FLOW BASED PRECEDENCE AND PREEMPTION METHODS WITHOUT A PRIORY SIGNALING

Derya Cansever
SI International
Reston, VA
and
Junaid Islam
Caspian
San Jose, CA

ABSTRACT

Precedence and Preemption (P&P) has its roots in the TDM voice networks, where it can be defined and implemented in a natural setting. Due to the hop-by-hop nature of IP networks, definition and implementation of P&P poses additional challenges. One of the main challenges is keeping the overall P&P architecture simple enough so that the management and operations of the networks do not become overly cumbersome, and that the end-to-end P&P requirements are satisfied across multiple autonomous domains. To this end, the DoD QoS Working Group has defined DoD wide use of the differentiated services code points (DSCP) to indicate the P&P level and the class of service associated with each packet. If P&P is based only on the DSCP value, this architecture may lead to inefficient use of network resources when the network is stressed.

Augmenting the information provided by the DSCP with a signaling protocol such as RSVP will overcome this inefficiency. But RSVP's scalability problems in backbone networks are well documented in RFC 2208. An alternative method is to treat the packets based on the micro-flow they belong to, in addition to their DSCP value. Advances in ASIC technology make it feasible to recognize and keep the state of the micro-flows for each router while maintaining scalability. Micro-flows are recognized based on attributes such as source and destination IP addresses and port numbers. In addition, other attributes such as packet sizes and packet inter-arrival times make it feasible to recognize the nature of the flows, e.g., VoIP, and thus allow for the allocation of the corresponding network resources without using signaling protocols. This paper will also discuss the application of flow identification techniques in DoD Black Core networks, where signaling and DSCP values may not be available.

INTRODUCTION

Support for Precedence and Preemption (P&P), also known as precedence-based assured service (PBAS), is a DoD requirement for all DoD owned and operated voice networks [1]. Work is underway to extend the P&P requirements to all Command and Control (C2) traffic [2]. The general approach adopted by the DoD Global Information Grid (GIG) networks is to use the Differentiated Services (DiffServ) approach for solving the P&P support problem. The DiffServ approach aggregates traffic from multiple users into broadly defined classes. We argue in Section 2 that DiffServ by itself is not sufficient for effective P&P support. In an ideal world, one would like to isolate every micro-flow, and provide for the appropriate P&P treatment on a per-flow basis. This approach would optimize the P&P goals, which can be stated as:

1. To assure that when the network is stressed, higher precedence messages will receive their QoS with higher probability than lower precedence messages
2. To assure that when the network is stressed, network resources are rationally utilized

The first goal is at the heart of the P&P. It requires that higher precedence traffic receive higher priority that is commensurate with its precedence level. The second goal is also important in that, when the network is stressed, (which is exactly the situation that makes the P&P mechanism a necessity,) the implementation of P&P should not waste any network resources. In other words, the P&P architecture should accommodate as many users as possible given the constraints of the network resources.

The first P&P goal is reachable using DiffServ. The differentiated services code point (DSCP) in IP Packets indicates the precedence level of the packets. This piece of information would enable the nodes in the transit path to implement the appropriate priority mechanisms, so that higher precedence mechanisms can receive their QoS with higher probability. The second P&P goal requires more elaborate mechanisms. We discuss in Section 2 that when the P&P architecture is based only on DiffServ tools, this architecture will yield to inefficient use of network resources. We also argue in the same section that, when the P&P traffic is treated on a micro-flow basis, it is possible to achieve the rational use of network resources. One way to implement micro-flow based treatment of P&P traffic is to use a signaling-based protocol such as RSVP [3]. Unfortunately, it is well known that signaling-based methods do not scale in backbone networks [4]. This paper discusses how advances in ASIC technology allow for the identification of individual flows in a network and the accounting of the flow state at each node. As a result, the micro-flow information can be used to devise a scalable P&P architecture that allows for rational use of network resources.

Section 2 discusses the shortcomings of relying solely on DiffServ to devise a P&P architecture. Section 3 reviews signaling-based approaches to P&P and the associated issues. Section 4 discusses the dynamic flow-based approach using ASIC technologies, and Section 5 indicates how this approach can be used to develop a scalable P&P architecture that makes rational use of network resources. Concluding remarks are found in Section 6.

