Priority Based Bandwidth Allocation and Load Balancing

Rekha V, Indiramma M
1Research Scholar, CSE Research Centre, Bangalore-19, India.
2Department of CSE, BMSCE, Bangalore-19, India.

ABSTRACT: Frequent Route failures are common in a multipath IP (Internet Protocol) network. Backup configuration is one of the techniques used to re-establish alternate path. There is a need to focus on priority based routing and load balancing during congestion in multipath networks. In this paper, we are proposing a priority based bandwidth allocation and load balancing (PBALB) approach for multipath routing. In order to reduce packet drop and to enhance fairness, throughput and lower delay transmissions, Traffic shaping based on different type of traffic flows in Differential service domain is proposed. Simulation result shows that the proposed PBALB technique improves the throughput.

KEYWORDS: IP, QoS, HTTP, FTP, Optimization, Compression, Robust, Load Balancing, Fairness, Priority, Scheduling.

I. Introduction

1.1. IP Networks:
A global computer network providing a variety of information and communication facilities becoming a part of real life. Normally, the Internet has yield only best effort service to all possible users without any deliberation to any requirements. A evolution of the Internet and use of current real-time multimedia services such as Real-time video and VoIP (Voice over IP) keep move upwards that have to lodge a distinct service levels and provide for QoS (Quality of Service) requirements [1]. Beneath the best effort approach, all packets are dealing with alike on a first-come-first-served basis and no scheme exists for differentiating the services for specific applications. Conventional Internet services effectively support applications such as HTTP (Hyper Text Transfer Protocol) or FTP (File Transfer Protocol) which are non-real-time multimedia services [2] [3] [4]. All-IP network not only bring down management costs and network deployment but also provide a great occasion able to be done for different recently developed services on the conventionally unconnected networks that are not possible. Nevertheless, running time-sensitive services such as VoIP on the packet switched networks may be affected by a below standard quality issues such as packet loss rate, large jitter and large delay [5]. Real-time traffic, such as interactive gaming, video conferencing and internet telephony [6] is time-dependent and need continuous information flow. Hence, there is a necessity to focus on QoS requirement for networks with mission-critical applications.

1.2. Traffic in IP networks:
Traffic engineering is a model, where network operator’s constraint the traffic and assign resources in sequence to reach goals, such as highest flow or smallest delay. One of the objectives is to permit distinct flows to share the network so that the entire flow will be maximized while maintaining fairness [7]. The simplicity and flexibility of IP protocol have led to the fact that “IP over everything”. But an intensive research is still under consideration how to provide best QoS using IP network for high priority traffic. Switching to IP as the main delivery method looks increasingly promising. Until now there is no standard method available including the application, transport and network layers for guaranteed QoS based on the application preference. Therefore, there is a need for recognized QoS framing and architecture to outline the application requirements in the available communication networks.
1.3. Resource Allocation in IP Networks:
Even though current communication networks, like the prevailing Internet, have come to be a considerable success in providing well organised data transmission services, e.g., electronic mail, web browsing and file transfer, it is even now a long way from enough hold up the rise demand for real-time services, e.g., audio, video and multimedia conveyance through the network. These real-time applications usually have strict Quality of Service (QoS) requirements and are sensitive to allocated bandwidth, time delay and packet loss ratio, which are not easy to be guaranteed in the TCP-based Internet service architecture [10] . Resource allocation in the multiservice communication networks presents a very important problem in the design of the future multi-class Internet. The future IP networks must carry a wide range of different traffic types being still able to provide performance guarantees to real-time sessions such as Voice over IP (VoIP), Video-on-Demand, (VoD)Video-Conferencing [11] [12]. Efficient and effective communication needs careful QoS design by means of appropriate resource allocation among competing for traffic flows with different service classes. On the other hand, for the future multi-class Internet, users will have to pay the network providers based on pricing strategies agreed in their Service-Level-Agreements. [13].

