



OPTIMIZING COST FOR ONLINE SOCIAL NETWORKS ON GEO-DISTRIBUTED CLOUDS

CH. SAI VIJAYA, VANAM GOPINATH

*PG SCHOLAR, **ASSOCIATE PROFESSOR.

DEPARTMENT OF CSE, SLC'S INSTITUTE OF ENGINEERING AND TECHNOLOGY, HYDERABAD

ABSTRACT

Geo-distributed clouds provide an intriguing platform to deploy online social network (OSN) services. To leverage the potential of clouds, a major concern of OSN providers is optimizing the monetary cost spent in using cloud resources while considering other important requirements, including providing satisfactory quality of service (QoS) and data availability to OSN users. In this paper, we study the problem of cost optimization for the dynamic OSN on multiple geo-distributed clouds over consecutive time periods while meeting predefined QoS and data availability requirements. We model the cost, the QoS, as well as the data availability of the OSN, formulate the problem, and design an algorithm named . We carry out extensive experiments with a large-scale real-world Twitter trace over 10 geo-distributed clouds all across the US. Our results show that, while always ensuring the QoS and the data availability as required, can reduce much more one-time cost than the state-of-the-art methods, and it can also significantly reduce the accumulative cost when continuously evaluated over 48 months, with OSN dynamics comparable to real-world cases

I. INTRODUCTION

INTERNET services today are experiencing two remarkable changes. One is the unprecedented popularity of online social networks (OSNs), where users build social relationships and create and share contents with one another. The other is the rise of clouds. Often spanning multiple geographic locations, clouds provide an important platform for deploying distributed online services. Interestingly, these two changes tend to be combined. While OSN services often have a very large user base and need to scale to meet demands of users worldwide, geodistributed clouds that provide Infrastructure-as-a-Service can match this need seamlessly and provide tremendous resource and cost efficiency advantages. Migrating OSN services toward geographically



distributed clouds must reconcile 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 inter-cloud communication cost when users at one cloud have to request the data of others that are hosted at a different cloud. Moreover, OSN providers hope to provide OSN users with satisfactory quality of service (QoS). 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 Abstract: Geo-distributed clouds provide an intriguing platform to deploy online social network (OSN) services. To leverage the potential of clouds, a major concern of OSN providers is optimizing the monetary cost spent in using cloud resources while considering other important requirements, including providing satisfactory quality of service (QoS) and data availability to OSN users. In this paper, we study the problem of cost optimization for the dynamic OSN on multiple geo-distributed clouds over consecutive time periods while meeting predefined QoS and data availability requirements. We model the cost, the QoS, as well as the data availability of the OSN, formulate the problem, and design an algorithm named . We carry out extensive experiments with a large-scale real-world Twitter trace over 10 geo-distributed clouds all across the US. Our results show that, while always ensuring the QoS and the data availability as required, can reduce much more one-time cost than the state-of-the-art methods, and it can also significantly reduce the accumulative cost when continuously evaluated over 48 months, with OSN dynamics comparable to real-world cases. experiences dynamics, e.g., new users join, old users leave, and the social relations also vary. Existing work on OSN service provisioning either pursues 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. The optimizing the monetary cost of the dynamic, multicloud-based OSN while ensuring its QoS and data availability. The QoS of the OSN service is better if more users have their data hosted on clouds of a higher preference. Our data availability model relates with the minimum number of replicas maintained by each OSN user. Algorithm named cosplay based on



our observations that swapping the roles (i.e., master or slave) of a user's data replicas on different clouds can not only lead to possible cost reduction, but also serve as an elegant approach to ensuring QoS and maintaining data availability. Compared to existing approaches, cosplay reduces cost significantly and finds a substantially good solution of the cost optimization problem, while guaranteeing all requirements are satisfied.

II. SYSTEM ARCHITECTURE Clouds and OSN users are all geographically distributed. Without loss of generality, we consider the singlemaster–multi-slave paradigm. Each user has only one master replica and several slave replicas of her data, where each replica is hosted at a different cloud. When signing in to the OSN service, a user always connects to her master cloud, i.e., the cloud that hosts her master replica, and every read or write operation conducted by a user goes to her master cloud first. Observing that most activities of an OSN user happen between the user and her neighbors (e.g., friends on Facebook or followers on Twitter), this scheme requires that a user's master cloud host a replica (either the master or a slave) of every neighbor of the user. This way, every user can read the data of her friends and her own from a single cloud, and the intercloud traffic only involves the write traffic for maintaining the consistency among a user's replicas at different clouds. Cloud computing is the use of computing resources (hardware and software) that are delivered as a service over a network (typically the Internet). Cloud computing consists of hardware and software resources made available on the Internet as managed third-party services. These services typically provide access to advanced software applications and high-end networks of server computers.

