Scientific Journal of Impact Factor (SJIF):

e-ISSN (O): 2348-4470
p-ISSN (P): 2348-6406

International Journal of Advance Engineering and Research Development

Volume 4, Issue 12, December -2017|

A Novel algorithm for improving Security and Storage for Online Social Networks on Geo-Distributed Clouds

[1]Suchita Choudhary, [2]Prof. Vikas N. Nandgaonkar, [3]Prof. Ashvini Jadhav

[1]Department of Computer Engineering, NMIET, SPPU, Pune
[2]Department of Computer Engineering, NMIET, SPPU, Pune
[3]Department of Computer Engineering, NMIET, SPPU, Pune

Abstract—*Geo-distributed clouds provide an excellent platform to deploy online social network (OSN) services. To take advantage of clouds, a major concern of OSN providers is optimizing the cost in case with money spent in using cloud resources while considering other important requirements, which includes providing quality of service (QoS), data availability and security to OSN users. In this paper, we will study the problem of cost optimization for the dynamic OSN on multiple geo-distributed clouds over consecutive time periods while also meeting the security using the heuristic algorithm. We model the cost, the QoS, as well as the data availability of the OSN, formulate the problem, and design an algorithm named heuristic for secure cloud storage.*

Index Terms—*Cloud computing,cost optimization models and methods,online social network, performance analysis and evaluation, security.*

I. INTRODUCTION

Internet services today are experiencing two remarkable changes. One is the popularity of online social networks (OSNs), where users build social relationships and share contents with one another. The other is the rise of clouds for storage. Often going through multiple geographical locations.

- Clouds provide an important platform for deploying distributed online services.
- OSN services have a very large user base and need to meet demands of users worldwide.
- Infrastructure-as-a-Service can match this need and provide tremendous resource and cost efficiency advantages.

Existing work on OSN service provide either least cost in a single site without the QoS concern as in the geo-distribution case or aims for least inter-data-center traffic in the case of multiple data centers without considering other dimensions of the service, e.g., data availability. More importantly, the models in all such work do not capture the cost of resource usage and thus cannot fit the cloud scenario. There are some works on cloud-based social video focusing on online social relationships to improve video distribution, which is only one of the many factors of OSN services; most optimization research on multicloud and multi-data-center services is not for OSN. They fail to capture the OSN features such as social relationships and user interactions, and thus their models are not applicable to OSN services.

In this paper, we will study the problem of optimizing the monetary cost of the dynamic, multicloud-based OSN while ensuring its Security and cost optimization using the heuristic algorithm.
Firstly model the cost, the QoS, and the data availability of the OSN service upon clouds. Cost model identifies different types of costs associated with multicloud OSN. Guided by existing research on OSN growth and analysis of real-world OSN dynamics, our model approximates the total cost of OSN over consecutive time periods when the OSN is large in user population but moderate in growth, enabling us to achieve the optimization of the total cost for Security mainly, Qos and data availability.

Migrating OSN services toward geographically distributed clouds must have the needs from several different aspects. First, OSN providers want to optimize the monetary cost spent in using cloud resources. For instance, they may wish to minimize the storage cost when replicating users' data at more than one cloud, or minimize the intercloud communication cost when users at one cloud have to request the data of others that are hosted at a different cloud. Moreover, OSN providers to provide OSN users with satisfactory quality of service (QoS). To this end, they may want a user's data and those of her friends to be accessible from the cloud to the user, for example. Last but not least, OSN providers may also be concerned with data availability, e.g., ensuring the number of users' data replicas to be no fewer than a specified threshold across clouds. Addressing all such needs of cost, QoS, and data availability is further complicated by the fact that an OSN continuously experiences dynamics, e.g., new users join, old users leave, and the social relations also vary.

Targeting the OSN service over multiple clouds, we begin with identifying the types of costs related to cloud resource utilization: the storage cost for storing users' data, the intercloud traffic cost for synchronizing data replicas across clouds, the redistribution cost incurred by the cost optimization mechanism itsel. As discussed and approximate the total cost of the multicloud OSN over time. Afterwards, proposed a vector model to capture the QoS of the OSN service, show the features of this model, and demonstrate its usage. Finally, model the OSN data availability by linking it with the number of each user's data replicas

A. Modelling the Storage and the Inter cloud Traffic Cost

OSN is commonly abstracted as a social graph, where each vertex represents a user and each edge represents a social relation between two users.