DIFFSERV APPROACH TO P&P

Chairman, Joint Chiefs of Staff Instruction (CJCSI) 6215.01B specifies the QoS and precedence requirements for voice traffic in the circuit-switched Defense Switched Network (DSN) [1]. As such, there are 5 levels of Precedence:

- Flash Override (FO)
- Flash (F)
- Priority (P)
- Immediate (I)
- Routine (R)

There is a sixth precedence level called Flash Override Override (FOO). This category is employed only at the user level. From a network point of view, any message labeled as FOO will get the same treatment as FO. This circuit-switched based requirement can be directly extended to IP networks for real-time services with signaling, such as the ones supported by the SIP protocol. Extension of these requirements to all C2 traffic is still under development [2]. The Differentiated Services (DiffServ) field in IPv4 and IPv6 packets contains 6 bits. As such, it can support up to 64 distinct DSCP values. The DiffServ field is normally used to indicate the QoS requirements of the data carried in the IP packet. Provided that the scoping is done appropriately, the 6 bit DiffServ field can be used to indicate both the QoS and precedence requirements of the traffic carried in a packet. However, there are issues associated with relying solely on the DiffServ field for the purposes of P&P support. An immediate concern is that QoS and precedence are two different entities indicated by the same field. Specification of an optimal queuing architecture to satisfy both the QoS and precedence requirements is an open issue. The DoD PBAS decision paper [5] recognizes this issue, and proposes the configuration depicted in Figure 1 as a way to satisfy both QoS and precedence requirements.

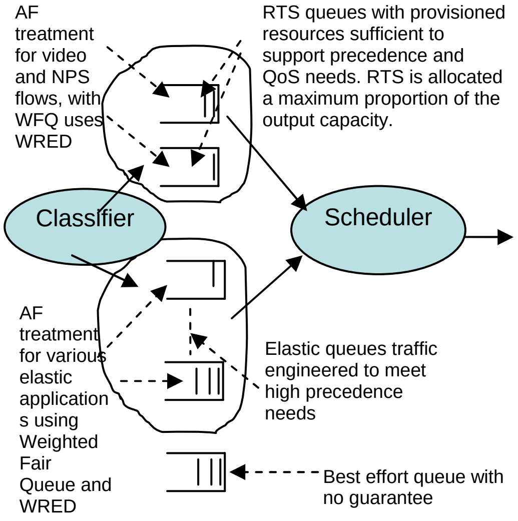
AF treatment for video and NPS flows, with WFQ uses WRED
RTS queues with provisioned resources sufficient to support precedence and QoS needs. RTS is allocated a maximum proportion of the output capacity.
Classifier
Scheduler
AF treatment for various elastic application s using Weighted Fair Queue and WRED
Elastic queues traffic engineered to meet high precedence needs
Best effort queue with no guarantee

Figure 1: Queuing Architecture for QoS and P&P Support (Source: MITRE)

In Figure 1, a Classifier places packets in one of the queues based on its DiffServ values, and a Scheduler picks which packet to serve at a given time. The above queuing architecture is only one of the alternatives, and there are at least four alternatives discussed in the latest version of the Decision Paper for PBAS. The choice of the optimal architecture for P&P support requires extensive modeling and simulation efforts and currently remains an open issue. This issue would be moot if the routers were able to identify each individual micro-flow and place packets belonging to a micro-flow in a separate queue. Then, the routers would allocate the appropriate network resources to each queue associated with the flows and execute the precedence requirements on an individual micro-flow basis.

There are other issues with the 'DiffServ only' P&P architecture. When the network is stressed, and preemption is required, this operation will be performed based on the DSCP value on a per-packet basis. As such, it does not discriminate between the end users that share the same DSCP value. In this case, the congested router may drop packets from several end users sharing the same DSCP value, and thus all of them would suffer. However, it is entirely possible that only one or two end users could be preempted, and the rest of the users sharing the same DSCP would continue to use the network resources. Since DSCP values do not discriminate against the end users, it is not possible to pre-empt a selected group of end users, and thus the second P&P goal cannot be satisfied.