III. PROBLEM DESCRIPTION

Existing work on OSN service provisioning either pursues least 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 monetary cost of resource usage and thus cannot fit the cloud scenario. There are some works on cloud-based social video, focusing on leveraging online social relationships to improve video distribution,



which is only one of the many facets of OSN services; most optimization research on multicloud and multi-data services is not for OSN. SPAR minimizes the total number of slave replicas while maintaining social locality for every user; S International Journal of Engineering and Techniques Existing work on OSN service provisioning either pursues least cost in a single site without the QoS concern as distribution case or aims for center traffic in the case of multiple data centers without considering other dimensions of the service, e.g., data availability. More e models in all such work do not capture the monetary cost of resource usage and thus cannot fit the There are some works on based social video, focusing on leveraging online social relationships to improve video distribution, which is only one of the many facets of OSN services; most optimization research on data-center services is not for OSN. SPAR minimizes the total number of slave replicas while maintaining social locality for every user; S-CLONE 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. PNUTS proposes selective replication at a per record granularity to minimize replication overhead and forwarding bandwidth while respecting policy constraints.

III PROPOSED SYSTEM: The problem of optimizing the monetary cost of the dynamic, multi cloud-based OSN while ensuring its QoS and data availability. We first model the cost, the QoS, and the data availability of the OSN service upon – Feb 2017 Page 64 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 center traffic, and others use a framework that captures and optimizes multiple dimensions of the OSN system objectives PNUTS proposes selective replication at a per record granularity to minimize replication overhead and forwarding bandwidth constraints. They fail to capture the OSN features such as social relationships and user interactions, and thus their models are not applicable to OSN The cost models in all the aforementioned existing work, do he monetary expense d cannot fit the cloud scenario. o not explore social locality to -data-center OSN is



unique in data access terms (i.e., social locality). The problem of optimizing the monetary cost of the dynamic, multi-based OSN while ensuring its QoS and data availability. We first model the cost, the QoS, and the data availability of the OSN service upon. Our cost model identifies different types of costs associated with multcloud OSN while capturing social locality, an important feature of the OSN service that most activities of a user occur between herself and her neighbors. Guided by existing research on OSN growth and our analysis of realworld 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 by independently optimizing the cost of each period. Our QoS model links the QoS with OSN users' data locations among clouds. For every user, all clouds available are sorted in terms of a certain quality metric (e.g., access latency); therefore, every user can have the most preferred cloud, the second most preferred cloud, etc. The QoS of the OSN service is better if more users have their data hosted on clouds of a higher preference. Our data availability model relates with the minimum number of replicas maintained by each OSN user. We then formulate the cost optimization problem that considers QoS and data availability requirements. This problem is NP-hard. We propose a heuristic algorithm named based on our observations that swapping the roles (i.e., master or slave) of a user's data replicas on different clouds can not only lead to possible cost reduction, but also serve as an elegant approach to ensuring QoS and maintaining data availability.

IV. METHODOLOGY

1. OSN System Construction Module
2. Modeling the Storage and the Intercloud Traffic Cost
3. Modeling the Redistribution

4.1 OSN System Construction Module : The Online Social Networking (OSN) system module. We build up the system with the feature of Online Social Networking. Where, this module is used for new user registrations and after registrations the users can login with their authentication. Where after the existing users can send messages to privately and publicly,



options are built. Users can also share post with others. The user can able to search the other user profiles and public posts. In this module users can also accept and send friend requests. With all the basic feature of Online Social Networking System modules is build up in the initial module, to prove and evaluate our system features. Clouds and OSN users are all geographically distributed. we consider the singlemaster–multi-slave paradigm.

4.2 Modeling the Storage and the Intercloud Traffic Cost: The Storage and intercloud Traffic Cost of OSN, which is commonly abstracted as a social graph, where each vertex represents a user and each edge represents a social relation between two users. In this module we calculate the Storage Cost. 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. Similarly, a user has a traffic cost, which is the monetary cost during a billing period because of the intercloud traffic (comments). Intercloud traffic, no matter read or write, as it is free of charge. A user has a sorted list of clouds for the purpose of QoS.

