ISSN: 2278-0181
ICACC - 2016 Conference Proceedings

Quality of Service in Cloud Computing by Traffic Redundancy Elimination and Double Quality Guarenteed Scheme

Y. Krishna Priyanka
PG Scholar, CSE
Sri Vasavi Engineering College
Tadepalligudem

G. Loshma
Associate Professor, CSE
Sri Vasavi Engineering College
Tadepalligudem

***Abstract*- Quality of service in the cloud environment is the ability to provide different priority to different applications, users, or data flows, or to guarantee a certain level of performance. With the given quality attributes like:**

- **Loss: probability that a flow's data is lost!**
- **Delay: time it takes a packet's flow to get from source to destination!**
- **Delay jitter: maximum difference between the delays experienced by two packets of the flow!**
- **Bandwidth: maximum rate at which the source can send data!**
- **QOS spectrum**

PACK (Predictive Acknowledgement's): Prediction-Based Cloud Bandwidth and Cost Reduction System is a novel end-to-end traffic redundancy elimination (TRE) system [1]. Here TRE algorithm was used to reduce traffic and to increase the efficiency [8].A Profit Maximization Scheme with Guaranteed Quality of Service in Cloud Computing was also used for Quality of Service along with maximising the profit. For this, Double renting scheme was used but it is does not consider profit and QOS together. So the single renting scheme was used to achieve profit and Quality of Service. Profit was better for all requests, but Quality of Service is not guaranteed every time [9].

The proposed system concentrates on the quality attributes like Profit, cost, Efficiency and Quality of Service. This work implements TRE algorithm and profit maximisation scheme together .Hence Quality of Service provided by the service provider will be increased .The profit maximisation scheme uses in this proposed scheme is Single renting scheme.
Objectives of the proposed system:

- **To maintain the cost of service.**
- **To eliminate the traffic and redundancy.**
- **To maximise profit.**
- **To reduce the resource wastage.**
- **To increase the quality.**

KeyWords- Qulity,renting scheme,profit,quality of service,traffic redundancy elimination.

I INTRODUCTION

Cloud computing offers its customers an economical and convenient pay-as-you-go service model, known also as usage-based pricing [2]. Cloud customers1 pay only for the actual use of computing resources, storage, and bandwidth,according to their changing needs, utilizing the cloud's scalable and elastic computational capabilities. In particular, data transfer costs (i.e., bandwidth) is an important issue when trying to minimize costs [2]. Consequently, cloud customers,applying a judicious use of the cloud's resources, are motivated to use various traffic reduction techniques, in particular traffic redundancy elimination (TRE), for reducing bandwidth costs. Traffic redundancy stems from common end-users' activities, such as repeatedly accessing, downloading, uploading (i.e., backup), distributing, and modifying the same or similar information items (documents, data, Web, and video). TRE is used to eliminate the transmission of redundant content and, therefore, to significantly reduce the network cost. In most common TRE solutions, both the sender and the receiver examine and compare signatures of data chunks, parsed according to the data content, prior to their transmission.When redundant chunks are detected, the sender replaces the transmission of each redundant chunk with its strong signature [3]–[5]. Commercial TRE solutions are popular at enterprise networks, and involve the deployment of two or more proprietary-protocol, state synchronized middle-boxes at both the intranet entry points of data centers and branch offices, eliminating repetitive traffic between them. Cloud computing turns information technology into ordinary commodities and utilities by the the pay-per-use pricing model [3, 4, 5]. In a cloud computing environment, there are always three tiers, i.e., infrastructure providers, services providers, and customers (see Fig. 1 and its elaboration inSection 3.1). An infrastructure provider Maintains the basic hardware and software facilities. A service provider rents resources from the infrastructure providers and providesservices to customers. A customer submits its request to a service provider and pays for it based on the amount and the quality of the provided service [6]. In this paper, we aim at researching the multiserver configuration of a service provider such that its profit is Maximized.Like all business, the profit of a service provider in cloud computing is related to two parts, which are the cost and the revenue. For a service provider, the cost is the renting Cloud computing turns information technology into ordinary commodities and utilities by the the pay-per-use pricing model [3, 4, 5]. In a cloud computing environment, there are always three tiers, i.e., infrastructure providers, services providers, and customers (see Fig. 1 and its elaboration in Section 3.1). An infrastructure provider maintains the basic hardware and software facilities. A service provider rents resources from the infrastructure providers and provides services to customers. A customer submits its request to a service provider and pays for it based on the amount and the quality of the provided service [6]. In this paper, we aim at researching the multiserver configuration of a service provider such that its profit is maximized. Like all business, the profit of a service provider in cloud computing is related to two parts, which are the cost and the revenue. For a service provider, the cost is the renting.

