This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE GLOBECOM 2005 proceedings.

Integrating Traffic Aggregation Mechanism into SIP Based IP Telephony over MPLS Network

Bo Rong, Bernard Tremblay, Maria Bennani, Michel Kadoch
LAGRIT, Department of Electrical Engineering
Ecole de technologie superieure, Universite du Quebec
Montreal, Quebec, Canada, H3C 1K3

Abstract—**In order to provide high quality multimedia service in a flexible and intelligent manner, future wired and wireless networks will employ SIP as the call control protocol and MPLS as the core network packet switching technology. In this paper, we propose a traffic aggregation based SIP over MPLS network architecture to integrate SIP protocol with the traffic engineering (TE) function of MPLS network seamlessly. Furthermore, the call admission and bandwidth prediction/re-negotiation algorithms running in this new network architecture are investigated intensively. We rely on OPNET simulation to assess the performance of different call admission and bandwidth prediction/re-negotiation algorithms. The simulation results show that the algorithm of *FCFS-LMS-BPR* is an eligible candidate for achieving the designated goals.**

I. Introduction

Recently there is a rapid growth in the popularity of SIP over MPLS technology mainly due to two reasons. Firstly, in the year 2000, Session Initiation Protocol (SIP) [1] was selected as the call control protocol for 3G IP-based mobile networks. Secondly, Multi-protocol Label Switching (MPLS) has been suggested as an apt packet switching technology for future core networks [2~5]. Despite the fact that MPLS development started with goal to expedite packet forwarding, the main benefit from MPLS in current network environment is from its traffic engineering (TE) capabilities including QoS guarantee, efficient usage of network resources, resilience, and economy [6~8]. In fact, the traffic engineering capabilities make MPLS network a good choice for carrying SIP signaled audio/video traffics.

Although the architecture of SIP over MPLS has been discussed in [9,10], according to our study, directly using SIP protocol over traffic engineering enabled MPLS network as previous literature suggested can cause an unbearable long SIP call setup delay. In traffic engineering enabled MPLS network, when a new SIP call request comes, usually Constraint-based Routing Label Distribution Protocol (CR-LDP) or Resource Reservation Protocol with Traffic Engineering Extensions (RSVP-TE) will be employed to set up a corresponding Label Switched Path (LSP) dynamically. This kind of dynamic LSP setup can contribute a lot to the whole SIP call setup delay in traffic engineering enabled MPLS network. If we use Label Distribution Protocol (LDP) instead of CR-LDP or RSVP-TE in MPLS network, the SIP call setup delay is short and acceptable. Nevertheless, LDP cannot support traffic engineering function in MPLS network, because it is only suitable for hop-by-hop label distribution and always selects the same physical path as conventional IP routing would select.

To solve the long SIP call setup delay problem discussed above, we introduce a traffic aggregation based SIP over MPLS network architecture in this paper. As a crucial component of this new architecture, the SIP-MPLS traffic aggregation server (TA server) is emphasized. The TA servers exchange traffic engineering signaling with MPLS network on behalf of a cluster of SIP clients. As a result, the SIP call setup delay in MPLS network is decreased substantially. Since the OPNET simulation study of SIP call setup delay in traffic aggregation based SIP over MPLS network architecture has been presented in [11], this paper focuses on the performance investigation of different call admission and bandwidth prediction/re-negotiation algorithms employed by the TA server.

The rest of the paper is organized as follows. Firstly, we introduce the traffic aggregation based SIP over MPLS network architecture in Section 2. Then, the call admission and bandwidth prediction/re-negotiation algorithms running on TA servers are suggested in Section 3. Furthermore, Section 4 presents the OPNET simulation results of different call admission and bandwidth prediction/re-negotiation algorithms. In the end, Section 5 summarizes the results.

II. Traffic Aggregation Based SIP over MPLS Network Architecture

In order to shorten the long SIP call setup delay caused by CR-LDP or RSVP-TE, a traffic aggregation based SIP over MPLS network architecture is proposed in this section. Fig.1 shows that this network architecture contains 2 parts, the local client network and the MPLS core network. Furthermore, there are two components in the client network, the SIP terminal and the TA server. As demonstrated in Fig.2, the TA server can be regarded as an enhanced SIP proxy server, and it negotiates/re-negotiates with the LER of MPLS core network about the overall QoS requirements of local client network on behalf of a cluster of SIP terminals (not only one SIP terminal). Then, the LER exchanges TE signaling with other routers inside the MPLS core network to set up the corresponding LSPs.