4.3 Modeling the Redistribution Cost: An important part of our cost model is the cost incurred by the optimization mechanism itself, which we call the redistribution cost. We generally envisage that an optimization mechanism is devised to optimize the cost by moving data across clouds to optimum locations, thus incurring such cost. The redistribution cost is essentially the inter-cloud traffic cost, for maintaining replica consistency, and treat the redistribution cost separately.

4.4 Approximating the Total Cost: Consider the social graph in a billing period. As it may vary within the period, we denote the final steady snapshot of the social graph in this period, and the initial snapshot of the social graph at the beginning of this period. The storage cost in is for storing users' data replicas, including the data replicas of existing users and of those who just join the service in this period. The intercloud traffic cost in is for propagating all users' writes to maintain replica consistency. The redistribution cost is the cost of moving data across clouds for optimization; it is only incurred at the beginning of a period.



CONCLUSION The problem of optimizing the monetary resources when deploying an online social network service over multiple geo-distributed clouds. The cost of OSN data placement, quantify the OSN quality of service with our vector approach, 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. By extensive evaluations with large-scale Twitter data is verified to incur substantial cost reductions over existing, state-of-the-art approaches. It is also characterized by significant one-time and accumulated cost reductions over 48 months such that the QoS and the data availability always meets predefined requirements.

REFERENCES

1. Lei Jiao, Jun Li, Senior Member, IEEE, ACM, Tianyi Xu, Wei Du, and Xiaoming Fu, Senior Member, IEEE, “Optimizing Cost for Online Social Networks on Geo-Distributed Clouds”, IEEE/ACM TRANSACTIONS ON NETWORKING, VOL. 24, NO. 1, FEBRUARY 2016.
2. Amazon, Seattle, WA, USA, “Case studies,” [Online]. Available:<http://aws.amazon.com/solutions/case-studies>.
3. Microsoft, Redmond, WA, USA, “Pricing overview: Microsoft.
4. Azure,”[Online].Available:<http://www.windowsazure.com/enus/pricing/details/>
5. Facebook, Menlo Park, CA, USA, “Facebook Newsroom Available: GitHub, “Twitter/Gizzard,” Available:<http://github.com/twitter/gizzard>
6. HubSpot, Inc., Cambridge, MA, USA, “State of the Twittersphere,” 2010.
7. A. Abou-Rjeili and G. Karypis, “Multilevel algorithms for partitioning power-law graphs,” in Proc. IPDPS, 2006, pp. 1–10.



8. S. Agarwal et al., “Volley: Automated data placement for geo-distributed cloud services,” in Proc. NSDI, 2010, p. 2
9. L. Backstrom, D. Huttenlocher, J. Kleinberg, and X. Lan, “Group formation in large social networks: membership, growth, and evolution,” in Proc. SIGKDD, 2006, pp. 44–54.
10. N. Tran, M. K. Aguilera, and M. Balakrishnan, “Online migration for geo-distributed storage systems,” in Proc. USENIX ATC, 2011, p. 15.
11. A. Vázquez et al., “Modeling bursts and heavy tails in human dynamics,” *Phys. Rev. E*, vol. 73, no. 3, p. 036127, 2006.
12. J. M. Pujol et al., “The little engine(s) that could: Scaling online social networks,” *IEEE/ACM Trans. Netw.* vol. 20, no. 4, pp. 1162–1175, Aug. 2012.
13. H. Roh, C. Jung, W. Lee, and D. “Resource pricing game in geo distributed clouds,” in Proc. IEEE INFOCOM, 2013, pp. 1519–1527.
14. K. Schloegel, G. Karypis, and V. Kumar, “Wavefront diffusion and LMSR: Algorithms for dynamic repartitioning of adaptive meshes,” *IEEE Trans. Parallel Distrib. Syst.*, vol. 12, no. 5, pp. 451–466, May 2001.
15. Y. Sovran, R. Power, M. K. Aguilera, and J. Li, “Transactional storage geo-replicated systems,” in Proc. SOSP, 2011, pp. 385–400.
16. D. A. Tran, K. Nguyen, and C. Pham, “S-clone: Socially-aware data replication for social networks,” *Comput. Netw.*, vol. 56, no. 7, pp. 2001–2013, 2012.
17. N. Tran, M. K. Aguilera, and M. Balakrishnan, “Online migration for geo-distributed storage systems,” in Proc. USENIX ATC, 2011, p. 15.



18. A. Vázquez et al., “Modeling bursts and heavy tails in human dynamics Rev. E, vol. 73, no. 3, p. 036127, 2006. International Journal of Engineering and Techniques - Volume 3 Issue 1, Jan

–