DEPENDABLE PEER TO PEER PROCESSING OF DATA BY ENABLING SECURED TRANSMISSION

[1]Mohamed Ibrahim Sait, [2]Dr.H.Venugopal, [3]Dr.Siddappa M

CSE, SSIT

Email:[1]Ibrahim.abby@gmail.com, [2]venussit@gmail.com, [3]siddappa.p@gmail.com

***Abstract*—The corporate network is often used for sharing information among the participating companies and facilitating collaboration in a certain industry sector where companies share a common interest. It can effectively help the companies to reduce their operational costs and increase the revenues. However, the inter-company data sharing and processing poses unique challenges to such a data management system including scalability, performance, throughput, and security. In this paper, we present BestPeer++, a system which delivers elastic data sharing services for corporate network applications in the cloud based on BestPeer—a peer-to-peer (P2P) based data management platform. By integrating cloud computing, database, and P2P technologies into one system, BestPeer++ provides an economical, flexible and scalable platform for corporate network applications and delivers data sharing services to participants based on the widely accepted pay-as-you-go business model. We evaluate BestPeer++ on Cloud platform. The benchmarking results show that BestPeer++ achieves near linear scalability for throughput with respect to the number of peer nodes.**

I. INTRODUCTION

Companies of the same industry sector are often connected into a corporate network for collaboration purposes. Each company maintains its own site and selectively shares a portion of its business data with the others. Examples of such corporate networks include supply chain networks where organizations such as suppliers, manufacturers, and retailers collaborate with each other to achieve their very own business goals including planning production-line, making acquisition strategies and choosing marketing solutions.

From a technical perspective, the key for the success of a corporate network is choosing the right data sharing platform, a system which enables the shared data (stored and maintained by different companies) network-wide visible and supports efficient analytical queries over those data. Traditionally, data sharing is achieved by building a centralized data warehouse, which periodically extracts data from the internal production systems (e.g., ERP) of each company for subsequent querying. Unfortunately, such a warehousing solution has some deficiencies in real deployment.

First, the corporate network needs to scale up to support thousands of participants, while the installation of a large-scale centralized data warehouse system entails nontrivial costs including huge hardware/software investments (a.k.a total cost of ownership) and high maintenance cost (a.k.a total cost of operations) [7]. In the real world, most companies are not keen to invest heavily on additional information systems until they can clearly see the potential return on investment (ROI) [8]. Second, companies want to fully customize the access control policy to determine which business partners can see which part of their shared data. Unfortunately, most of the data warehouse solutions fail to offer such flexibilities. Finally, to maximize the revenues, companies often dynamically adjust their business process and may change their business partners. Therefore, the participants may join and leave the corporate networks at will. The data warehouse solution has not been designed to handle such dynamicity.

To address the aforementioned problems, this paper presents BestPeer++, a cloud enabled data sharing platform designed for corporate network applications. By integrating cloud computing, database, and peer-to-peer (P2P) technologies, BestPeer++ achieves its query processing efficiency and is a promising approach for corporate network applications, with the following distinguished features.

- BestPeer++ is deployed as a service in the cloud. To form a corporate network, companies simply register their sites with the BestPeer++ service provider, launch BestPeer++ instances in the cloud and finally export data to those instances for sharing. BestPeer++ adopts the pay-as-you-go business model popularized by cloud computing [9]. The total cost of ownership is therefore substantially reduced since companies do not have to buy any hardware/software in advance. Instead, they pay for what they use in terms of BestPeer++ instance's hours and storage capacity.

- BestPeer++ extends the role-based access control for the inherent distributed environment of corporate networks [4]. Through a web console interface, companies can easily configure their access control policies and prevent undesired business partners to access their shared data.

- BestPeer++ employs P2P technology to retrieve data between business partners. BestPeer++ instances are organized as a structured P2P overlay network named BATON [1]. The data are indexed by the table name, column name and data range for efficient retrieval.

- BestPeer++ employs a hybrid design for achieving high performance query processing [5]. The major workload of a corporate network is simple, low overhead queries. Such queries typically only involve querying a very small number of business partners and can be processed in short time. Best-Peer++ is mainly optimized for these queries. For infrequent time-consuming analytical tasks, we provide an interface for exporting the data from Best-Peer++ to Hadoop and allow users to analyse those data using MapReduce.

In summary, the main contribution of this paper is the design of BestPeer++ system that provides economical, flexible and scalable solutions for corporate network applications. We demonstrate the efficiency of BestPeer++ by benchmarking BestPeer++ against public cloud, over a set of queries designed for data sharing applications. The results show that for simple, low-overhead queries, the performance of BestPeer++ is significantly better than HadoopDB [6].