 0-7803-9415-1/05/$20.00

This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE GLOBECOM 2005 proceedings.

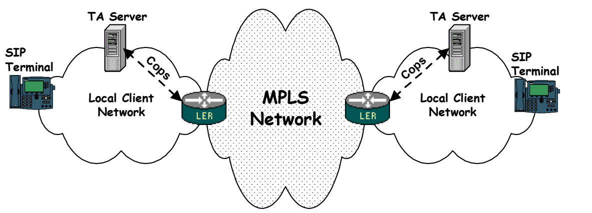


Fig. 1. Traffic Aggregation Based SIP over MPLS Network Architecture

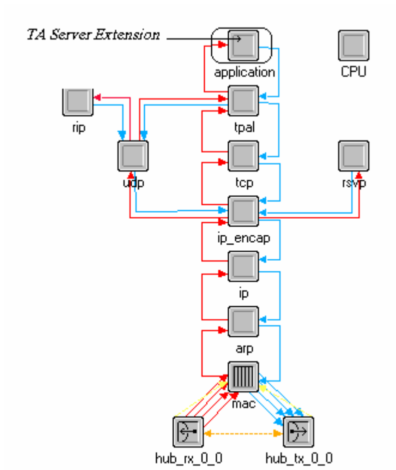


Fig.2. Extending a Traditional SIP Proxy Server to a TA Server in the Perspective of OPNET Node Model

We use a set of time marks $\{t_0, t_1, t_2, \ldots, t_{n-1}, t_n, t_{n+1}, \ldots\}$ to describe the time in system. If the TA server knows exactly that the overall bandwidth requirement of its local client network during $[t_{n-1}, t_n]$ is $B_{n-1,\ n}$, then at time t_{n-1}, the TA server negotiates with MPLS network to get $B_{n-1,\ n}$ outgoing bandwidth by using Common Open Policy Service (COPS) messages. As time goes by, if the TA server knows that the overall bandwidth requirement of its local client network during $[t_n, t_{n+1}]$ changes to $B_{n,\ n+1}$, then at time t_n, the TA server would re-negotiate with MPLS network to increase/decrease the bandwidth requirement to $B_{n,\ n+1}$. By this way, the LSPs needed by SIP telephony are setup in MPLS core network before SIP calls are made. As a result, the SIP call setup delay in MPLS network is decreased greatly. However, it is impossible for the TA server to know the exact value of $B_{n-1,\ n}$ before the time of t_{n-1}. Usually, the TA server can only employ a certain bandwidth prediction algorithm to give an approximate value of $B_{n-1,\ n}$, which can be defined as $B^{pred}_{n-1,\ n}$. If $B^{pred}_{n-1,n} < B_{n-1,n}$ during $[t_{n-1}, t_n]$, the local client network doesn't have enough outgoing bandwidth to accommodate all the SIP calls, and the TA server has to utilize admission control algorithm to decline some of the call requests. On the contrary, if $B^{pred}_{n-1,n} > B_{n-1,n}$ during $[t_{n-1}, t_n]$, some outgoing bandwidth resource of the local client network would be wasted. From the above discussion, we can conclude that the call admission and bandwidth prediction/re-negotiation algorithm running on TA servers is very important to the overall performance of traffic aggregation based SIP over MPLS network architecture.

Fig.3 shows the signaling flow in this new network architecture. The call setup starts with a standard SIP INVITE message sent by the caller to the local TA server. This message carries the callee URL in the SIP header and the session specification (session specification describes the QoS requirements of the SIP call) in the body SDP. The caller treats the TA server as a standard SIP proxy server. Regarding the caller ID, the session information and the remaining outgoing bandwidth in local client network, the TA server decides whether this SIP call request is admitted. If the call request is admitted, the TA server would forward the original INVITE message to the callee; Otherwise, it simply sends the caller a DECLINE message to drop the call. Furthermore, no matter the call is admitted or not, it would be registered in the TA server for the purposes of call admission and bandwidth prediction in the future. In addition, if the bandwidth re-negotiation is needed, the TA server re-negotiates with MPLS core network to adjust the bandwidth requirement by using COPS messages.