1.4. Research Objectives:
In previous work [14], a novel QoS based framework over the Fast IP reroute networks is proposed. It includes optimization of bandwidth with traffic shaping. A compression mechanism is proposed over the existing Multi Route Configurations (MRC) algorithm [4] for bandwidth optimization. When there is a huge amount of traffic in the network, existing MRC follows default scheduling which results in large transmission delay for high priority packets. In our previous work [15], an Agent based Monitoring and Scheduling Technique (AMST) is proposed which provides a generic framework for packet scheduling. AMST schedules the packets based on their priority and by selecting the transmission path with lowest transmission delay. Though both the works provide bandwidth optimization and scheduling for MRC, effective resource allocation and load balancing are to be provided in order to ensure QoS for all types of traffic. In BASMIN [3], a dynamic bandwidth allocation scheme along with load balancing is applied for various traffic priorities. But it fails to consider the congestion scenario that may occur at any link along a data transmission path. The bandwidth allocation becomes more complicated within the multipath network; where the origin rate is composed of add up to the path rates [10]. Moreover, the network priority level defined in the weight function is not so accurate. For this reason, in this paper, we design a Priority Based Bandwidth Allocation and Load Balancing (PBALB) technique for MRC [4]. The rest of the paper is organized as follows. In Section 2, we first discuss some related work with respect to Traffic shaping and resource allocation in IP networks, and in Section 3, we brief out about the design of PBALB technique with respect to Traffic Classification, Admission control and Path selection and Load balancing. The Experimental Results are discussed in Section 4. The concluding remarks are provided in Section 5.

II. RELATED WORKS

2.1 Traffic Shaping (Diffserv) in IP Networks:
Sladana Zoric and Melika Bolic [1] have proposed, a scheme of providing QoS (Quality of Service) within the DiffServ related IP network. DiffServ systematically provide QoS in networks of DiffServ shore up routers, since it is based on scalable IP based technology. Furthermore, the system shows the consequence on jitter and queuing delays of both traffic types when their WFQ weights differ and obtain and best weights that provide the optimal overall jitter and delay for real-time traffic.

Anh Nguyenand et al. [3] have proposed, to expand network user’s satisfaction by designing bandwidth allocation scheme using BASMIN technique. A metric is proposed for estimating the network user happiness degree network worth. A three-step estimation procedure was then direct to compare BASMIN outcome with other three approved bandwidth allotment strategy.
Shyan Hwang and et al. [6] have suggested a scheduler to provide QoS (Quality of Service) guarantees to Expedited Forwarding (EF), Assured Forwarding (AF) and Best Effort (BE) classes known as Adaptive Weighted Fair Queueing with the Priority (AWFQP). The objective of the scheduler was to make higher priority classes better than the lower priority classes.

Sanaa Sharafeddine and Zaher Dawy [8] have proposed an outcome approach to providing a eminent resilient network opposed to traffic divergence. Hence the distributions of real-time associations over an IP network enterprise and to show how ability of real-time traffic were affected when different resource constraint method were made used.

2.2 Allocation of Resource in IP Networks:
Throughput and providing fairness are two major goals of the network. The main challenge here is to allocate resource in such a manner that none of the flows are starving and gives satisfying levels of throughput.

Muhammad Naeem [2] have proposed a Session border controller (SBC) architecture, that actively keep track of the traffic data as well as connect with access network operator for getting user's SLA and arrange in advance resources as per. In addition adjustment in SIP invites and re-invite message to include requested information was used. These changes will be helpful in make sure availability of user's handset ability to IMS core network.

Miriam Allalouf and Yuval Shavitt [7] have presented an algorithm to locate an best and whole per commodity max-min fair rate vector in a polynomial number of course of action. In inclusion, they have shown a quick and give out an algorithm where each origin router could develop the routing and the fair rate allotment for its commodities while keeping the nearby best max-min fair allotment standard. The polynomial epsilon-approximation (FPTAS) algorithm is distributed algorithm and was fully based on a primal-dual alternation method.

Jiong Jin and et al. [10] have proposed the resource allocation and flow control issue as appeal to the general multipath communication networks with various applications. Furthermore the distributed algorithm, be revealed and show that included all the sources with optimistic grow and compelled utilities in the stable state, the utility max–min fairness was accomplish, which was crucial for stable Quality of Service (QoS) for distinct applications. By merging the first form Lagrangian technique and filtering method, the selected approach removes classic oscillation response in multipath networks and has a quick coincide property. In inclusion, the algorithm was able to resolve the optimal routing approach and pass the overall traffic fairly out of the available paths.

