

ADMISSION CONTROL OF MULTIMEDIA SESSIONS TO A SET OF MULTIMEDIA SERVERS CONNECTED BY AN ENTERPRISE NETWORK

*Md. Monirul Islam*¹, *Md. Mostofa Akbar*¹, *Hemayet Hossain*¹ and *Eric G. Manning*²

¹Dept. of CSE, BUET, Dhaka-1000, Bangladesh, ²Dept. of CSc and ECE, Uvic, Victoria, BC, Canada
Email: {mmislam, mostofa, hemayet}@cse.buet.ac.bd, eric.manning@engr.uvic.ca

ABSTRACT

The satisfactory electronic delivery of multimedia material not only requires a particular Quality of Service (QoS) level to be maintained to fulfill users' expectations; there is in addition another goal: maximizing the revenue earned for the network owner. In this paper, we designed an Admission Controller (AC) along with the necessary algorithms to allow an Enterprise Network (EN) to satisfy both of these goals. We consider the selection of both a server from several alternatives, and the selection of a delivery route, plus the reservation of resources both in the selected server and on the delivery route to achieve the two goals of guaranteed absolute QoS and revenue optimization. We simulated the AC; simulation results are provided together with explanations.

1. INTRODUCTION

Overloaded servers and congested network paths are among the main reasons for excessive delay in accessing content from a remote server. Excessive delay or variance of delay in data transmission are annoying to users and definitely unacceptable to customers paying real money to enjoy streaming multimedia content. Hence networks must be able to carry multimedia streams with guaranteed absolute QoS. The current best effort, connectionless IP datagram service is incapable of absolute service guarantees unless the network is heavily overprovisioned. Hence some form of connection-oriented service atop the IP datagram service is a necessary condition for guaranteed QoS. Moreover, the necessary resources must be reserved both on the network links and in the servers for all admitted sessions. However, finite resources implies that an unbounded number of users cannot be admitted while respecting all guarantees of absolute QoS. Hence admission criteria are required for the selection of a subset of multimedia sessions from the set seeking admission, so that the QoS of all admitted sessions can be guaranteed, while maximizing the owner's revenue.

This paper is organized as follows. Section 2 provides a literature review of admission control problems

and Section 3 defines the present problem and redefines the notion of Service Level Agreement (SLA). Section 4 presents a new Admission Control Algorithm (ACA) and Section 5 describes its complexity. Section 6 presents and discusses the results obtained. The paper is concluded in Section 7, mentioning major contributions and further research work.

2. RELATED WORK

Finding a near optimal selection of both a server and of a delivery route from server to customer, is an important topic in content routing research [1]. Khan et al [2] study the delivery of multimedia streams with guaranteed QoS and maximized utility from a single multimedia server. Replicating multimedia data among multiple servers and selecting a nearly optimal one based on server parameters has been studied by Akbar et al [3]. However, their solution does not find an optimal delivery route. The transmission of multimedia streams through the links of an EN [EN- a network of 50 nodes or fewer operated by a single entity and used for video and teleconferencing] by controlling admission based on SLAs concluded between the customers and the network operator has been proposed by Watson [4] and further studied by Akbar et al [5]. However, the problem in their approach is that it determines the path and QoS level based solely on network resources, without considering the server parameters, and therefore, it cannot find the optimal server.

3. PROBLEM DEFINITION AND SLA

Limited resources in servers and on network paths are bottlenecks for multimedia transmissions, especially when there are many parties contesting for resources. This paper presents an AC for a group of multimedia server systems in an EN, as shown in Fig 1. Multimedia content such as videos is replicated in multiple servers, and multiple paths exist between a server and a customer. Customers at workstations send service requests to the AC in the form of proposed SLAs. The basic definition of SLA, taken from [6], is redefined in this paper. Here, SLA_i the i th SLA, contains

- multimedia data, m_i to be enjoyed by the customer
- customer's location, d_i
- QoS level vector, q_i . The j th level of q_i contains
 - network delay bound, D_{ij}
 - revenue offered by user, r_{ij}
 - network Bandwidth (BW) requirements, b_{ij}
 - server resource requirements like memory, CPU cycles, I/O BW, etc.

The SLA does not include the identity of the source of the media stream. The AC must find both the appropriate source server, and the delivery route to the user for each of the optimally selected SLAs, subject to the resource constraints and the goal of revenue maximization.