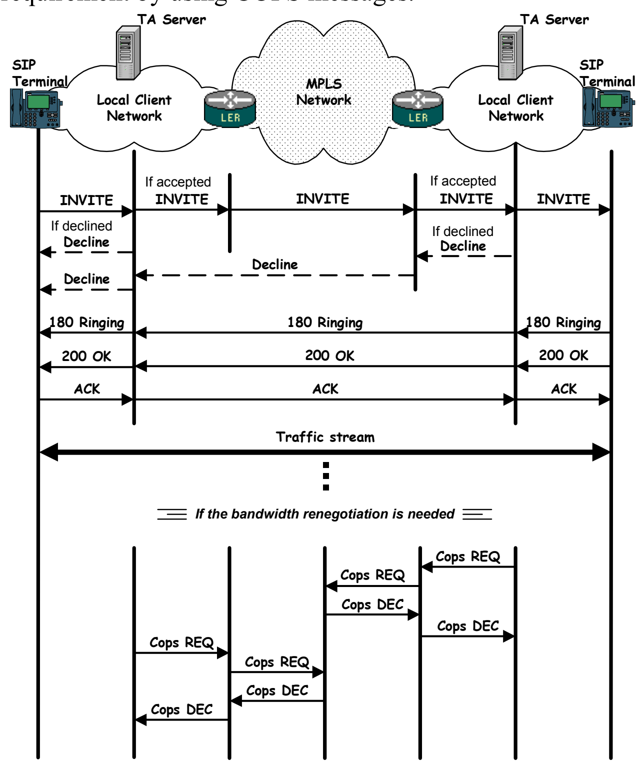


Fig. 3. The Signaling Flow in Traffic Aggregation Based SIP over MPLS Network Architecture

III. The Call Admission and Bandwidth Re-negotiation Algorithms on TA Server

In this section, the call admission and bandwidth prediction/re-negotiation algorithms running on TA server are discussed. Fig.4 shows that a call admission and bandwidth prediction/re-negotiation algorithm consists of 3 parts, i.e., Admission Control Module, Bandwidth Prediction Module, and Bandwidth Re-negotiation Module. These three modules work cooperatively to offer the call admission and bandwidth prediction/re-negotiation function.

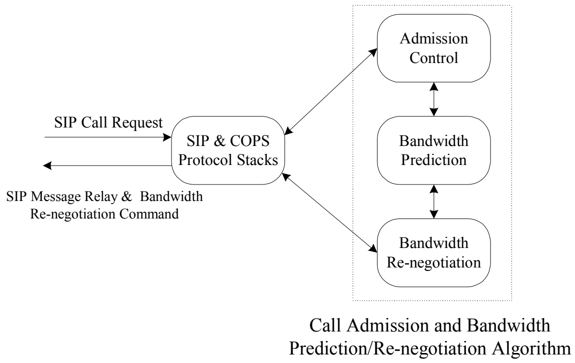


Fig.4. Architecture of Call Admission and Bandwidth Prediction/Re-negotiation Algorithm

Based on the architecture in Fig.4, three algorithms, including both simple and complicated ones, are depicted as follows.

(1) *FCFS with fixed bandwidth contract (FCFS-FB)*

In this case, the local client network has a fixed bandwidth contract with the MPLS core network, and the TA server accepts/declines SIP calls using *FCFS* policy. This algorithm has obvious drawbacks such as high call blocking probability or low bandwidth utility efficiency, because it has no bandwidth prediction/re-negotiation mechanisms.

