrating happens. Data replication HDFS is consisted of one Name Node and some Data Nodes. The Name Node makes all decisions regarding replication of blocks. The necessity for re-replication may arise due to many reasons: a Data Node may become unavailable, a replica may become corrupted, a hard disk on a Data Node may fail, or the replication factor of a file may be increased.

Cluster rebalancing A HDFS cluster can easily become imbalanced, for example, when a new data node joins the cluster, thus increasing the use of network bandwidth. On the other hand, when some data nodes become full, new data blocks are placed on only non-full data nodes, thus reducing their read parallelism. It is important to redistribute data blocks when imbalance occurs. Label-Based Data Migration: According to the security labels marked on all the Data Nodes, and the security rules as one new attribute in all files in HDFS, Name Node would determine which block has the intra-migration right to move into another Data Node except just by some rules HDFS has now [6].

Implementation of Label Marking

- Each Data Node has an array of labels to take these Security labels. Each File has one Security rule.
- The rule is made by some Security labels and three logic symbols, such as "&&", "||", "^".
- The rule is expressed by binary tree. The Name Node need to take the value of each label on each Data Node to the file expression, when the result equals to 1, this file has the right, otherwise has no rights.

Data Node 1	Label 1	Label 2	Label 3	Label 4
	0	1	1	0

Data Node 2	Label 1	Label 2	Label 3	Label 4
	1	0	1	1

Figure4: Label Marking

3. Temporary-Ticket based Inter-Cloud Data Transmission Policy: Temporary-Ticket based Inter-Cloud Data Transmission Policy is to ensure that 1) data blocks could be recognized and protected during the migrating from the source cloud to target cloud, 2) the source and target's trustworthiness before the migrating could be attested, and 3) the parallel transmission could be protected and completed reliably. This policy demands: First, before the migration process, the master servers of the source and target clouds should attest each other to ensure their trustworthiness. Second, the master servers of two clouds should do security negotiation over SSL and return a series of encrypted temporary tickets for all source data nodes to hold(This ticket is associated with the source node ID, the source path of being migrated data block, and a one-time usage information). Once the target master server decrypts the ticket and verifies its validity, then it will return the target data node ID for the data block transmission. Third, the data block transmission is in parallel between the source and target data nodes directly and the data block would be encrypted with the key during the previous phase of security negotiation over SSL.

Using SSL to confirm security parameters including temporary ticket using temporary ticket to indicate the role of worker during transmission process Source cloud should distribute ticket to trusted worker Objective cloud should keep record of temporary tickets to check workers[7].

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 3, March 2015

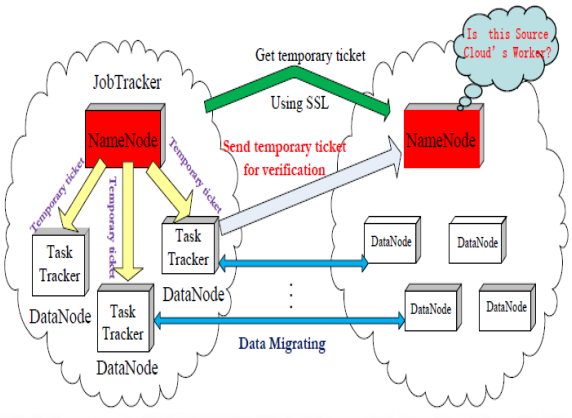

Figure5: Securing Data Migration between private Cloud Storage Systems

V. RESULTS

Deployed a public cloud and a partner cloud with above security services supported HDFS and evaluated their effects on the first cloud system supported HDFS. For information isolation management, there ar 3 major aspects which will have influence on read/write time value, including policy computing, price tag} creation and ticket en/decrypting.By testing on an individual basis, And discover that the overhead is all among 0.15ms~15.96ms, that is nearly negligible, because the information access time is, e.g. 5889ms for reading a 64M data(1block) and 67847ms for writing a 512M data(8 blocks) within the system(figure 6)

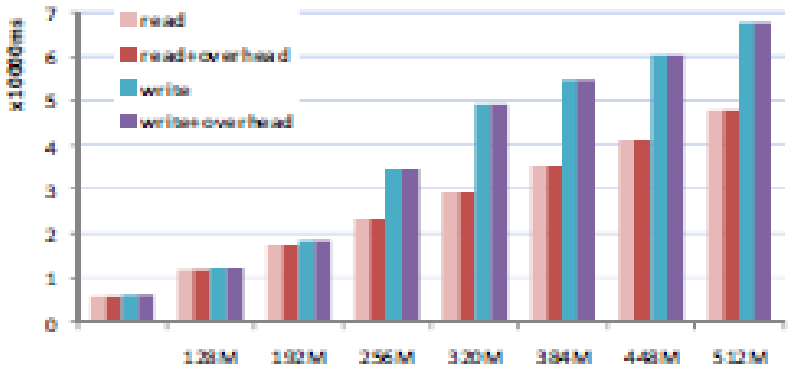

Figure 6. Isolation Control Cost for read and write

For the intra-cloud information migration from block replication,a file uploaded by a user. however the quantity of DataNodes will less impact on the value of security call as every block's uploading is set by the NameNode. and also the average cost of this operation is a smaller amount than five-hitter. For rebalancing, the migrated information capability depends on the common usage rate(the used capability/total capacity altogether DataNodes), and also the total of all DataNodes' excess capability over the common usage rate is our migrated information capability. So, the scale of information capacity has obvious influence on the value of migration. And the average value of rebalancing is regarding 100% because it computes the migrated information capability altogether DataNodes(figure 7)