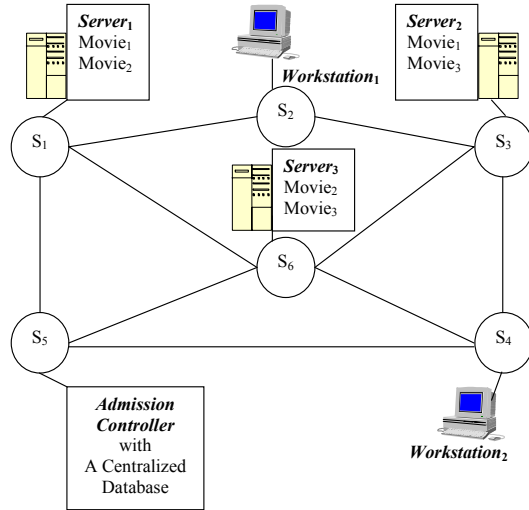


Fig 1: Multimedia server system

4. ADMISSION CONTROL METHODOLOGY

Fig 2 shows our ACA. Admission control is done on batches of SLAs. SLAs are batched (Step 1) during time intervals called epochs. Step 2 maps SLAs to an MMKP, a variant of the classical Knapsack Problem. A Multidimensional Multiconstraint KP contains multiple resource dimensions and multiple groups, each containing multiple items. [Here, each group of items represents a proposed SLA and each item of the group represents admission of the SLA at that level of QoS; and different servers and resources represent resource dimensions.] Each selected item therefore consumes resources from each of the resource dimensions. At most one item is picked from each group, [i.e., an SLA is admitted at most one level of QoS], subject to multiple resource constraints, and one objective function – operator revenue – is maximized. Step 2.1 finds the sources of the data and Step 2.2 finds k shortest paths for every source–destination pair using Eppstein's k shortest paths algorithm [4]. Step 2.3 creates items for a group of the MMKP, from every pair comprising a delay-constrained path and a QoS level.

These are inserted into an array named the QoS_path_list . For each new SLA, we add a null or dummy QoS level to the QoS profile (Step 2.4). Selection of this level means rejection of an SLA. Step 3 applies a heuristic algorithm I-HEU [5], to solve the MMKP, which may admit some new SLAs from the current batch.

```

Procedure admission()
Begin
1 batched_sla_list ← new_batched_sla_list
2 for i ← 1 to size(batched_sla_list[ ]) do
2.1 sources ← get_sources(batched_sla_list[i].data)
2.2 path_list ← null
d ← batched_sla_list[i].destination
for s ← 1 to size(sources) do
t ← sources[s]
temp ← determine_K_candidate_paths(t, d)
path_list ← path_list + temp
endfor
2.3 for j ← 1 to size(batched_sla_list[i].QoS_list) do
for k ← 1 to size(path_list[ ]) do
tLevel ← batched_sla_list[i].QoS_list[j]
if delay(path_list[k]) < delay(tLevel) then
add ( path_list[k], bid ( tLevel)) in
batched_sla_list[i].QoS_path_list
endif
endfor
endfor
2.4 add (null, null) in the
batched_sla_list[i].QoS_path_list
endfor
3 temp_sla_list ← active_sla_list + batched_sla_list
I-HEU(temp_sla_list)
accepted ← accepted_slas(temp_sla_list)
active_sla_list ← active_sla_list + accepted
rejected_sla_list ← rejected_slas(temp_sla_list)
endProcedure

```

Fig 2: Admission Control Algorithm

5. COMPLEXITY

Suppose that, n proposed SLAs, each with l QoS levels, are submitted to the AC. The number of nodes and edges in the EN are N and $O(N)$, respectively. Each multimedia stream is replicated on each of s servers. The number of alternate paths from each source to customer destination is K . Using Epstein's algorithm [4] to find the shortest paths we need $O(sN \log N + sKN)$ time [7]. Next we map the problem to an MMKP with $O(N)$ resource dimensions, n groups, and a maximum of lsK items in each group. The worst-case complexity of I-HEU [5] to solve this MMKP is $O(Nn^2 (lsK - 1)^2)$. So, the total complexity is $O(sN \log N + sKN) + O(Nn^2 (lsK - 1)^2)$.

