Transport Working Group   F. Baker
Internet-Draft   J. Polk
Updates: 4542,4594 (if approved)   Cisco Systems
Intended status: Informational   M. Dolly
Expires: April 2, 2007   AT&T Labs
                                                     September 29, 2006

An EF DSCP for Capacity-Admitted Traffic
draft-baker-tsvwg-admitted-voice-dscp-01

Status of this Memo

Copyright Notice

Abstract

This document requests a DSCP from the IANA for a class of real-time traffic conforming to the Expedited Forwarding Per Hop Behavior and admitted using a CAC procedure involving authentication, authorization, and capacity admission, as compared to a class of real-time traffic conforming to the Expedited Forwarding Per Hop Behavior but not subject to capacity admission or subject to very

coarse capacity admission.

One of the reasons behind this is the need for classes of traffic that are handled under special policies, such as the non-preemptive Emergency Telecommunication Service, the US DoD's Assured Service (which is similar to MLPP), or e-911. These do not need separate DSCPs or separate PHBs that are separate from each other, but they need a traffic class from which they can deterministically obtain their service requirements from including SLA matters.

Table of Contents

1. Introduction

This document requests a DSCP from the IANA for a class of real-time traffic conforming to the Expedited Forwarding [RFC3246][RFC3247] Per Hop Behavior and admitted using a CAC procedure involving authentication, authorization, and capacity admission, as compared to a class of real-time traffic conforming to the Expedited Forwarding Per Hop Behavior but not subject to capacity admission or subject to very coarse capacity admission.

One of the reasons behind this is the need for classes of traffic that are handled under special policies, such as the non-preemptive Emergency Telecommunication Service, the US DoD's Assured Service (which is similar to MLPP and uses preemption), or e-911, in addition to normal routine calls that use call admission. It is possible to use control plane protocols to generally restrict session admission such that admitted traffic should receive the desired service, and the policy (e.g., routine, NS/EP, e-911, etc) need not be signaled in a DSCP. However, service providers need to distinguish between special-policy traffic and other classes, particularly the existing VoIP services that perform no capacity admission or only very coarse capacity admission and can exceed their allocated resources.

This DSCP applies to the Telephony Service Class described in [RFC4594]. WIthin an ISP and on inter-ISP links (i.e., within networks whose internal paths are uniform at hundreds of megabits or faster), one would expect this traffic to be carried in the Real Time Traffic Class described in [I-D.ietf-tsvwg-diffserv-class-aggr].

1.1. Definitions

The following terms and acronyms are used in this document.

PHB: A Per-Hop-Behavior (PHB) is the externally observable forwarding behavior applied at a DS-compliant node to a DS behavior aggregate [RFC2475]. It may be thought of as a program configured on the interface of an Internet host or router, specified drop probabilities, queuing priorities or rates, and other handling characteristics for the traffic class.

DSCP: The Differentiated Services Codepoint (DSCP), as defined in [RFC2474], is a value which is encoded in the DS field, and which each DS Node MUST use to select the PHB which is to be experienced by each packet it forwards [RFC3260]. It is a 6-bit number embedded into the 8-bit TOS field of an IPv4 datagram or the Traffic Class field of an IPv6 datagram.

CAC: Call Admission Control, which includes concepts of authorization (an identified and authenticated user is determined to also be authorized to use the service) and capacity admission (at the present time, under some stated policy, capacity exists to support the call). In the Internet, these are separate functions, while in the PSTN they and call routing are carried out together.

UNI: A User/Network Interface (UNI) is the interface (often a physical link or its virtual equivalent) that connects two entities that do not trust each other, and in which one (the user) purchases connectivity services from the other (the network). Figure 1 shows two user networks connected by what appears to each of them to be a single network ("The Internet", access to which is provided by their service provider) which provides connectivity services to other users.