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 3, March 2015

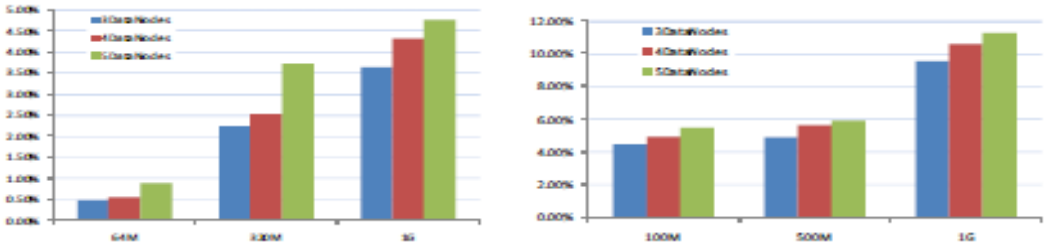

Figure 7. Intra-Cloud Migration Cost for Block Replication(left graph) and Rebalancing(right graph)

In HDFS, distcp may be a typical operation for inter-cloud migration. For secure distcp, the time value is principally from three aspects: SSL negotiation, temporary price ticket distribution and verification, file encrypting/decrypting for secure transmission despite platform attestation, which is optional for important knowledge migrating. By testing in our prototype(see figure 8), the temporary price ticket distribution and verification solely want zero.15-3ms overhead, and SSL negotiation 890ms. the price of en/decrypting operation is larger than the price of on top of 2 phases, e.g. 290726ms for a 384M huge file. however scrutiny to the confidentiality of information, this value is value for many users.

Compared to HDFS while not a PDE service, the time we tend to will used for PDE service (when uploading a file) includes the time it prices to come up with a isosceles key, the time it costs for cryptography of the file transfer filter, and also the time it costs for cryptography of the isosceles key and persistence of the isosceles key. Especially, the time value for encrypting the file transfer filter is offset by the time value for uploading the file. Table 1 shows the time we tend to value once we transfer various styles of knowledge. With the PDE service examination the uploading while not a PDE service. Table 1 shows the time we tend to cost once we transfer files of varied amounts. As shown in the tables, the common overhead of PDE service is concerning 13%, and this is often acceptable for the personal knowledge protection.

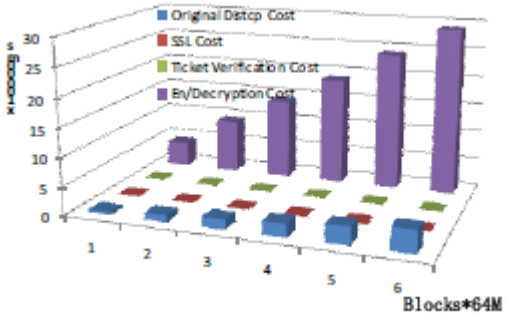

Figure 8. Secure Inter-Cloud Migration Cost for distcp

File size	1K	1M	32M	128M	1G
Upload Cost (%)	21.4	21.1	14.1	4.54	1.21
Download Cost (%)	23.6	20.7	11.7	8.5	3.7

Table1: Time consumption of up/downloading a file

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 3, March 2015

VI.CONCLUSION AND FUTURE WORK

This paper describes security issues on data isolation, intra-cloud data migration and inter-cloud data migration under the environment of a Private Storage Cloud (PSC) extended with a Partner/Public Cloud. In future will make Multi-tenants Data Isolation & Sharing Secure Massive Data Migration Data Server Virtualization Security

REFERENCES

[1] http://en.wikipedia.org/wiki/Cloud_storage, 2011.2
[2] Konstantin Shvachko, Hairong Kuang, Sanjay Radia, Robert Chansler. The Hadoop Distributed File System. In Proceedings of the 26th IEEE Symposium on Mass Storage Systems and Technologies, pp:1~10, 3-7 May 2010, Incline Village, NV.
[3]Qingni Shen, Yahui Yang, Zhonghai Wu, Xin Yang, Lizhe Zhang, Xi Yu, Zhenmin Lao, Dandan Wang, Min Long. SAPSC: Security Architecture of Private Storage Cloud Based on HDFS. In Proc. of the 26th IEEE International Conference on Advanced Information Networking and Applications Workshops (WAINA-2012), pp: 1292-1297, Fukuoka, Japan, March 26-29, 2012
[4] Qingni Shen, Xin Yang, Xi Yu, Yahui Yang, Zhonghai Wu. Towards Data Isolation and Collaboration in Storage Cloud. The 2011 IEEE Asia-Pacific Services Computing Conference (APSCC2011), pp: 139-146. December 12-15, 2011, Jeju, Korea.
[5] Qingni Shen, Lizhe Zhang, Xin Yang, Yahui Yang, Zhonghai Wu, Ying Zhang. SecDM: Securing Data Migration between Cloud Storage Systems. In Proceedings of 2011 Ninth IEEE International Conference on Dependable, Autonomic and Secure Computing (CDAS2011), pp: 636-641. December 12-14, Sydney, Australia
[6] Ying Chen, Qingni Shen*, Pengfei Sun, Yangwei Li, Sihan Qing. Reliable Migration Module in Trusted Cloud based on Security Label-Design and Implementation. In Proc. of the 26th IEEE International Parallel & Distributed Processing Symposium Workshops (IPDPS 2012). pp: 2230-2236, May 21-25, 2012, Shanghai, China
[7] Qingni Shen, Lizhe Zhang, Xin Yang, Yahui Yang, Zhonghai Wu, Ying Zhang. SecDM: Securing Data Migration between Cloud Storage Systems. In Proceedings of 2011 Ninth IEEE International Conference on Dependable, Autonomic and Secure Computing (CDAS2011), pp: 636-641. December 12-14, Sydney, Australia.