II. RELATED WORK

Several TRE techniques have been explored in recent years.A protocol-independent TRE was proposed in [4]. The paper describes a packet-level TRE, utilizing the algorithms presented in [3].Several commercial TRE solutions described in [6] and [7] have combined the sender-based TRE ideas of [4] with the algorithmic and implementation approach of [5] along with protocol specific optimizations for middle-boxes solutions. In particular,[6] describes how to get away with three-way handshake between the sender and the receiver if a full state synchronization is maintained.References and present redundancy-aware routing algorithm. These papers assume that the routers are equipped with data caches, and that they search those routes that make a better use of the cached data.A large-scale study of real-life traffic redundancy is presented in and In the latter, packet-level TRE techniques are compared [3],. Our paper builds on their finding that “an end to end redundancy elimination solution, could obtain most of the middle-box’s bandwidth savings,” motivating the benefit of low cost software end-to-end solutions. Wanax is a TRE system for the developing world where storage and WAN bandwidth are scarce. It is a software-based middle-box replacement for the expensive commercial hardware. In this scheme, the sender middle-box holds back the TCP stream and sends data signatures to the receiver middle-box.

The receiver checks whether the data is found in its local cache. Data chunks that are not found in the cache are fetched from the sender middle-box or a nearby receiver middle-box. Naturally,such a scheme incurs a three-way-handshake latency for noncached data. The pricing strategies are divided into two categories,i.e., static pricing and dynamic pricing. Static pricing means that the price of a service request is fixed and known in advance, and it does not change with the conditions.With dynamic pricing a service provider delays the pricing decision until after the customer demand is revealed, so that the service provider can adjust prices accordingly . Static
pricing is the dominant strategy which is widely used in real world and in research [2], Ghamkhari *et al.* adopted a flat-rate pricing strategy and set a fixed price for all requests, but Odlyzko in argued that the predominant flat-rate pricing encourages waste and is incompatible with service differentiation. Another kind of static pricing strategies are usage-based pricing. For example, the price of a service request is proportional to the service time and task execution requirement (measured by the number of instructions to be executed) in and [2], respectively.

Usage-based pricing reveals that one can use resources more efficiently . Dynamic pricing emerges as an attractive alternative to better cope with unpredictable customer demand . Mac´ıas *et al.* used a genetic algorithm to iteratively optimize the pricing policy. Amazon EC2 has introduced a ”spot pricing” feature, where the spot price for a virtual instance is dynamically updated to match supply and demand. However,consumers dislike prices to change,especially if they perceive the changes to be”unfair”. After comparison, we select the usage-based pricing strategy in this paper since it agrees with the concept of cloud computing mostly.

III.STRATEGY TO INCREASE THE QUALITY OF SERVICE

PACK algorithm and Double Quality Guarenteed Scheme combinely used together to enhance the quality of service.

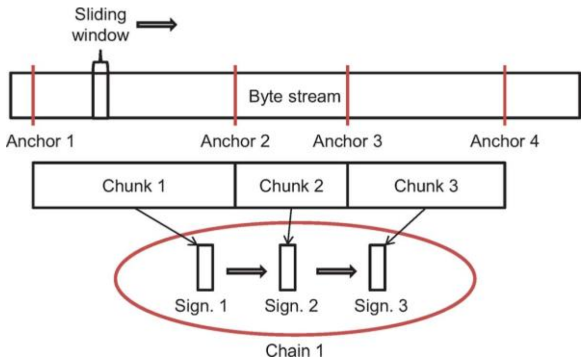


Fig:from stream to chain

PACK ALGORITHM

For the sake of clarity, we first describe the basic receiverdriven operation of the PACK protocol. Several enhancements and optimizations are introduced in Section IV.

The stream of data received at the PACK receiver is parsed to a sequence of variable-size, content-based signed chunks similar to [3] and [5]. The chunks are then compared to the receiver local storage, termed chunk store. If a matching chunk is found in the local chunk store, the receiver retrieves the sequence of subsequent chunks, referred to as a chain, by traversing the sequence of LRU chunk pointers that are included in the chunks’ metadata. Using the constructed chain, the receiver sends a prediction to the sender for the subsequent data. Part of each chunk’s prediction, termed a hint, is an easy-to-compute function with a small-enough false-positive value, such as the value of the last byte in the predicted data or a byte-wide XOR checksum of all or selected bytes. The prediction sent by the receiver includes the range of the predicted data, the hint, and the signature of the chunk. The sender identifies the predicted range in its buffered data and verifies the hint for that range. If the result matches the received hint,it continues to perform the more computationally intensive SHA-1 signature operation. Upon a signature match, the sender sends a confirmation message to the receiver, enabling it to copy the matched data from its local storage.