II. EXISTING SYSTEM

The local administrator at each normal peer can assign the new user with an existing role if the access privilege of that role is applicable to the new user. If none of the existing roles satisfies the new user, the local administrator can create new roles by three operators. Note that Best Peer++ does not collect the information of existing users in the collaborating ERP databases, since it will lead to potential security issues. Instead, the user management module of Best Peer++ provides interfaces for the local administrator at each participating organization to create new accounts for users who desire to access Best Peer++ service.

III. PROPOSED SYSTEM

Specifically, we use the two-tier partial replication strategy to provide both data availability and load balancing, as proposed in our recent study. To enhance the usability of conventional P2P networks, database community have proposed a series of PDBMS [2] (Peer-to-Peer Database Manage System) by integrating the state-of-art database techniques into the P2P systems. We have discussed the unique challenges posed by sharing and processing data in an inter-businesses environment and proposed Best Peer++, a system which delivers elastic data sharing services [3], by integrating cloud computing, database, and peer-to-peer technologies.

IV. ENHANCEMENT

Cloud computing provides a scalability environment for growing amounts of data in different Peers and processes that work on various applications and services by means of on-demand self-service. One of the strength of cloud computing is that data are being centralized and outsourced in clouds. This kind of outsourced storage in clouds has become a new profit growth point by providing a comparably low-cost, scalable, location independent platform for managing clients' data. The cloud storage service (CSS) relieves the burden for storage management and maintenance.

$$Cbasic = (\alpha+\beta)\ N + \gamma t$$

However, if such an important service is vulnerable to attacks or failures, it would bring irretrievable losses to the clients since their data or archives are stored in an uncertain storage pool outside the enterprises. These security risks come from the following reasons: the cloud infrastructures are much more powerful and reliable than personal computing devices.

$$Cbasic = (\alpha+\beta)\ N + \gamma\frac{N}{\theta}$$

However, they are still facing all kinds of internal and external threats; for the benefits of their possession, there exist various motivations for cloud service providers (CSP) to behave unfaithfully towards the cloud users; furthermore, the dispute occasionally suffers from a lack of trust on CSP. Consequently, their behaviors may not be known by the cloud users, even if this dispute may result from the users' own improper operations. Therefore, it is necessary for cloud service providers to offer an efficient audit service to check the integrity and availability of the stored data.

$$W(i) = \tau(Ti) * \sigma(i+1)$$

In this paper, we introduce a dynamic audit service for integrity verification of un-trusted and outsourced storages. Our audit system, based on a novel audit system architecture, can support dynamic data operations and timely abnormal detection with the help of several effective techniques, such as fragment structure, random sampling, and index-hash table.

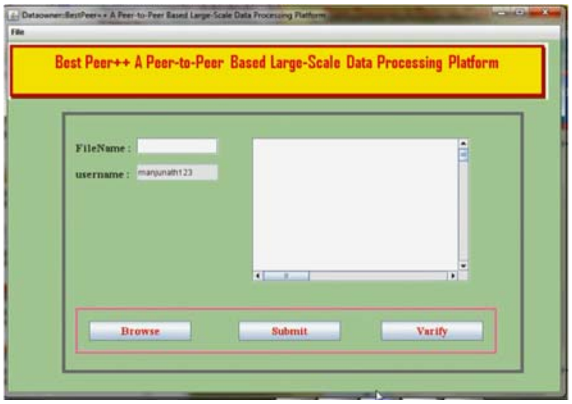
$$\sigma(i) = \sigma(i + 1) * S(Ti) *g(i)$$

Furthermore, we propose an efficient approach based on probabilistic query and periodic verification for improving the performance of audit services. A proof of- concept prototype is also implemented to evaluate the feasibility and viability of our proposed approaches. Our experimental results not only validate the effectiveness of our approaches, but also show our system has a lower computation cost, as well as a shorter extra storage for integrity verification.

V. IMPLEMENTATION

A. Data Owner

In this module, the data owner has to register in a cloud server and fire walls (fire wall1 and fire wall2), after registration he has to login. Data owner uploads their data file into the cloud server and the cloud server will connect to fire wall, the data file is stored in a normal peer and the backup file will be stored in Bootstrap peer. The Data owner can have capable of manipulating the stored data file. If any attacker is modify data file in a cloud server, then Data owner will verify the data file.