Zitoune Land et al. [11] have represented a sensitive constraint policy which accommodates the source bit rate to the set of resources in sequence to make sure performance assurance for multimedia applications. The technique called flatness based approach tracking concern with extreme traffic flow rate updates and restricts the traffic in sequence to respect the time limit.

III. PBALB Proposed Technique

3.1. Overview:
In this paper, we present a priority basis bandwidth allocation and load balancing techniques for MRC [4] as an additive to our earlier works [14] [15].

Figure 1 depicts the block diagram of the PBALB proposed technique. It consists of 3 phases. In phase 1, various types of traffic flows are classified and shaped by the edge routers in the DiffServ domain. In phase-2, admission control process is performed according to BASMIN. Here, the quality of convinced network appeal is enhanced with the assigned priorities. In phase-3, for path selection and load balancing, the path prices and path rates are estimated for all the paths and the source rate is adjusted using the source prices. The proposed solution will be implemented in NS-2 and the performance is evaluated in terms of assigned bandwidth, fairness index, throughput and delay.
3.2 System Model:

The DiffServ architecture transmits the traffic to the edge routers (ER) of the DiffServ domain. The core routers then forward the data packets as per the per hop behavior (PHP) related to each traffic class. ERs maintain all the user traffic profiles such that the traffic aligns with the Service Level Agreement (SLA) mentioned by a network operator.
3.2.1 Traffic Classification:
The architecture demonstrates per hop behavior (PHBs) that includes the EF, AF and BE.

a) Expedited Forwarding (EF):
This offers minimum-loss, minimum-latency, minimum-jitter, and guaranteed bandwidth service.
Priority queue with rate limited to the traffic class is considered.
VOIP and interactive gaming are some of the real-time applications involving EF. Assured Forwarding (AF):
This offers different forwarding assurances and PHBs define the following four traffic classes: Assured Forwarding 1, Assured Forwarding 2, Assured Forwarding 3 and Assured Forwarding 4.
For every class of traffic allocated with a specific portion of network bandwidth and buffer space to assure a precise Quality of Service.
The service profile is clearly defined as maximum drop precedence and dropped first when packet exceeds threshold during congestion.
The Assured Forwarding service is utilized by non-real time applications.

b) Best Effort (BE):
No service assured by PHBs.
The entire packets related to BE traffic is treated with equal fairness. BE traffic is forward in with equal fairness for entire packets.

3.2.3 Prioritizing the Traffic Class (PTC):

Algorithm 1: PTC

<table>
<thead>
<tr>
<th>Algorithm 1: PTC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Input: Packets with different traffic classes</td>
</tr>
<tr>
<td>Output: Assignment of Scheduler</td>
</tr>
<tr>
<td>Assign the Packets with different traffic classes into priority queuing scheduler (PQS)</td>
</tr>
<tr>
<td>If PQS = E</td>
</tr>
<tr>
<td>then AQS PQS Else</td>
</tr>
<tr>
<td>If PQS = BE</td>
</tr>
<tr>
<td>Then AQS FQS</td>
</tr>
<tr>
<td>Else</td>
</tr>
<tr>
<td>If Traffic Load &gt;Threshold Value</td>
</tr>
<tr>
<td>then BE shares bandwidth with AF based on FQS</td>
</tr>
<tr>
<td>AQS = FQS</td>
</tr>
</tbody>
</table>

Copyright to IJIRCCE DOI: 10.15680/IJIRCCE.2015.0401001
In above mention algorithm 1 initially the traffic class is fed into the priority queuing scheduler (PQS). In PQS scheduler, EF packets are provided with the higher priority. BE packets are then served with the condition that there are no pending EF packets to be transferred. Since AF contains its own bandwidth in Fair queuing scheduler (FQS), it is fed into it. Collating the traffic, the Adaptive Queuing Scheduler is applied, where the traffic load is compared with the traffic threshold value. If the traffic load is less than a traffic threshold, then the bandwidth offered to EF class is shared by the BE class as per the Priority Queuing Scheduler. On the other hand, if PQS results in BE, then AQS is changed to FQS. Otherwise, if Traffic load (TL) is greater than the threshold value, then BE shares bandwidth with AF class based on FQS and AQS is changed to PQS.