When preemption occurs, it is desirable that the user that is preempted would be notified, so that it stops sending packets while the stress condition continues to exist. Such a mechanism would improve the use of network resources under stress conditions, as stipulated in the second P&P goal. Based on DSCP values only, this mechanism cannot be realized.

Another issue is related to precedence inversion in P&P. Precedence inversion occurs when a flow or flow aggregate does not receive its required QoS when a flow or flow aggregate at a lower precedence does receive its QoS. This can happen when traffic at a given precedence level exceeds its allocated network resources, and thus preemption occurs; while traffic at a lower precedence level is not affected because network resources allocated to this precedence level turns out to be adequate at this particular time. The precedence inversion is an artifact of the aggregation resulting from the DSCP-based treatment. Precedence inversion can be avoided if each micro-flow is identified and buffered in a separate (virtual) queue and allocated with its appropriate network resources, such as buffer space and share of the bandwidth in the outgoing link. In this case, since each micro-flow is allocated its own resources, precedence inversion will be avoided. This is because at each node, the network would have the ability to identify the micro-flows and their respective priorities, along with their QoS requirements. As such, the P&P implementation system will have the capability to allocate network resources on a per-micro-flow basis and according to their precedence, and thus avoid situations where lower priority users go through the network while higher priority users are blocked.

SIGNALING BASED APPROACHES

Discussion of Section 2 indicates that per micro-flow treatment of packets in P&P is desirable for rational use of network resources when the network is stressed. One way to ensure per micro-flow treatment is to signal the existence of flows and reserve the appropriate network resources along the path of the flow. RSVP is a standard protocol that can be used for this purpose [3]. However, it is well-known that RSVP does not scale in backbone networks [4]. Furthermore, there is a setup time delay associated with signaling. Except for low bandwidth access networks with a limited number of users, the use of RSVP is generally not feasible for the identification of micro-flows and allocation of micro-flow based network resources.

Another signaling based approach to P&P is to associate the SIP protocol with RSVP. In general, real-time applications such as voice over IP (VoIP) and video over IP (ViIP) make use of the SIP protocol [6]; this setup can be exploited for P&P implementation. The SIP server can examine each call request and determine their precedence value, as indicated in their DSCP field. Depending on the network congestion levels, the SIP server can reject and/or preempt SIP calls, thus satisfying the P&P requirements in a similar way to a circuit-switched network. One of the issues associated with this approach is that it is limited to SIP-based applications only. When the P&P requirements are extended to all C2 traffic, some of the high priority messages will not use SIP; thus this approach will not generalize. Scalability issues of RSVP still remain with this approach.

INTRODUCTION TO FLOW BASED QOS AND ROUTING

Conceptually, flow-based QoS is based on the simple observation that individual IP packets belong to a group (i.e. a micro-flow, or a group of micro-flows) and that having knowledge of a flow’s existence and characteristics enables better network performance. The concept of a flow-based router or traffic management device is not new, but until recently, it has been considered too complex and costly. However, the development of inexpensive memory and advances in ASIC technology allows packet processing at the hardware level to implement line rate scheduling of packets and to store flow-state information on a per-micro-flow basis, in the order of millions of micro-flows. Routing and traffic management products capable of 30 million flows over multiple 10 Gbps interfaces are commercially available today. Note that the ability of processing millions of micro-flows at the hardware level provides an infrastructure for micro-flow based P&P architecture.

An IP micro-flow is defined as a series of packets exhibiting the same values for a set of parameters known collectively as the 5-tuple. These five parameters include source IP address, destination IP address, IP protocol ID, source port address, and destination port address. Once a micro-flow is identified, a flow-based router can assign the micro-flow a QoS characteristic. The assigned QoS characteristic may be a function of the following:

- Specifics of the 5-tuple
- DiffServ code point (DSCP),
- VLAN tag
- Other properties, as defined by the network administrator.