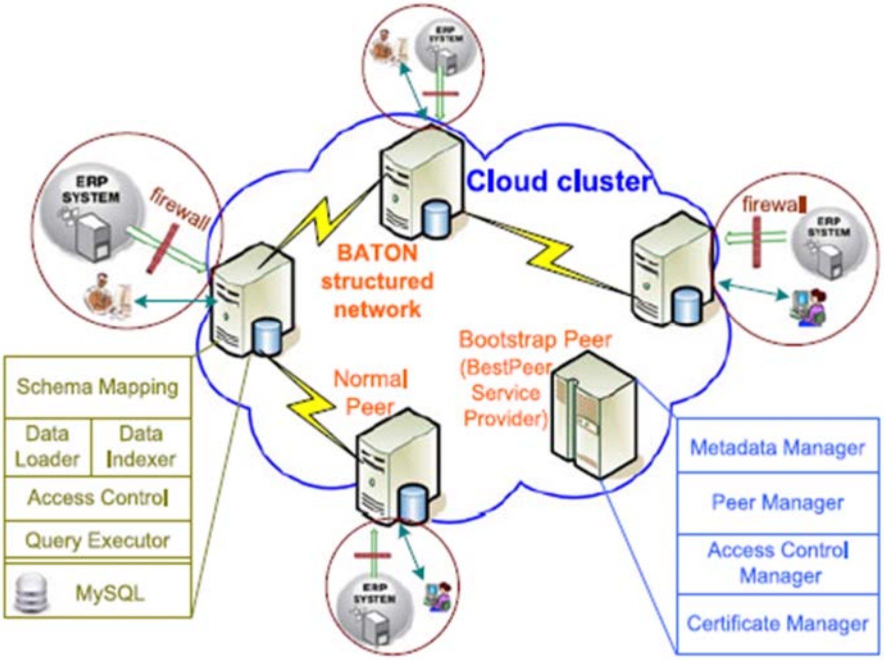


Dataowner:BestPeer++ A Peer-to-Peer Based Large-Scale Data Processing Platform

Best Peer++ A Peer-to-Peer Based Large-Scale Data Processing Platform

FileName :

username : manjunath123

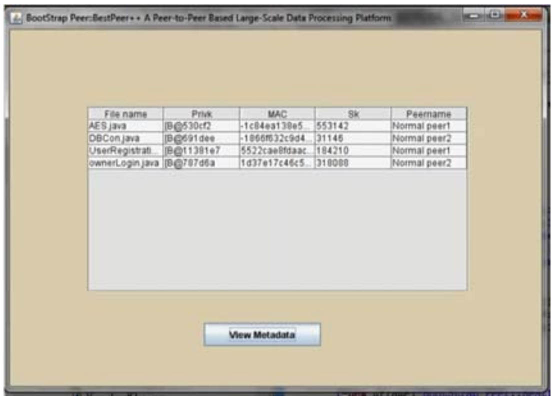


B. Cloud Cluster

The cloud cluster consists of fire walls (firewall1 and firewall2) and peers (normal peer1, normal peer2 and Bootstrap peer).The cloud server is responsible for data storage and file authorization for an end user. The normal peer software consists of five components: schema mapping, data loader, data indexer, access control, and query executor. The data file will stored in normal peer with their tags such as file name, secret key, digital sign, and owner name. If the end user requested file is correct then the data will be sent to the corresponding user and also will check the file name, end user name and secret key. If all are true then it will send to the corresponding user or he will be captured as attacker. The peer which is not having more count in attacker list is called as best peer. And also the best peer will become per++ when it is not in the attacker list.

C. Bootstrap Peer

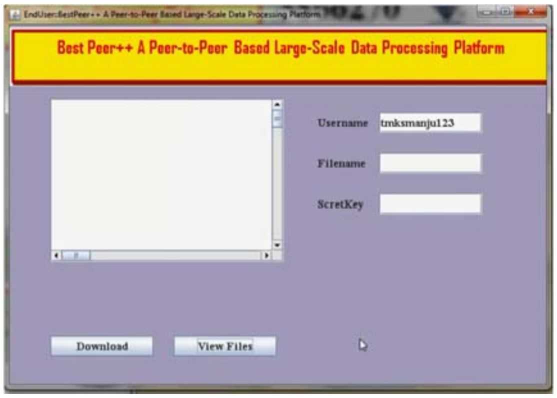
In this module, the bootstrap peer is the entry point of the whole network. It has several responsibilities. First, the bootstrap peer serves for various administration purposes; including monitoring and managing normal peers and also scheduling various network management events. Second, the bootstrap peer acts as a central repository for meta data of corporate network applications, including shared global schema, participant normal peer list, and role definitions. In addition, Bootstrap Peer employs the standard PKI encryption scheme to encrypt/decrypt data transmitted between normal peers in order to further increase the security of the system.