1) A user has a storage cost, which is the monetary cost for storing one replica of her data (e.g., profile, statuses) in the cloud for one billing period.

2) Similarly, a user has a traffic cost, which is the monetary cost during a billing period because of the inter cloud traffic. 3) A user has a sorted list of clouds for the purpose of QoS. 4) A user has a security cost, which is the monetary cost during a billing period because of the data stored on cloud should be secured and data integrity should be maintained.

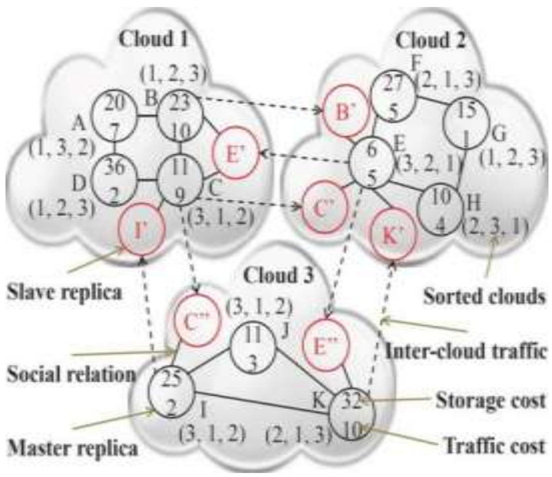

Fig. 1. Storage and intercloud traffic cost.

Fig. 1 is an example where 11 users are hosted by three clouds. Black circles represent each user's master replica, and red ones represent the slave replicas of neighbors to ensure social locality. Solid lines are social relations, and dotted arrows are the traffic. Within each black circle, the value on the top is the storage cost of a user, and the value at the bottom is the traffic cost. For Fig. 1, the total storage cost is 330, and the total intercloud traffic cost is 50.

1) *Approximating the Total Cost:* Consider the social graph in a billing period. As it may vary within the period, the final of the social graph in this period as , and the initial of the social graph at the beginning of this period as G=(V,E).
2) *Comparing QoS:* There can be different data placements upon clouds. Each may result in a different corresponding QoS vector. For two QoS vectors and representing two qa and qb placements respectively, we observe that the former placement provides QoS no better than the latter, i.e.qa¡=qb.
3) *Data Availability Requirement:* An OSN provider specifies the data availability requirement by indicating the minimum number of every user's slave replicas.
4) *Security:* Where the data is stored on cloud should be stored in encrypted format, while accessing the data a key will be generated using algorithm and sent to authorized user who can decrypt data at receiver side. This will be done by using algorithm for maintaining security as well as cost optimization for maintaining security.

II. REVIEW OF LITERATURE

We compared different security algorithms like RSA, SHA, AES, etc for how to make cloud storage secure, but all this algrithm were increasing the cost for online social geo distributed cloud which was not suitable and efficient for the cloud storage, so we decided to use the heuristic algorithm for security which chooses best possible solution from many.

A. Data Preparation: We gathered different security relatedalgorithms, then we studied for each algorithm how much bits are used then how will it effect on cost for cloud storage.

B. Heuristic algorithm can be used in security protocols andprovide interesting cases of how the effiency is increased for the cloud storage also the dats integrity is maintained.Its also observed that it will cut all the possible attacks on the cloud storage with respect to security.

Our approach seems to be rather promising, as this research is in an initial stage, it can be extended in many directions.