6. SIMULATION RESULTS AND DISCUSSION

Figures 3 through 8 show the results obtained by simulation of the effects of system size, epoch duration and the number of shortest paths. We ran the Java simulation of the AC on a single processor Pentium-IV

PC with 512 MB RAM. We used a Harray graph, $H_{31,4}$ as an EN with 31 nodes and 62 links [8]. A set of 10 movies, each of length 3 hours, were replicated in different movie servers of the network. The number of replications of each movie was half of the number of servers. So, each server stored 5 different movies. Network link resource was held at 50 times the number of servers. Epoch duration and the number of shortest paths were generally held constant at 15 seconds and 3, respectively. Otherwise, they were varied as shown on abscissae of the corresponding figures. System size is defined by the number of servers and the underlying network as a whole. The SLA arrival rate was assumed governed by an exponential distribution with mean $Expected_No_of_Users/Service_duration$, and departures were controlled by uniform distribution for service duration. We observe from the simulation results that the number of admissible SLAs (n), and thereby revenue, depends on the available resources which are almost proportional to the number of servers (s) in the system. The variation in epoch or the number of shortest paths does not change the total available resources, so total admission and revenue does not change much in these cases. [The observed results were not shown in this paper due to the limitation on the total length of this paper.] The complexity of the admission controller is approximately $O(Nn^2l^2s^2K^2)$. As n is proportional to s in this case, the admission time (T) is nearly $O(s^4)$ where N , l and K are kept constant. This is depicted in Fig 3. On the other hand, T is a quadratic function of K , when K is varied while leaving the other parameters unchanged, as shown in Fig 7. However, the change in epoch does not change most of the terms of the complexity expression, except the number of SLAs contesting for admission. That is why Fig 5 shows only small increases in admission time when the epoch value increases. The average resource utilization is affected by various reasons. At small system sizes, links nearby servers become saturated quickly keeping most of the other links underutilized. On the other hand, with very large system size many SLAs get admitted, and many of these find servers in nearby nodes or even in their own nodes. So they consume very little or no link bandwidth. Hence resource utilization is lower in these cases, compared to the case of moderate system size. Fig 4 clarifies our views on system size. An increase in the time value of epochs allows larger batches of SLAs to arrive at the AC together. Hence the AC has more SLAs to choose among and resources are more effectively used, which increases resource utilization. That is why Fig 6 shows increasing utilization with increased values of epoch. On the other hand, with an increase in the number of shortest paths, SLAs with a larger set of source and destination nodes are likely to be admitted, leading to increased link utilization and thus increased overall resource utilization. Fig 8 shows this effect.

To validate our simulation results, we present a graph theoretical analysis. Let, there be N nodes and $N \times r$ links in the EN, where r is the degree of nodes. The

average distance between two nodes is $L = \left\lceil \left(\frac{N}{2r} + 1 \right) / 2 \right\rceil$.

On average, L links, each with B_L BW, will be exhausted by $\lfloor B_L/B_q \rfloor$ admitted SLAs, where B_Q is the average BW requirement per SLA. So, the total number of admissible SLAs has mean value of $\left\lceil \frac{B_L}{B_Q} \right\rceil \times \lfloor N \times r / L \rfloor$. On the other hand, each movie is replicated in s servers and each SLA requires a fraction q of total server resources. So, considering both server and link parameters, the average number of total admissible SLAs will be $\min \left(\left\lceil \frac{B_L}{B_Q} \right\rceil \times \lfloor N \times r / L \rfloor, \left\lfloor \frac{s}{q} \right\rfloor \right)$. Now, in our simulation, $N=31$ and $r=2$. So, L becomes 5 and we have the mean values- $B_Q=5$, $q=0.01$. So, the expected number of admissions is 200 for $s=2$ and $B_L=100$, where-as this number is 165 from our simulation, which is 82.5% of the expected value.

7. CONCLUSION

A central challenge facing network operators is maintaining customer guarantees while sustaining profitability. In this study we derived a new ACA for a multimedia system with multiple servers supported by an EN. This is able to admit enough customers to efficiently utilize the system resources, while maximizing the earned revenue from the customers. We have also studied the AC by extensive simulations. Finally, we presented a simple graph theoretic analysis for number of admissions and revenue. More analysis is required to predict the average resource utilization. Moreover, further study is required to upgrade and analyze the AC if customers are allowed to make future reservations. Our current methodology can be extended for the cases where different portions of movies are kept in different servers or movies are multicast to users having common QoS requirements.