3.3 Admission Control:

Algorithm 2: Admission control:

Let B be the network path
Let AB be the available bandwidth
Let BWmin be the minimum bandwidth

The admission control is executed at the ER. It corresponds to a decision making the scenario where the network request is either accepted or rejected. When a new network connection request arrives at edge routers, it performs the following:

If B possesses ABW (ABW > BWmin)
Then ER accepts the new request
Else
ER analyzes if there is any existing path B
If yes,
Then ER accepts the new request
Else
ER rejects the new request.
End if
End if

In above algorithm 2, the decision is made assuming that the worth of the network request (NW) is less than the new value (NWnew). The network worth is estimated using the
following equation: \[ NW = 2^p U \] (1)

- \( p \) = network priority level (Estimated in section 3.2)
- \( U \) = Utility function
- \( 2^p \) = relative importance of the request.

### 3.4 Path Selection and Balancing Load:

For route selection and balancing load, the path prices and path rates are estimated for all the paths and the source rate is adjusted using the source prices.

During each time \( T+1 \), every link \((g)\) upgrade its link price \((O_g)\) using the following equation (2):

\[
O_g(T+1) = [O_g(T) + \xi (PR^g(T) Q_g)]
\] (2)

Where \( \xi \) = step size \((\xi > 0)\)

- \( PR^g(T) = R_g.PR \)
- \( T=Time \)
- \( R_g = link rate \)
- \( PR^g = aggregate pat\)h rate at link \( g \).

Eq. (2) reveals that if the combined path at link rate \( g \) goes beyond the capacity of link \( Q_g \), \( O_g \) will be get larger; it will be reduced otherwise.

For each source \( S \), first-form of Lagrangian algorithm is used to upgrade its rate of path:

\[
pr_{S,i}(T+1) = \left[ pr_{S,i}(T) + \xi \left( \frac{1}{A_i(pr_{S,i}(T))} + \delta_S(t) - \varepsilon_S(f) - Z_{S,i}(T) \right) \right]
\] (3)

Source rate is estimated using the following equation:

\[
pr_S(T+1) = \sum_{i=1}^{c} pr_{S,i}(T + 1)
\] (4)

where

\[
\delta_S(T+1) [\delta_S(T) \xi (f_S - pr_S(T))]
\] (5)

\[
\varepsilon_S(T+1) [\varepsilon_S(T) \xi (f_S - pr_S(T))]
\] (6)

The above equations are the smaller bound and higher bound price to impose the origin constaint
The Smaller bound and higher bound price are shown in above equation (5) and (6) impose the control on source $f_S \leq pr_S \leq F_S$

$$Z_{S,j}(T) = \max_{g \in K_S} (T)$$

(7)

The above equation reveals the route cost which includes the maximum of the link cost across the relevant route.

IV. EXPERIMENTAL RESULTS OF PBALB TECHNIQUE

4.1 Representational Model and Criteria:
Section 4.1, we brief out the performance of PBALB technique upon the NS-2 network simulator [16]. We collate our outcome with Multi Route Configurations (MRC) algorithm and BASMIN [3]. The design of network used in the simulation is illustrated in Figure 2. There are 26 nodes and frame works consists of main routers C1 and C2 and also contain 3 sets of ingress and egress routers indicated by IE1-EE1, IE2-EE2 and IE3-EE3 illustrated in Figure 2.

We use a combination of TCP, Video and CBR traffic flows. The size of packet is 512 bytes and entirely there are only 12 flows. The link delay and bandwidth are assigned with 10ms and 10Mb individually.

4.2 PBALB Performance Metrics:
The bottleneck bandwidth and rate of traffic are changed and We measure the following below specified metrics for 12 set of TCP, Video and CBR traffic flows in our PBALB simulation experiments.

a) Throughput: It is an amount of data received by all the receivers and measured in Mbits/sec.

b) Fairness: It is the share of bandwidth obtained by each receiver and measured in Mbits.

c) Delay: It is the end-to-end average delay obtained at all the receivers.

The outcome are narrated in the next section.