III. SYSTEM ARCHITECTURE / SYSTEM OVERVIEW

A Resource Allocation Methodology By Which The Resource Scheduling Becomes Optimum And Enhancing The Computing Experience. The multiplicity revealing and enhancement revealing operators are used in the proposed algorithm to determine which low-level heuristic is to be used in finding better solutions for job scheduling Hyper-heuristics aim to discover some algorithms that are capable of solving a whole range of problems, with little or non-direct human control. Heuristic techniques are often referred to as search algorithms. The problems are solved by discovering a solution from the set of all possible solutions for a given problem, which is regarded as the search space. Non-deterministic search methods such as local search methods, evolutionary algorithms, simulated annealing, and other search algorithms offer an alternative approach to an exhaustive search to solve complicated computational problems within a sensible amount of time. These methods guarantee for finding a solution at any time, but it may not be optimum.:

The heuristic function is defined on the security formula expressed. The heuristic is efficient as it will not only guide the searching algorithm towards promising regions of the graph but can also prune those parts of the state space where attack cannot happen under a given security formula. The heuristic is defined in terms of two mutually recursive functions Hs and Ht which assign weights to states and transitions respectively. The state space is obtained according to the semantics of cIP defined in.

A state consists of a tuple (C, x, k) where,

1. C is a context containing principal instances which joinedthe session,
2. x is a mapping of variables to messages, and3. k is a set of messages representing the intruder knowledge. Role-Swap Versus Master Migrations: It appears that roleswaps limit the solution to the initial placement, i.e., a role swap does not seem to be able to migrate a user's data to a cloud that has none of her replicas in the initial placement. We will demonstrate and explain in the following that: 1) allowing such master migrations does not help much in reducing the cost; and 2) only role-swaps can move a considerable amount of users to the clouds that do not host their replicas in the initial placement.

Optimizing Multicloud Services: The work most related to OSN services may be those on social media that leverage online social relationships to improve media delivery. Finds out the best data center for each data item based on access interdependencies, the identity, and time stamp of data access, while balancing storage capacity across data centers; proposes selective replication at a per record granularity to minimize replication overhead and forwarding bandwidth while respecting policy constraints. A substantial body of literature studies cloud resource pricing and allocation, request mapping, and content routing in the multicloud or multi-data-center scenario. Although our work also focuses on multicloud services, OSN is unique in data access patterns (i.e., social locality), making this group of existing work inapplicable to our scenario.

IV. SYSTEM ANALYSIS

When calculating the costs, we assume that all clouds have the same billing prices. In reality, resource usage of clouds from different providers or at different locations may be charged at different prices. Such cases can be easily addressed, and our proposed algorithm, as shown later, can also straightforwardly adapt to these cases. We also assume that each cloud can provide infinite resources on demand to an OSN service provider, a guarantee often provided by a cloud provider to its customers.

An important part of our cost model is the cost incurred by the optimization mechanism itself, which we call the redistribution cost. We generally observe that an optimization mechanism is to optimize the cost by moving data across clouds to optimum locations, thus increasing such cost. The redistribution cost is essentially the inter cloud traffic cost, but in this paper we use the term inter cloud traffic to specifically refer to the inter cloud write traffic for maintaining replica consistency, and treat the redistribution cost separately and to maintain security as well.

The above fig 2. is almost same as the previous fig 1., but here it shows the security will be implemented and the cost as well be also optimized for online geo-distributed cloud storage.Along with other cost opTimization features of QOS, Data availability also the security will be implemented for the same and the cost will be optimized.

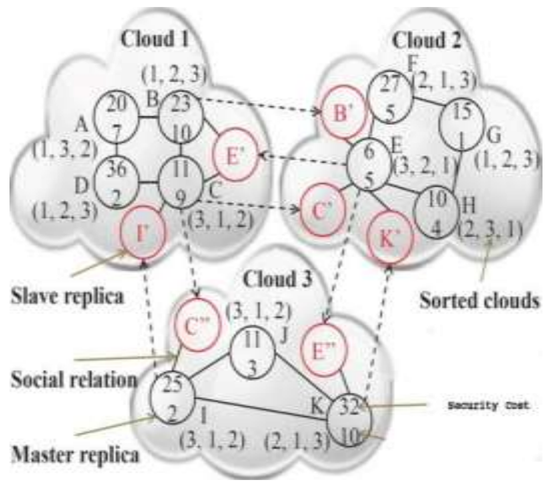

Fig. 2. Storage Security for cloud storage.

Data Availability: Whether to remove a slave or not does not only depend on social locality, but also on the data availability requirement. Creating slaves is always fine because it never violates the data availability requirement.