(2) *FCFS with static adaptive bandwidth contract (FCFS-SAB)*

This algorithm divides one day into a couple of periods, and in each period the local client network has a specific fixed bandwidth contract with the MPLS core network. Same to *FCFS-FB*, *FCFS* is employed for call admission in this algorithm. If the fixed bandwidth contract of each period can be set up reasonably according to the statistics of everyday traffic load in the local client network, this algorithm has better performance than *FCFS-FB* due to its simple bandwidth prediction/re-negotiation function. However, the traffic load of local client network often behaves differently from one day to another. In this respect, this algorithm is not flexible enough.

(3) *FCFS with LMS based bandwidth prediction/re-negotiation (FCFS-LMS-BPR)*

Similar to *FCFS-FB* and *FCFS-SAB*, *FCFS-LMS-BPR* takes *FCFS* as its admission control policy. Differently, to improve the overall performance, *FCFS-LMS-BPR* employs *LMS based bandwidth prediction/re-negotiation* to enhance its bandwidth requirement tracing capability.

Before explaining the algorithm of *LMS based bandwidth prediction/re-negotiation*, basic knowledge of Least Mean Square (LMS) linear predictor [12] is introduced as follows. The k-step linear predictor is concerned with the prediction of $x(n+k)$ using a linear combination of the current and previous $x(n)$. A pth-order linear predictor to calculate $x'(n+k)$ (the predicted value of $x(n+k)$) has the following form:

$$x'(n+k) = \sum_{i=0}^{p-1} w_i(n)x(n-i) \quad (1)$$

where, x(n)represents the time series, $w_i(n)$ denotes the weights of the LMS predictor.

The LMS linear predictor is an adaptive approach, which does not require prior knowledge of the autocorrelation structure of a sequence. Therefore, it can be used as an on-line algorithm for forecasting bandwidth requirement. The LMS algorithm is initialized by setting all weights to zero at time n=0. Then, the weights are updated using the following relationship:

$$w_i(n+1) = w_i(n) + \mu * e(n) * x(n) \qquad i=0,1,\ldots,p-1 \quad (2)$$

where, $e(n)=x(n)-x'(n)$ is the error signal, and μ represents the step size.

By utilizing the LMS predictor, the algorithm of *LMS based bandwidth prediction/re-negotiation* can be implemented as follows.

LMS Based Bandwidth Prediction/Re-negotiation

Parameter Definitions:

$B_{req}'(t_n, t_{n+1})$: The predicted value of overall outgoing bandwidth requirement from local client network in the period of $[t_n, t_{n+1}]$.
$TR_{exp}'(t_n, t_{n+1})$: The predicted value of overall expired outgoing traffic load in local client network in the period of $[t_n, t_{n+1}]$.
$B_{con}(t_n, t_{n+1})$: The overall outgoing bandwidth that the TA server contracts with the MPLS core network in the period of $[t_n, t_{n+1}]$.
$B_{free}(t_n)$: The free outgoing bandwidth resource for local client network at time t_n.
B_{thresh}: The bandwidth re-negotiation threshold.

The Algorithm:

At time t_n, use two *1*-step LMS predictors to forecast the values of $B_{req}'(t_n, t_{n+1})$ and $TR_{exp}'(t_n, t_{n+1})$ respectively

If $B_{req}'(t_n, t_{n+1}) - TR_{exp}'(t_n, t_{n+1}) - B_{free}(t_n) > B_{thresh}$

/* make bandwidth re-negotiation with MPLS core network for more bandwidth*/
$B_{con}(t_n, t_{n+1})=B_{con}(t_{n-1}, t_n)+B_{req}'(t_n, t_{n+1})-TR_{exp}'(t_n, t_{n+1})-B_{free}(t_n)$

Else if $B_{req}'(t_n, t_{n+1}) - TR_{exp}'(t_n, t_{n+1}) - B_{free}(t_n) < -B_{thresh}$

/* make bandwidth re-negotiation with MPLS core network for less bandwidth*/
$B_{con}(t_n, t_{n+1})=B_{con}(t_{n-1}, t_n)-B_{req}'(t_n, t_{n+1})+TR_{exp}'(t_n, t_{n+1})+B_{free}(t_n)$

Else
No bandwidth re-negotiation is taken

The algorithm of *FCFS-LMS-BPR* is designed to have more flexibility, higher bandwidth utility efficiency, and lower call blocking probability than other two algorithms. The advantages of this algorithm mainly come from its prediction capability.