NNI: A Network/Network Interface (NNI) is the interface (often a physical link or its virtual equivalent) that connects two entities that trust each other within limits, and in which the two are seen as trading services for value. Figure 1 shows three service networks that together provide the connectivity services that we call "the Internet". They are different administrations and are very probably in competition, but exchange contracts for connectivity and capacity that enable them to offer specific services to their customers.

Queue: There are multiple ways to build a multi-queue scheduler. Weighted Round Robin (WRR) literally builds multiple lists and visits them in a specified order, while a calendar queue (often used to implement Weighted Fair Queuing,or WFQ) builds a list for each time interval and enqueues at most a stated amount of data in each such list for transmission during that time interval. While these differ dramatically in implementation, the external difference in behavior is generally negligible when they are properly configured. Consistent with the definitions used in the Differentiated Services Architecture [RFC2475], these are treated as equivalent in this document, and the lists of WRR and the classes of a calendar queue will be referred to uniformly as "queues".

```
                                        _.--------.
                                    ,-''           `--.
                                 ,-'                   `-.
           ,-------.           ,',-------.                `.
         ,'         `.       ,','         `.                `.
        /  User       \ UNI / /   Service   \                 \
       (    Network    +-----+    Network    )                 `.
        \             /  ;    \             /                    :
         `.         ,'   ;     `.         .+                    :
           '-------'    /        '-------'  \ NNI                \
                       ;                     \                     :
                       ;     "The Internet"   \  ,-------.         :
                      ;                        +'         `.        :
        UNI: User/Network Interface           /   Service   \       |
                     |                       (    Network    )      |
        NNI: Network/Network Interface        \             /       |
                      :                        +.         ,'        ;
                       :                      /  '-------'         ;
                       :                     /                     ;
           ,-------.    \        ,-------.  / NNI                 /
         ,'         `.   :     ,'         `+                     ;
        /  User       \ UNI   /   Service   \                    ;
       (    Network    +-----+    Network    )                 ,'
        \             /     \ \             /                 /
         `.         ,'       `.`.         ,'                ,'
           '-------'           `.'-------'                ,'
                                 `-.                   ,-'
                                    `--.           _.-'
                                        `--------''
```

Figure 1: UNI and NNI interfaces

1.2. Problem

In short, the Telephony Service Class described in [RFC4594] permits the use of capacity admission in implementing the service, but present implementations either provide no capacity admission services or do so in a manner that depends on specific traffic engineering. In the context of the Internet backbone, the two are essentially equivalent; the edge network is depending on specific engineering by the service provider that may not be present.

However, services are being requested of the network that would specifically make use of capacity admission, and would distinguish among users or the uses of available Voice-on-IP capacity in various ways. Various agencies would like to provide services as described in section 2.6 of [RFC4504] or in [RFC4190]. This requires the use of capacity admission to differentiate among users (which might be

911 call centers, other offices with preferential service contracts, or individual users gaining access with special credentials) to provide services to them that are not afforded to routine customer-to-customer IP telephony sessions.

1.3. Proposed Solution

The IETF is asked to differentiate, in the Telephony Service, between sessions that are originated without capacity admission or using traffic engineering and sessions that are originated using more robust capacity admission procedures. Sessions of the first type use a traffic class in which they compete without network-originated control as described in Section 2.2.1 or Section 2.2.2, and in the worst case lose traffic due to policing. Sessions of the second type cooperate with network control, and may be given different levels of preference depending on the policies that the network applies. In order to provide this differentiation, the IETF requests that the IANA assign a separate DSCP value to admitted sessions using the Telephony service (see Section 4).

2. Implementation of the Admitted Telephony Service Class

2.1. Potential implementations of EF in this model

There are at least two possible ways to implement the Expedited Forwarding PHB in this model. They are to implement separate classes as a set of

- o Multiple data plane traffic classes, each consisting of a policer and a queue, and the queues enjoying different priorities, or

- o Multiple data plane traffic classes, each consisting of a policer but feeding into a common queue or multiple queues at the same priority.