The specific functions of a flow-based router include:

- Inspection. When a packet is received at a line card, it is inspected to determine whether it is part of a flow that has already been established across the platform.
- State Creation. If no information about this packet exists, information about the flow is extracted from the first packet, and a flow-state information block is created and stored locally. The flow-state information includes forwarding information taken from the forwarding information base (FIB) tables and QoS assignments, as a result of the QoS classification for that flow. The forwarding information may include:
  - Switch Fabric Route, Nexthop, Priority, Class, Maximum Rate, Minimum Rate, Shaping Enabled, Admission Control Enabled, Delay Variation, Burst Tolerance, Packets Sent / Received / Dropped, Bytes Sent / Received / Dropped, Duration, Bit Rate, Packet Rate, Packet Size
- Scheduling. In addition to forwarding information, flow-state information includes traffic type, rate information [available rate (AR) or guaranteed rate (GR)], switch fabric path, size of the packet, and the time that the last packet of the flow was sent.
- Processing. Traffic conditioning and traffic control are checked for this flow. Depending upon the checked results, packets are dropped or scheduled to be sent immediately or scheduled to be sent at a future time. If a packet does not arrive for a given flow, or a time-out occurs (time-out parameters are configurable), the flow is deemed inactive and the flow-state information is deactivated.

Once the IP micro-flow has been determined and classified, it is necessary to allocate a traffic type in order to assign a service level appropriate for that flow. This QoS allocation can be defined using the network management system, and possibly using a policy-based management (PBM) system. For example, the router may identify a flow with the following properties: It consists of a packet stream of constant packet size, with constant inter-arrival times. Such a flow would correspond to a VoIP session, and the router would allocate appropriate network resources, without having to make use of signaling protocols.

Note the processes of micro-flow identification and network resource (QoS) allocation do not require the use of signaling protocols. Rules for flow identification and the corresponding network resource allocation are pre-determined, and these parameters are configurable. The actions of micro-flow identification and QoS allocation are performed at the hardware level, scaling to millions of flows at aggregate speeds in the order of gigabits per second. Note that, short messages with high precedence would not necessarily benefit from this mechanism. High priority Messages that do not form a discernable flow, e.g. 1-2 packet long, will be identified based on their DSCP values alone. Since there is no need to reserve network resources to such messages, the fact that they are not identified as a flow does not create a problem.

FLOW BASED P&P IMPLEMENTATIONS WITHOUT SIGNALING

As discussed in the previous section, dynamic flow identification (DFI) allows the identification and classification of micro-flows based on traffic characteristics. The traffic characteristics of the micro-flow can be based on the following parameters:

- Duration of the micro-flow
- Average packet size of the micro-flow
- Average rate of the micro-flow
- Byte count of the micro-flow
- Pattern of packet inter-arrival gaps

Once a micro-flow is dynamically identified, the following traffic conditioning and traffic control can be applied:

- Network resource (QoS) allocation
- Remark DSCP at the egress

A typical example is to identify non-interactive peer-to-peer (P2P) traffic dynamically. When the P2P traffic is identified, the flow-based router will migrate the corresponding micro-flow to a class where rate limits apply, based on pre-determined network policies. In another example, suppose that the end user devices are not capable of marking the DSCP Field. DFI identifies VoIP traffic based on packet size and inter-packet time duration, and the DSCP is remarked to a value that upstream routers can treat adequately. Also, certain applications emanating from certain IP addresses can be assigned priorities based on pre-defined policies.

Since DFI allows for the identification of micro-flows and the proper allocation of network resources, P&P requirements can be met without having to use signaling techniques such as RSVP or the SIP servers. Therefore, DFI allows for scalable P&P implementation and extends the P&P concepts to all C2 traffic, including applications that do not use signaling.

When the GIG is deployed, it is plausible that signaling traffic will not be allowed through the encrypted Black Core. By the same token, there may be instances where DSCP values will not be exposed at the Black Core due to information assurance (IA) reasons. In this case, the DFI method described in this paper can be used to fulfill the P&P requirements without signaling and without having to inspect the DSCP field. Flow-based routers at the Black Core can leverage state information to extrapolate application information autonomously via DFI. They may also augment the state information with the addresses of ingress and egress HAIPE devices. As such, Black Core networks can devise limited P&P capabilities without signaling and without DSCP inspection. Figure 2 depicts an example of P&P in a Black Core network.

