ISSN:2249-5789

a file uploaded by a user. But the number of *Data Nodes* does less impact on the cost of security decision as each block's uploading is decided by the *Name Node*. And the average cost of this operation is less than 5%. For rebalancing, the migrated data capacity depends on the average usage rate(the used capacity/total capacity in all *Data Nodes)*, and the sum of all *Data Nodes*' excess capacity over the average usage rate is our migrated data capacity. So, the size of data capacity has obvious influence on the cost of migration. And the average cost of rebalancing is about 10% as it computes the migrated data capacity in all *Data Nodes* (see figure 6).

In HDFS, *distcp* is a typical operation for inter-cloud migration. For secure *distcp*, the time cost is mainly from three aspects: SSL negotiation, temporary ticket distribution and verification, file encrypting/decrypting for secure transmission regardless of platform attestation, which is optional for critical data migrating. By testing in our prototype (see figure 7), the temporary ticket distribution and verification only need 0.15-3ms overhead, and SSL negotiation 890ms. The cost of en/decrypting operation is larger than the cost of above two phases, e.g. 290726ms for a 384M big file. But comparing to the confidentiality of data, this cost is worth for most users.

Table 2. Time consumption of up/downloading a file

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

Compared to *HDFS* without a *PDE* service, the time we will used for *PDE* service (when uploading a file) includes the time it costs to generate a symmetric key, the time it costs for encryption of the file upload filter, and the time it costs for encryption of the symmetric key and persistence of the symmetric key. Especially, the time cost for encrypting the file upload filter is offset by the time cost for uploading the file. *Table 2* shows the time we cost when we upload various kinds of data. With the PDE service comparing the uploading without a PDE service. *Table 2* shows the time we cost when we download files of various amounts. As shown in the tables, the average overhead of PDE service is about 13%, and this is acceptable for the private data protection.

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. The security solutions based on the HDFS layer, with master/slave architecture, for the PSC are proposed. And an implementation of these security services is given with AOP method. The performance analysis of them proves the efficiency of the security design. In future, we will make our security services compatible with other cloud storage software systems.

ACKNOWLEDGMENT

This work is supported by IBM SUR Project 2009 and the National Natural Science Foundation of China under Grant No.60873238, 61073156, and 61070237.

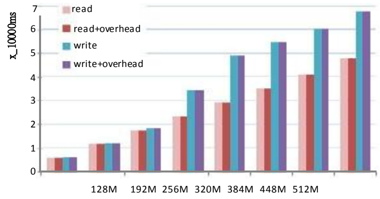

Figure 5. Isolation Control Cost for read and write

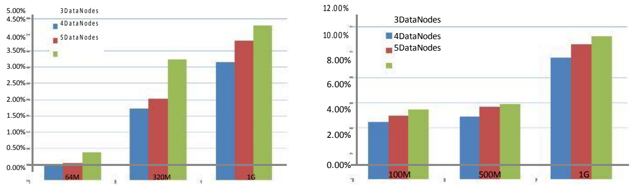

Figure 6. Intra-Cloud Migration Cost for Block Replication(left graph) and Rebalancing(right graph)

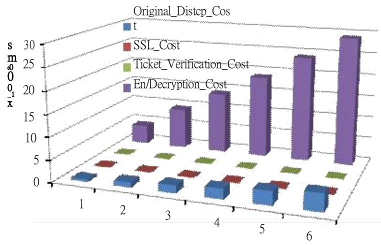

Figure 7. Secure Inter-Cloud Migration Cost for distcp

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] S. Ghemawat, H. Gobioff, S. Leung. “The Google file system,” In Proceedings of ACM Symposium on Operating Systems Principles, Lake George, NY, Oct 2003, pp 29–43.

[4] Parascale Cloud Storage Software - Reference and Whitepapers, 2010. Available at: http://www.hds.com/index.php/technology/how-it-works/73-parascale-cloud-storage/138-parascale-cloud-storage-reference-papers.

[5] Cloudera, http://www.cloudera.com/, 2011

[6] SINA. Cloud Data Management Interface (CDMI) v1.0, 2010. Available at: http://www.snia.org/tech_activities/standards/curr_standards/cdmi/.

[7] Owen O'Malley, Kan Zhang, Sanjay Radia, Ram Marti, and Christopher Harrell. Hadoop Security Design. Technical Report, 2009.10

[8] X. Yang, Q. Shen, Y. Yang, S. Qing, A Way Of Key Management In Cloud Storage Based On Trusted Computing. In Proc of the 8th IFIP International network and parallel computing, pp:135-145, Oct. 2011.

[9] Reiner Sailer, Xiaolan Zhang, Trent Jaeger, and Leendert van Doorn, C.: Design and Implementation of a TCG-based Integrity Measurement Architecture. In: SSYM'04: Proceedings of the 13th conference on USENIX Security Symposium, pages 16–16, Berkeley, CA, USA.

[10] CA, USA, 2004. USENIX Association.Kiczales, G.; Lamping, J; Mehdhekar, A; Maeda, C; Lopes, C. V.; Loingtier, J; Irwin, J. Aspect-Oriented Programming. In proceedings of the European Conference on Object-Oriented Programming (ECOOP), Springer-Verlag LNCS 1241. June 1997.