We will explain the difference, and describe in what way they differ in operation. The reason this is necessary is that there is current confusion in the industry, including a widely reported test for NS/EP services that implemented the policing model and described it as an implementation of the multi-priority model, and discussion in other environments of the intermixing of voice and video traffic at relatively low bandwidths in the policing model.

The multi-priority model is shown in Figure 2. In this model, traffic from each service class is placed into a separate priority queue. If data is present in both queues, traffic from one of them will always be selected for transmission. This has the effect of

transferring jitter from the higher priority queue to the lower priority queue, and reordering traffic in a way that gives the higher priority traffic a smaller average queuing delay. Each queue must have its own policer, however, to protect the network from errors and attacks; if a traffic class thinks it is carrying a certain data rate but an abuse sends significantly more, the effect of simple prioritization would not preserve the lower priorities of traffic, which could cause routing to fail or otherwise impact an SLA.

```
                                             .
                     policers    priorities  |`.
           EF ---------> <=> ----------||----+  `.
                                         high|    `.
           EF2---------> <=> ----------||----+     .'-----------
                           .             medium  .'
              rate queues  |`.         +-----+ .' Priority
           AF1------>||----+  `.      /  low |'   Scheduler
                           |    `.   /
           AF2------>||----+     .'-+
                           |   .'
           CS0------>||----+ .' Rate Scheduler
                           |'   (WFQ, WRR, etc)
```

Figure 2: Implementation as a data plane priority

The multi-policer model is shown in Figure 3. In this model, traffic from each service class is policed according to its SLA requirements, and then placed into a common priority queue. Unlike the multi-priority model, the jitter experienced by the traffic classes in this case is the same, as there is only one queue, but the sum of the traffic in this higher priority queue experiences less average jitter than the elastic traffic in the lower priority.

```
                       policers    priorities  .
             EF ---------> <=> -------\        |`.
                                       --||----+  `.
             EF2---------> <=> -------/    high|    `.
                             .                 |     .'--------
                rate queues  |`.         +-----+   .'
             AF1------>||----+  `.      /  low | .' Priority
                             |    `.   /       |'   Scheduler
             AF2------>||----+     .'-+
                             |   .'
             CS0------>||----+ .' Rate Scheduler
                             |'   (WFQ, WRR, etc)
```

Figure 3: Implementation as a data plane policer

The difference between the two operationally is, as stated, the issues of loss due to policing and distribution of jitter.

If the two traffic classes are, for example, voice and video, datagrams containing video data are relatively large (generally the size of the path MTU) while datagrams containing voice are relatively small, on the order of only 40 to 200 bytes, depending on the codec. On lower speed links (less than 10 MBPS), the jitter introduced by video to voice can be disruptive, while at higher speeds the jitter is nominal compared to the jitter requirements of voice. At access network speeds, therefore, [RFC4594] recommends separation of video and voice into separate queues, while at optical speeds [I-D.ietf-tsvwg-diffserv-class-aggr] recommends that they use a common queue.

If, on the other hand, the two traffic classes are carrying the same type of application with the same jitter requirements, then giving one preference in this sense does not benefit the higher priority traffic and may harm the lower priority traffic. In such a case, using separate policers and a common queue is a superior approach.

2.2. Capacity admission control

There are five major ways that capacity admission is done or has been proposed to be done in Internet Voice applications:

- o Capacity admission control by assumption,
- o Capacity admission control by call counting,
- o End-point capacity admission performed by probing the network,
- o Centralized capacity admission control, and
- o Distributed capacity admission control

2.2.1. Capacity admission control by assumption

The first approach is to ignore the matter entirely. If one assumes that the capacity available to the application is uniformly far in excess of its requirements, it is perhaps overhead that can be ignored. This assumption is currently made in Internet VoIP offerings such as Skype and Vonage; the end user is responsible to place his service on a LAN connected to the Internet backbone by a high speed broadband connection and use capable ISPs to deliver the service. There is an authorization step in the sense that the service ensures that the user pays his bills, but no capacity admission is considered because there is a clear separation from the

voice application service provider admitting the calls and the access network provider admitting the traffic. The two have no way of knowing about each other, except maybe in the abstract sense.