4.3. Case -1:
a) Fairness:
Fairness is a share of bandwidth among all nodes. The objective of this paper is to consider demand for QoS by different applications; it is not fair to directly share the bandwidth of the network. Rather, the network should share out the
bandwidth to the participating applications as stated by their different QoS utilities according to Equation NW = 2p U. Since purpose of our work is to research the impact of higher priority EF traffic on the fairness percentage by individual flow with respect to the equation.

CBR flow has the highest priority compared to video and TCP traffics according to prioritizing the traffic class in section 3.2. In QoS networks, CBR traffic has priority for its QoS requirement over non-real-time traffic TCP flow, which uses the bandwidth left unused by CBR flows. Hence, therefore, there is a possibility that CBR traffic with priority will cause TCP throughput degradation in QoS networks. Figure 4 and 5 manifest the PBALB fairness and BASMIN fairness index in percentage share of bandwidth obtained for individual CBR, TCP and Video traffic flows.

Figure 4: Fairness Index % for PBALB.

Figure 5: Fairness Index % for BASMIN.

Figure 6 and Figure 7 shows graph fairness share in (%) with respect to varying Traffic Rates for CBR, TCP and Video flows. Increasing the rate of each traffic flow will also cause overload and packet drop? Hence, the varying the rates of traffic as 1Mbit, 2Mbit, 3Mbit and 4Mbit keeping bottleneck bandwidth as 8Mb. PBALB in fig 6 obtains better fairness with respect to BASMIN in Figure 7.

Figure 6: Throughput for PBALB.

Figure 7: Throughput for BASMIN.
B. Throughput:
Figure 8, 9 and 10 shows throughput achieved by PBALB and BASMIN for 3 types of flows. The graph shows throughput measured on a scale of percentage with respect to varying bottleneck bandwidth. Throughput achieved by BASMIN is less compared to PBALB for all types of traffic flows. As the bottleneck bandwidth is increased, PBALB obtains 18% more throughput for CBR, 72% for TCP and 17.5% for Video flows, when compared to BASMIN. This is due to the fact that PBALB performs load balancing in addition to admission control.

C. Delay:
The Delay is evaluated by beginning transference of packets from origin to target end. In our simulation transmission delay measured in the sec with respect to varying bottleneck bandwidth 1Mb, 2Mb, 3Mb and 4Mb and also delay is measured with respect to varying transmission rates of 2Mbps, 3Mbps and 5Mbps. Simulation Results in Figure 11, 12, 13 and 14 shows transmission delays has been obtained for CBR and Video flows for PBALB and BASMIN. A delay of 0.2sec for CBR and 0.75sec for video flows exhibited by PBALB technique and a delay of 0.6sec for CBR and 1.5sec for video flows is exhibited by BASMIN technique. The delay increases linearly beyond 4Mb for both the techniques. However, PBALB has 78% lesser delay for CBR, 44% lesser delay for Video flows when compared to BASMIN, as it uses traffic shaping for reducing the queue waiting time.
4.4 Case-2:
In the second set of experiments, the PBALB technique is compared with MRC and the results are evaluated by varying the transmission rate. We vary the traffic rate as 1, 2, 3, 4 and 5Mb keeping the bottleneck bandwidth as 8Mb.
Figure 17: Rate Vs MRC Transmission Delay.  
Figure 18: Rate Vs PBALB Transmission delay.  

Figure 15 and 16 shows throughput obtained for PBALB and MRC, respectively. PBALB obtains more throughput for video and CBR flows when compared to BASMIN. This is due to the fact that PBALB performs load balancing and admission control as opposed to MRC. Figure 17 and 18 shows the delay obtained for MRC and PBALB. The delay increases linearly for MRC and increases beyond 3Mb for PBALB. However, PBALB has 88% lesser delay when compared to MRC, as it uses traffic shaping for reducing the queue waiting time. Figure 19 and 20 shows the fairness obtained for PBALB and MRC, respectively.

V. CONCLUSION
A PBALB technique has been proposed for MRC in this paper. Initially, the traffic flows are classified and shaped by the edge routes in the DiffServ domain. Then the admission control process is performed where the network request is either accepted or rejected. For path selection and load balancing, the path prices and path rates are estimated for all the paths and the source rate is adjusted using the source prices. Our simulation results show proposed technique enhances fairness, throughput and reduces transmission delay.
REFERENCES