PACK CHUNKING ALGORITHM

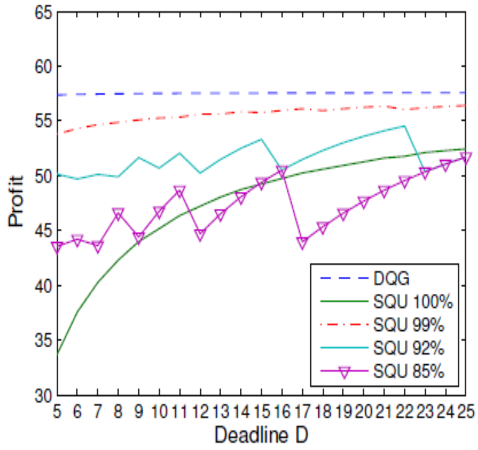
1. *mask* ⇐ $0x00008A3110583080$ {48 bytes window; 8 KB chunks}
2. *longval* ⇐ 0 {has to be 64 bits}
3. **for all** byte ∈ stream **do**
4. shift left *longval* by 1 bit {lsb ← 0; drop msb}
5. *longval* ⇐ *longval* bitwise-xor *byte*
6. **if** processed at least 48 bytes **and** (*longval* bitwise-and *mask*) == *mask* **then**
7. found an anchor
8. **end if**
9. **end for**

IV.DOUBLE QUALITY GUARANTEED SCHEME

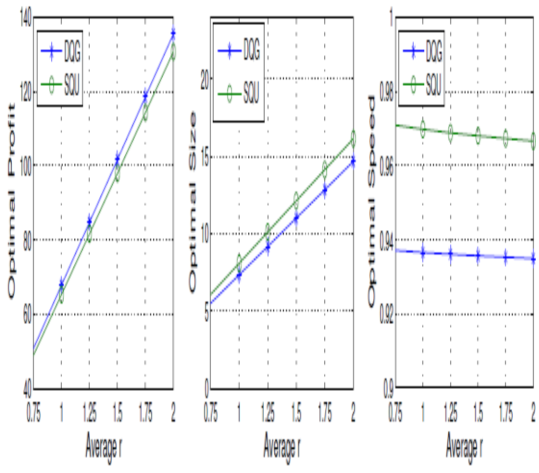
Algorithm 1 Double-Quality-Guaranteed (DQG) Scheme

1: A multiserver system with m servers is running and waiting for the events as follows
2: A queue Q is initialized as empty
3: Event – A service request arrives
4: Search if any server is available
5: **if** true **then**
6: Assign the service request to one available server
7: **else**
8: Put it at the end of queue Q and record its waiting time
9: **end if**
10: End Event
11: Event – A server becomes idle
12: Search if the queue Q is empty
13: **if** true **then**
14: Wait for a new service request
15: **else**
16: Take the first service request from queue Q and assign it to the idle server
17: **end if**
18: End Event
19: Event – The deadline of a request is achieved
20: Rent a temporary server to execute the request and release the temporary server when the request is completed
21: End Event

The Double-Quality-Guaranteed (DQG) resource renting scheme which combines long-term renting with short-term renting. The main computing capacity is provided by the long-term rented servers due to their low price. The short-term rented servers provide the extra capacity in peak period. The proposed DQG[8] scheme adopts the traditional FCFS queueing discipline. For each service request entering the system, the system records its waiting time. The requests are assigned and executed on the long-term rented servers in the order of arrival times. Once the waiting time of a request reaches *D*, a temporary server is rented from infrastructure.

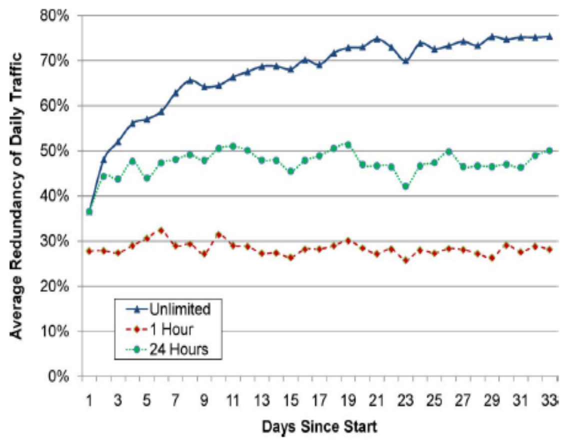


Profit comparison



(a) Comparison of Profit. (b) Comparison of Server Size. (c) Comparison of Server Speed.