2.2.2. Capacity admission control by call counting

The H.323 gatekeeper, originally specified in 1996, operates on the model that the considerations of Section 2.2.1 generally apply, and that it is therefore sufficient to count calls in order to ensure that any bottlenecks in the network are never overloaded. . Which phone is calling which phone is configured information into the Gatekeeper, ensuring it doesn't admit too many calls across a low speed link. The area of influence of a Gatekeeper is called a Zone, and limits how far away a Gatekeeper can influence calls. This is because call counting doesn't scale when more than one server is admitting flows across the same limited speed links. This approach is consistent with the original design of H.323, which in 1996 was a mechanism for connecting H.320 media gateways across a LAN. VoIP has come a long way since then, however, and the engineering trade-offs this approach requires in complex networks are unsatisfactory.

SIP provides the option to go down another path, to admit its servers at layer 7, have no awareness of lower layer connectivity, resulting in a divorce from infrastructure knowledge - save for [RFC3312], which binds the two, but only at the endpoints.

In short, if there is a bottleneck anywhere in the network that might be used to connect two gatekeepers, SIP proxies that do not implement or do not configure the use of [RFC3312], or other call management systems, the amount of traffic between the two must be contained below that bottleneck even if the normal path is of much higher bandwidth. In addition, the multiplexing of traffic streams between different pairs of gatekeepers over a common LAN infrastructure is not handled by the application, and so must be managed in the engineering of the network.

2.2.3. End-point capacity admission performed by probing the network

[I-D.briscoe-tsvwg-cl-architecture] is one of many proposals that have looked at probing of the network by the end system to determine its capacity to accept a new session. Such proposals have been made a number of times by the likes of NTT Labs, UIUC researchers, Cisco Systems (which used its Service Assurance Architecture to probe capacity using pings and report when network delay variability increased), and others. Many of the proposals tested in research have fared reasonably well in high bandwidth environments where actual network congestion is unusual, but have not scaled down to slower access links.

The problem has been, in essence, that variable rate codecs can be on the quiet side of the average for lengthy periods of time and then become noisier. New sessions can be disrupted or disrupt existing sessions if they perform their capacity admission procedures at a quiet time and find themselves overrunning the allocated capacity during a noisy time. In addition, for a service in which the network must exercise control and differentiate among users, the users cannot be depended on to differentiate among themselves in the network's favor. The network must manage that service.

For this reason, [I-D.briscoe-tsvwg-cl-architecture] is only proposed as a solution within backbone networks, leaving access networks to provide other forms of capacity admission, and more generally such techniques are only recommended in high bandwidth contexts. What is not addressed, is when these quite times become not-at-all quite due to some event occurring, leading to great amounts of traffic. A means of maintaining existing critical calls is essential to retain a given service. Times of disaster can be such times of extreme bursts of the number of call attempts. Once a call is established, that call needs to be retained.

2.2.4. Centralized capacity admission control

The concept of a Bandwidth Broker was first discussed in the research world surrounding ESNET and Internet II in the late 1990's, and has been discussed in the literature pertaining to the Differentiated Services Architecture [RFC2475]. It is, in short, a central system that performs a variety of services on behalf of clients of a network including applying AAA services (as in [RFC2904])and authorizing them to use specified capacity at specified times. Its strength is that it is relatively simple, at least in concept, and can keep track of simple book-keeping functions apart from network elements such as routers. Its weakness is that it has no idea what the specific routing of any stated data flow is, or its capacity apart from services such as MPLS Traffic Engineering or engineering assumptions specified by the designers of a network, and obtaining that information from the network via SNMP GET or other network management action can impose a severe network overhead, and is obviously not in real-time.

For scaling reasons, operational Bandwidth Brokers generally take on a semi-distributed or fully distributed nature. They are implemented on a per-point-of-presence basis, and in satellite networks might be implemented in each terminal. At this point, they become difficult to operationally distinguish from distributed capacity admission services such as described in Section 2.2.5.