IV. OPNET SIMULATION OF DIFFERENT CALL ADMISSION AND BANDWIDTH PREDICTION/RE-NEGOTIATION ALGORITHMS

In this section, the simulation results are presented to demonstrate the performance of the algorithms suggested in Section 3. Although analytical approach can be applied to study these algorithms, it quickly becomes too complicated and intractable when the source model and the algorithm become complex. For this reason, we use OPNET simulation to study their performances. We suppose in the local client network there are 4 groups of SIP call requests, whose attributes are shown as follows.

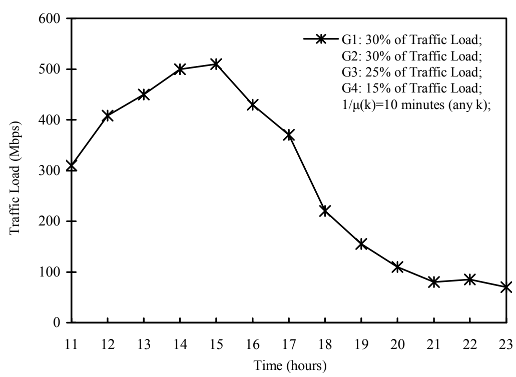
(a)group G_1: 64Kbps; (b)group G_2: 500Kbps;
(c)group G_3: 1.5Mbps; (d)group G_4: 4Mbps.

Obviously, G_1 and G_2 are groups of low-bandwidth call requests, while G_3 is the medium-bandwidth call group and G_4 is high-bandwidth call group. Moreover, call requests of every group are assumed to arrive according to Poisson process independently, and their service time is exponentially distributed. The parameters used in this model are defined below:

$\lambda(k)$(calls/hour)**:** the average arrival rate of G_k calls.
$1/\mu(k)$(hours/call)**:** the average service time of G_k calls.
Pb_k(%)**:** the blocking probability of group G_k.
Eff(%)**:** the bandwidth utility efficiency of a certain algorithm.

With OPNET simulation, numerical results are attained from Fig.5 to Fig.9 to illustrate the performance of *FCFS-FB*, *FCFS-SAB*, and *FCFS-LMS-BPR*. Parameter setup of the simulation is shown as follows:

(1)*Traffic load in the local client network (a typical scenario from Bell Canada):*



(2) *The parameter setup of these 3 algorithms*

(a) *FCFS-FB*: Fixed Bandwidth=250Mbps;

(b) *FCFS-SAB*:

Time of Day	Bandwidth Contract (Mbps)
11:00~12:30	350
12:30~15:30	390
15:30~17:30	210
17:30~19:30	65
19:30~23:00	30

(c) *FCFS-LMS-BPR*:

Global Parameter Setup of *FCFS-LMS-BPR*		Local Parameter Setup of 2 *LMS* Predictors	
Initialized Bandwidth Contract $B_{con}(0, t_1)$:	320*Mbps*	Order of *LMS* Predictor p:	7
Bandwidth Re-negotiation Threshold B_{thresh}:	10*Mbps*	Step Size of *LMS* Predictor μ:	0.006
Time Interval of Algorithm Execution $T_{exe\text{-}int}$:	10*Minutes*		

The simulation results from Fig.5 to Fig.9 demonstrate the bandwidth contract changing, blocking probability, and bandwidth utility efficiency of these 3 algorithms during half day. Firstly, we discuss the performance of *FCFS-FB*. As shown in Fig.5, *FCFS-FB* has no capability of bandwidth prediction. As a result, Fig.6 shows that the blocking probability of *FCFS-FB* is not stable. When the traffic load in local client network overweighs the outgoing bandwidth contract, the blocking probability is high. On the other hand, when the traffic load is lower than the outgoing bandwidth contract, the blocking probability is low, but a lot of bandwidth resource is wasted. Furthermore, from Fig.9 we know the bandwidth utility efficiency of *FCFS-FB* is varying with the traffic load in local client network arbitrarily. In a word, the performance of *FCFS-FB* is terrible. Secondly, the performance of *FCFS-SAB* is investigated. Fig.5, Fig.7, and Fig.9 demonstrate that *FCFS-SAB* has a good bandwidth utility efficiency but poor blocking probability when its bandwidth contract is lower than the real traffic load in most time of half day. Actually, it is very difficult for *FCFS-SAB* to have both good bandwidth utility efficiency and good blocking probability due to its inflexibility. Thirdly, we focus on the performance of *FCFS-LMS-BPR*. Fig.5 shows *FCFS-LMS-BPR* has high bandwidth prediction accuracy, while Fig.8 and Fig.9 illustrate it also possesses satisfying blocking probability and bandwidth utility efficiency. Obviously, *FCFS-LMS-BPR* has the best performance among these three algorithms, and it is an apt candidate to fulfill call admission and bandwidth prediction/re-negotiation task on TA servers.