Comparison between the schemes



Social networking site redundancy elimination

V. CONCLUSION

Cloud computing is expected to trigger high demand for TRE solutions as the amount of data exchanged between the cloud and its users is expected to dramatically increase. The cloud environment redefines the TRE system requirements, making proprietary middle-box solutions inadequate. Consequently, there is a rising need for a TRE solution that reduces the cloud's operational cost while accounting for application latencies, user mobility, and cloud elasticity.In order to guarantee the quality of service requests and maximize the profit of service providers, this paper has proposed a novel Double-Quality-Guaranteed (DQG) renting scheme for service providers. This scheme combines short-term renting with long-term renting, which can reduce the resource waste greatly and adapt to the dynamical demand of computing capacity. An *M/M/m+D* queueing model[9] is build for our multiserver system with varying system size. And then, an optimal configuration problem of profit maximization is formulated in which many factors are taken into considerations, such as the market demand,the workload of requests, the server-level agreement, the rental cost of servers, the cost of energy consumption, and so forth. The optimal solutions are solved for two different situations, which are the ideal optimal solutions and the actual optimal solutions. In addition, a series of calculations are conducted to compare the profit obtained by the DQG renting scheme with the Single-Quality-Unguaranteed (SQU)[10] renting scheme. The results show that our scheme outperforms the SQU scheme in terms of both of service Quality and profit.

REFERENCES

[1] E. Zohar, I. Cidon, and O. Mokryn, “The power of prediction: Cloud bandwidth and cost reduction,” in Proc. SIGCOMM, 2011, pp. 86–97.

[2] M. Armbrust, A. Fox, R. Griffith, A. D. Joseph,R.Katz, A. Konwinski,G. Lee, D. Patterson, A. Rabkin, I. Stoica, and M. Zaharia, “A view of cloud computing,” Commun. ACM, vol. 53, no. 4, pp. 50–58, 2010.

[3] U. Manber, “Finding similar files in a large file system,” in Proc.USENIX Winter Tech. Conf., 1994, pp. 1–10.

[4] N. T. Spring and D. Wetherall, “A protocol-independent technique for eliminating redundant network traffic,” in Proc. SIGCOMM, 2000, vol.30, pp. 87–95.

[5] A. Muthitacharoen, B. Chen, and D. Mazières, “A low-bandwidth network file system,” in Proc. SOSP, 2001, pp. 174–187.

[6] E. Lev-Ran, I. Cidon, and I. Z. Ben-Shaul, “Method and apparatus for reducing network traffic over low bandwidth links,” US Patent7636767, Nov. 2009.

[7] S.Mccanne andM. Demmer, “Content-based segmentation scheme for data compression in storage and transmission including hierarchical segment representation,” US Patent 6828925, Dec. 2004.

[8] K. Hwang, J. Dongarra, and G. C. Fox, Distributed and Cloud Computing. Elsevier/Morgan Kaufmann, 2012.

[9] J. Cao, K. Hwang, K. Li, and A. Y. Zomaya, “Optimal multiserver configuration for profit maximization in cloud computing,” IEEE Trans. Parallel Distrib. Syst., vol. 24, no. 6, pp. 1087–1096, 2013.

[10] A. Fox, R. Griffith, A. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, and I. Stoica, “Above the clouds: A berkeley view of cloud computing,” Dept. Electrical Eng. and Comput. Sciences, vol. 28, 2009.

[11] R. Buyya, C. S. Yeo, S. Venugopal, J. Broberg, and Brandic, “Cloud computing and emerging it platforms: Vision, hype, and reality for delivering computing as the 5th utility,” Future Gener. Comp. Sy., vol. 25, no. 6, pp. 599– 616, 2009.

[12] P. Mell and T. Grance, “The NIST definition of cloud computing.national institute of standards and technology,” Information Technology Laboratory, vol. 15, p. 2009, 2009.

[13] J. Chen, C. Wang, B. B. Zhou, L. Sun, Y. C. Lee, and Y. Zomaya, “Tradeoffs between profit and customer satisfaction for service provisioning in the cloud,” in Proc. 20th Int'l Symp. High Performance Distributed Computing. ACM, 2011, pp. 229–238.

[14] J. Mei, K. Li, J. Hu, S. Yin, and E. H.-M. Sha, “Energyaware preemptive scheduling algorithm for sporadic tasks on dvs platform,” MICROPROCESS MICROSY., vol. 37,no. 1, pp. 99–112, 2013.

[15] P. de Langen and B. Juurlink, “Leakage-aware multiprocessor scheduling,” J. Signal Process. Sys., vol. 57, no. 1, pp.73–88, 2009.

[16] G. P. Cachon and P. Feldman, “Dynamic versus static pricing in the presence of strategic consumers,” Tech. Rep.,2010.

[17] Y. C. Lee, C. Wang, A. Y. Zomaya, and B. B. Zhou, “Profit driven scheduling for cloud services with data access awareness,” J. Parallel Distr. Com., vol. 72, no. 4, pp. 591–602, 2012.