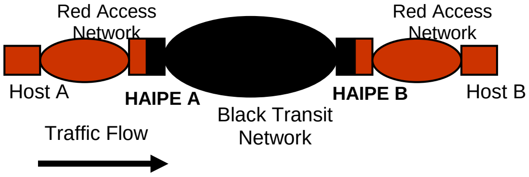

Figure 2: P&P in Black Core Networks

Suppose Host A generates high priority VoIP traffic destined for Host B. Red Access Network can implement P&P mechanisms using conventional methods, such as DSCP and/or RSVP. When the packets reach the HAIPE A device, the contents will be encrypted, including the source and destination IP addresses and the DSCP field. Instead, the IP addresses of HAIPE B and HAIPE A will be appended to the encrypted packet as the destination and source IP addresses, respectively. Flow-based routers at the Black Core can recognize the VoIP micro-flow based on traffic characteristics. Using this information in conjunction with the IP addresses of the HAIPE devices, it is possible to implement P&P mechanisms at the Black Core network without leaving any signature that can be recognized by a potential attacker. Of course, such a P&P mechanism will have to rely on information that includes:

- Packet size,
- Packet inter-arrival times,
- Duration of identified flows,
- Source HAIPE Address,
- Destination HAIPE Address,

As such, DFI based P&P at the Black Core will not be as granular as a P&P implementation in an unencrypted network. The efficiency of such a DFI based Black Core P&P mechanism is subject to further research.

CONCLUSIONS

If P&P is based only on the DSCP value, the architecture may lead to inefficient use of network resources when the network is stressed. Augmenting the information provided by the DSCP with a signaling protocol such as RSVP will overcome this inefficiency. But RSVP’s scalability problems in backbone networks are well documented in RFC 2208. Another approach is to associate the SIP protocol with RSVP. Non-signaled C2 traffic will not benefit from this approach. Furthermore, RSVP’s scalability issues still remain valid.

Ideally, one would like to treat the packets based on the micro-flow they belong to, in addition to their DSCP value. Advances in ASIC technology make it feasible to recognize and keep the state of the micro-flows for each router in a scalable way. Micro-flows are recognized based on attributes such as source and destination IP addresses and port numbers. In addition, other attributes such as packet size and packet inter-arrival times make it feasible to recognize the nature of the flows, e.g., VoIP, and thus allow for the allocation of the corresponding network resources without using signaling protocols. This architecture will allow for granular and scalable P&P implementation, which optimizes the use of network resources under stressed conditions. Furthermore, hardware-based flow identification techniques can also be used for secure P&P implementation in DoD Black Core networks, where signaling and DSCP information may not be allowed.

REFERENCES

[1] Chairman of the Joint Chiefs of Staff - Instructions (CJCSI), "Policy for Department of Defense Voice Networks," CJCSI 6215.01B, 23 September 2001.

[2] Chairman of the Joint Chiefs of Staff – Instructions (CJCSI), "Policy, Responsibilities, Processes, and Administration for the Department of Defense Global Information Grid (GIG) Networks," CJCSI 6215.02A (draft), 31 July 2004.

[3] Braden, B.; Zhang, L.; Berson, S.; Herzog, S.; and S. Jamin, "Resource ReSerVation Protocol (RSVP) -- Version 1 Functional Specification," RFC 2205, September 1997

[4] Mankin, A. et al, “Resource ReSerVation Protocol (RSVP) Version 1 Applicability Statement Some Guidelines on Deployment,” RFC 2208, September 1997.

[5] DoD Global Information Grid Net-Centric Implementation Document (NCID) Technical Decision Paper: GIG QoS – E2E: Precedence-based Assured Services (PBAS) June 2006.

[6] Rosenberg, J. et al, “SIP: Session Initiation Protocol,” RFC 3261, June 2002.