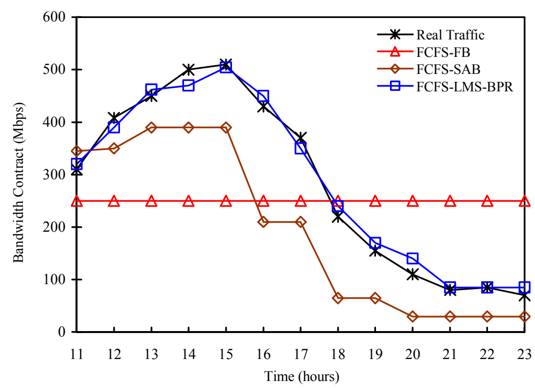


Fig. 5. Bandwidth Contract Changing of All 3 Algorithms during Half Day

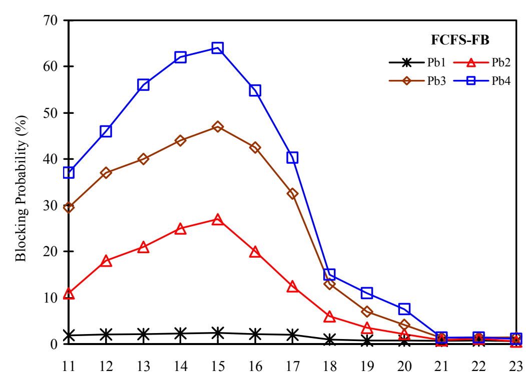


Fig. 6. Blocking Probability of *FCFS-FB* during Half Day

This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE GLOBECOM 2005 proceedings.