8. REFERENCES

- [1] M. Gritter and D. R. Cheriton. An Architecture for Content Routing Support in the Internet. <http://www.dsg.stanford.edu/papers/contentrouting/>, 2001.
- [2] S. Khan, E. G. Manning, & K. F. Li. The Utility Model for Adaptive Multimedia Systems. Intl. Conf. Multimedia Modeling, Singapore, pp. 111-126, Jan. 1998.
- [3] M M Akbar, E. G. Manning and G. C. Shoja. Application of UM-D to Optimal Server Selection for Content Routing. ICECE 2002. Dhaka, December 27-28, 2002.
- [4] D. Eppstein. Finding the k Shortest Paths. 35th IEEE Symp. Foundations of Computer Science, Santa Fe, pp. 154-165, 1994.
- [5] M. M. Akbar, E. G. Manning, G. C. Shoja and S. Khan. Heuristic Solutions for the Multidimensional

Multiple-Choice Knapsack Problem, International Conference on Computational Science. pp 659-668, Part II, San Francisco, U.S.A, May 2001.

[6] M M Akbar, E. G. Manning, R. K. Watson, G. C. Shoja, S. Khan and K. F. Li. Optimal Admission Controllers for Service Level Agreements in Enterprise Networks. 6th SCI Conference. Orlando, FL July 14-18, 2002.

[7] M. M. Islam. Admission Control of Multimedia Sessions to a Set of Multimedia Servers Connected Through an Enterprise Network. M. Sc. Engg. Thesis, Department of CSE, BUET, December, 2004.

[8] D. B. West. Introduction to Graph Theory. Prentice-Hall of India Private Limited, page no. 135, 2000.

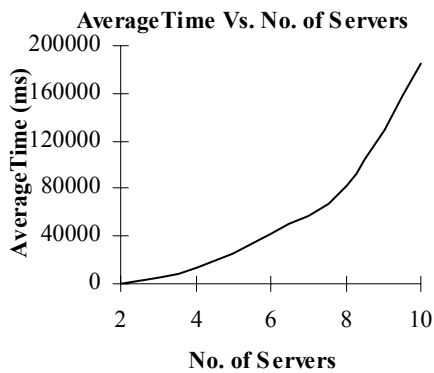


Fig 3: Effect of system size on average admission time

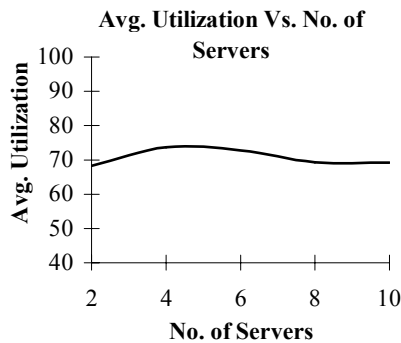


Fig 4: Effect of system size on average resource utilization

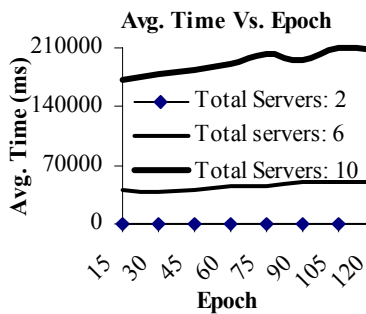


Fig 5: Effect of epoch on average admission time

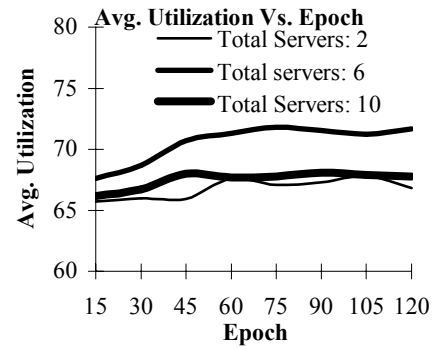


Fig 6: Effect of epoch on average resource utilization

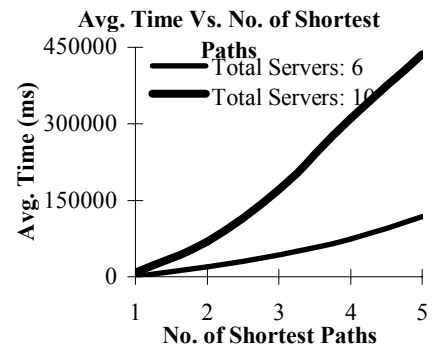


Fig 7: Effect of the No. of Shortest Paths on average admission time

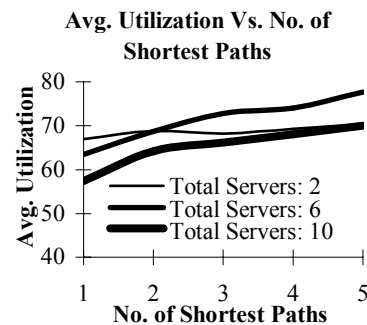


Fig 8: Effect of the No. of Shortest Paths on average resource utilization