BootStrap Peer:BestPeer++ A Peer-to-Peer Based Large-Scale Data Processing Platform
View Metadata

File name	Privk	MAC	Sk	Peername
AES.java	[B@530cf2	-1c84ea138e5...	553142	Normal peer1
DBCon.java	[B@691dee	-1866f632c9d4...	31146	Normal peer2
UserRegistrati...	[B@11381e7	5522cae8fdaac...	184210	Normal peer1
ownerLogin.java	[B@787d6a	1d37e17c46c5...	318088	Normal peer2

D. Data Consumer(End User)

The data consumer is nothing but the end user who will request and gets file contents response from the corresponding cloud servers and fire walls. End user should register before downloading any files from the cloud server. After registration he should login. If the file name and secret key is correct then the end user is getting the file response from the cloud or else he will be considered as an attacker and also he will be blocked in corresponding cloud and fire wall. If he wants to access the file after blocking he wants to UN block from the cloud.


EndUser:BestPeer++ A Peer-to-Peer Based Large-Scale Data Processing Platform

Best Peer++ A Peer-to-Peer Based Large-Scale Data Processing Platform

Username tmksmanju123

Filename

ScretKey

E. Attacker

If user is entered a wrong secrete key, then considered as an attacker. Attacker is one who is integrating the cloud file by adding malicious data to the corresponding cloud. The peer which is not having more count in attacker list is called as best peer. And also the best peer will become per++ when it is not in the attacker list.

F. Data Integrity

The data owner performs a dynamic audit service for verifying the integrity of un-trusted and outsourced storage. The data owner audit service, constructed based on the techniques, fragment structure, random sampling and index-hash table, can support provable updates to outsourced data, and timely abnormal detection. In addition, we propose an efficient approach based on probabilistic query and periodic verification for improving the performance of audit services.

VI. CONCLUSION

In the enhancement work, the system has developed to facilitate the client in getting a proof of integrity of the data which he wishes to store in the cloud storage servers with bare minimum costs and efforts. Our scheme was developed to reduce the computational and storage overhead of the client as well as to minimize the computational overhead of the cloud storage server. We also minimized the size of the proof of data integrity so as to reduce the network bandwidth consumption. At the client we only store two functions, the bit generator function g, and the function h which is used for encrypting the data. Hence the storage at the client is very much minimal compared to all other schemes that were developed. Hence this scheme proves advantageous to thin clients like PDAs and mobile phones. The operation of encryption of data generally consumes a large computational power. In our scheme the encrypting process is very much limited to only a fraction of the whole data thereby saving on the computational time of the client.

REFERENCES

[1] H.V. Jagadish, B.C. Ooi, and Q.H. Vu, *“BATON: A Balanced Tree Structure for Peer-to-Peer Networks”,* Proc. 31st Int’l Conf. Very Large Data Bases (VLDB ’05), pp. 661-672, 005.

[2] W.S. Ng, B.C. Ooi, K.-L. Tan, and A. Zhou, *“PeerDB: A P2P-Based System for Distributed Data Sharing,”* Proc. 19th Int’l Conf. Data Eng., pp. 633-644, 2003.

[3] H.T. Vo, C. Chen, and B.C. Ooi, *“Towards Elastic Transactional Cloud Storage with Range Query Support,”* Proc. VLDB Endowment, vol. 3, no. 1, pp. 506-517, 2010.

[4] S. Wu, S. Jiang, B.C. Ooi, and K.-L. Tan, *“Distributed Online Aggregation,”* Proc. VLDB Endowment, vol. 2, no. 1, pp. 443-454, 2009.

[5] S. Wu, J. Li, B.C. Ooi, and K.-L. Tan, *“Just-in-Time Query Retrieval over Partially Indexed Data on Structured P2P Overlays,”* Proc. ACM SIGMOD Int’l Conf. Management of Data (SIGMOD ’08), pp. 279-290, 2008.

[6] A. Abouzeid, K. Bajda-Pawlikowski, D.J. Abadi, A. Rasin, and A.Silberschatz, *“HadoopDB: An Architectural Hybrid of MapReduce and DBMS Technologies for Analytical Workloads,”*1 Proc. VLDB Endowment, vol. 2, no. 1, pp. 922-933, 2009.

[7] Oracle Inc., *“Achieving the Cloud Computing Vision,”* White Paper, 2010.

[8] Saepio Technologies Inc., *“The Enterprise Marketing Management Strategy Guide,”* White Paper, 2010.

[9] Google Inc., “Cloud Computing-What is its Potential Value for Your Company?” White Paper, 2010.