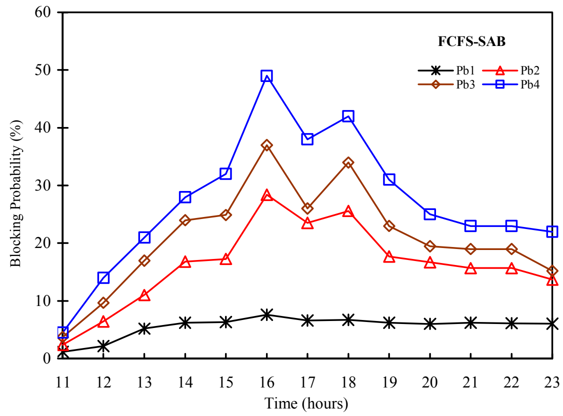


Fig. 7. Blocking Probability of *FCFS-SAB* during Half Day

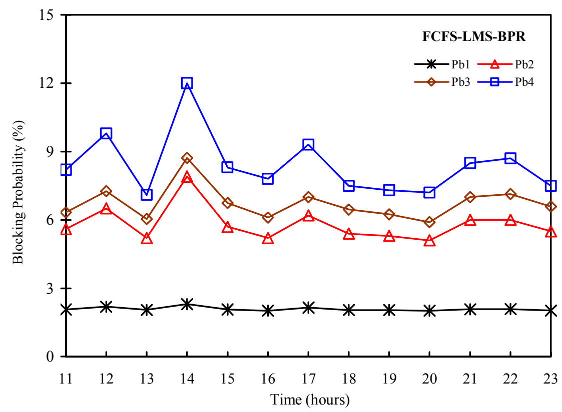


Fig. 8. Blocking Probability of *FCFS-LMS-BPR* during Half Day

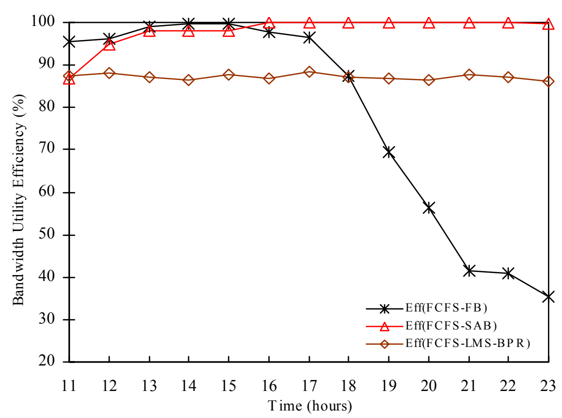


Fig. 9. Bandwidth Utility Efficiency of All 3 Algorithms during Half Day

V. Conclusion

We have presented a new network architecture of SIP telephony over MPLS in this paper. In order to integrate SIP protocol with the traffic engineering function of MPLS network seamlessly and facilitate SIP call setup, the TA server is emphasized as a key part of this network architecture. Furthermore, the call admission and bandwidth prediction/re-negotiation algorithms running on the TA server are investigated intensively. This paper relies on OPNET simulation to show the performance of different call admission and bandwidth prediction/re-negotiation algorithms. From the simulation results we can conclude that the algorithm of *FCFS-LMS-BPR* is suitable for TA servers, due to its satisfying performance of high bandwidth prediction accuracy, low call blocking probability, and high bandwidth utility efficiency.

Acknowledgment

The authors would like to thank Bell Canada and NSERC for their financial support.

References

[1] Faccin S.M., Lalwaney P., Patil B., " IP multimedia services: analysis of mobile IP and SIP interactions in 3G networks," *IEEE Communications Magazine*, Volume: 42, Issue: 1, Jan. 2004, pp. 113-120.

[2] Hyeong Ho Lee, Bu Ihl Kim, Jae Sup Lee, Chu Hwan Yim, "Structures of an ATM switching system with MPLS functionality," *GLOBECOM'99*, Volume: 1B, 1999, pp. 616-620.

[3] Li T., "MPLS and the evolving Internet architecture," *IEEE Communications Magazine*, Volume: 37, Issue: 12, Dec. 1999, pp. 38-41.

[4] Armitage G., "MPLS: the magic behind the myths [multiprotocol label switching]," *IEEE Communications Magazine*, Volume: 38, Issue: 1, Jan. 2000, pp. 124-131.

[5] Kankkunen A., "MPLS and next generation access networks," *ECUMN 2000*, Oct. 2000, pp. 5-16.

[6] Awduche D.O., "MPLS and traffic engineering in IP networks," IEEE Communications Magazine, Volume: 37, Issue: 12, Dec. 1999, pp. 42-47.

[7] Xipeng Xiao, Hannan A., Bailey B., Ni L.M., "Traffic engineering with MPLS in the Internet," IEEE Network, Volume: 14, Issue: 2, March-April 2000, pp. 28-33.

[8] Jong-Moon Chung, "Analysis of MPLS traffic engineering," *Proceedings of the 43rd IEEE Midwest Symposium on Circuits and Systems*, Volume: 2, Aug. 2000, , pp. 550-553.

[9] Zhang C., Guy C.G., "TE-SIP server design for a SIP-over-MPLS based network," *Proceedings of International Conference on Communication Technology 2003 (ICCT 2003)*, Volume: 2, April 2003, pp. 1758-1761.

[10] Salsano S., Veltri L., "QoS control by means of COPS to support SIP-based applications," *IEEE Network*, Volume: 16, Issue: 2, March-April 2002, pp. 27-33.

[11] Bo Rong, Lebeau J., Bennani M., Kadoch M., Elhakeem A. K., "Traffic Aggregation Based SIP over MPLS Network Architecture," *AINA 2005*, Volume: 1, 25-30 March 2005, pp. 827-832.

[12] Haykin S., *Adaptive Filter Theory*. Englewood Cliffs, NJ: Prentice Hall, 1991.