Optimizing OSN Services: For OSN at a single site, using distributed method to partition the data across servers potentially leads to poor performance. Recent work proposes maintaining social locality to address this issue: minimizes the total number of slave replicas while maintaining social locality for every user; maximizes the number of users whose social locality can be maintained, given a fixed number of replicas per user. For OSN across multiple sites, some propose selective replication of data across data centers to reduce the total inter-data-center traffic, and others propose a framework that captures and optimizes multiple dimensions of the OSN system objectives simultaneously. The work does not have the concern of QoS as in our geo-distribution case. Besides, the cost models in all the aforementioned existing work, except, do not capture the monetary expense and cannot fit the cloud scenario, while it do not explore social locality to optimize the multi-data-center OSN service.

V. CONCLUSION

In this paper, we will study the problem of optimizing the monetary cost spent on cloud resources when deploying an online social network service over multiple geodistributed clouds. We model the cost of OSN data storage on cloud,quality of service with and address OSN data availability by ensuring a minimum number of replicas for each user. Based on these models, we present the optimization problem of minimizing the total cost while ensuring the QoS and the data availability. We propose as our heuristic algorithm. Which will show that this algorithm will be implemented for maintaining security on cloud and to avoid monetary cost i.e optimize the cost in case of security

REFERENCES

[1] S. Agarwal et al., Volley: Automated data placement for geo-distributed cloud services, in Proc. NSDI, 2010, p. 2.

[2] . Benevenuto, T. Rodrigues, M. Cha, and V. Almeida, Characterizing user behavior in online social networks, in Proc. IMC, 2009, pp. 4962.

[3] H. Chun et al., Comparison of online social relations in volume vs interaction: a case study of cyworld, in Proc. IMC, 2008, pp. 5770. [14] H. Hu and X. Wang, Evolution of a large online social network, Phy

[4] L. Jiao, J. Li, W. Du, and X. Fu, Multi-objective data placement for multi-cloud socially aware services, in Proc. IEEE INFOCOM, 2014, pp. 2836.

[5] G. Liu, H. Shen, and H. Chandler, Selective data replication for online social networks with distributed datacenters, in Proc. IEEE ICNP, 2013, pp. 110.

[6] J. M. Pujol et al., The little engine(s) that could: Scaling online social networks, IEEE/ACM Trans. Netw., vol. 20, no. 4, pp. 11621175, Aug. 2012.

[7] D. A. Tran, K. Nguyen, and C. Pham, S-clone: Socially-aware data replication for social networks, Comput. Netw., vol. 56, no. 7, pp. 20012013, 2012.

[8] Z. Wang et al., Propagation-based social-aware replication for social video contents, in Proc. ACM Multimedia, 2012, pp. 2938.

[9] Y. Yang, Q. Chen, and W. Liu, The structural evolution of an online discussion network, Physica A, Statist. Mech. Appl., vol. 389, no. 24, pp. 58715877, 2010.

[10] L. Jiao, J. Li, T. Xu, and X. Fu, Cost optimization for online social networks on geo-distributed clouds, in Proc. IEEE ICNP, 2012, pp. 110. [11] HubSpot, Inc., Cambridge, MA, USA, State of the Twittersphere, 2010.

[12] Microsoft, Redmond, WA, USA, Pricing overview: Microsoft Azure, [Online]. Available: http://www.windowsazure.com/en-us/ pricing/details/

[13] A. L. Barabasi, The origin of bursts and heavy tails in human dynamics, Nature, vol. 435, no. 7039, pp. 207211, 2005.

[14] H. Hu and X. Wang, Evolution of a large online social network, Phys.

[15] Lett. A, vol. 373, no. 1213, pp. 11051110, 2009. A. Vzquez et al., Modeling bursts and heavy tails in human dynamics, Phys. Rev. E, vol. 73, no. 3, p. 036127, 2006

Author Biography

Name: **Suchita Zope**, Completed BE in Information technology, currently pursuing ME in Computer Networks.
Name: **Prof. Vikas N. Nandgaonkar**, Completed his UG and PG in Computer Engineering from Pune University, Pune, MH, India. He is working as dean Research and Development (R&D) in the Department of Computer Engineering in NMIET, University of Pune, India, He is also a life member of professional body ISTE. His research interest includes Adhoc Networks and Security.