2.2.5. Distributed capacity admission control

The IETF developed the Integrated Services Model [RFC1633]and the RSVP capacity admission protocol [RFC2205] in the early 1990's, and then integrated it with the Differentiated Services Architecture in [RFC2998]. Since then, the IETF has worked to describe a next generation capacity admission protocol, which is calls NSIS, and which is limited in scope to considering unicast sessions. [RFC4542] looked at the issue of providing preferential services in the Internet, and determined that RSVP with its defined extensions could provide those services to unicast and multicast sessions.

As with the Bandwidth Broker model, there are concerns regarding scaling, mentioned in [RFC2208]. Present implementations that have been measured have been found to not display the scaling concerns, however, and in any event the use of RSVP Aggregation enables the backbone to handle such sessions in a manner similar to an ATM Virtual Path, bundling sessions together for capacity management purposes.

2.3. Prioritized capacity admission control

Emergency Telecommunication Service, the US DoD's Assured Service, and e-911 each call for some form of prioritization of some calls over others. Prioritization of the use of bandwidth is fundamentally a matter of choices - at a point where one has multiple choices, applying a policy that selects among them. In the PSTN, GETS operates in favor of an authorized caller either by routing a call that would otherwise be refused by a path unavailable to the general public or by queuing the call until some existing call completes and bandwidth becomes available. e-911 is similar, but the policy is based on the called party, the emergency call center. MLPP operates by preempting an existing call to make way for the new one.

In the Internet, routing is not performed on a per-call basis, so, apart from interconnections to the PSTN, re-routing isn't an option. On the other hand, in the Internet there are more classes of traffic than in the PSTN. In the PSTN, all calls are uses of circuits, while in the Internet some bandwidth is always reserved for elastic applications - at least, it must be available for routing, and there is generally significant consideration of the web, instant messaging, and other applications. In essence, any capacity admission policy that differentiates between calls has the option of temporarily borrowing bandwidth from the capacity reserved for elastic traffic by accepting new sessions under some prioritized policy while refusing sessions of lower priority because the threshold at that priority has been reached.

For example, regardless of the type of capacity admission that is used (apart from "no admission process"), one might admit prioritized sessions using a higher bandwidth threshold than one admits lower priority sessions.

If capacity admission as described in Section 2.2.2 is in use, the thresholds must be set low enough that bandwidth would be available anywhere in the network. This greatly limits the utility of such a service due to the level of bandwidth waste that results.

If capacity admission as described in Section 2.2.3 is in use, then either multiple thresholds must be applied in marking the traffic, multiple traffic marks must be applied, or there must be multiple ways to interpret the result. In any event, this is only applicable in domains in which the law of large numbers applies.

If capacity admission as described in Section 2.2.4 is in use, thresholds can be applied related to a general policy or SLA, or related to the network ingress and egress in use. It requires them to maintain state regarding network traffic routing separate from the network; to the extent that is variable, it requires direct monitoring in the OSS.

If capacity admission as described in Section 2.2.5 is in use, thresholds can be applied to the critical points of the path that the traffic in question actually takes because one is asking the equipment that the path traverses.

3. Recommendations on implementation of an Admitted Telephony Service Class

It is the belief of the authors that either data plane PHB described in Section 2.1, if coupled with adequate AAA and capacity admission procedures as described in Section 2.2.5, are sufficient to provide the services required for an Admitted Telephony service class. If preemption is in view, as described in section 2.3.5.2 or [RFC4542], this provides the tools for carrying out the preemption. If preemption is not in view, or in addition to preemptive services, the application of different thresholds depending on call precedence has the effect of improving the probability of call completion by admitting preferred calls at a time that other calls are being refused. Routine and priority traffic can be admitted using the same DSCP value, as the choice of which calls are admitted is handled in the admission procedure executed in the control plane, not the policing of the data plane.

On the point of what protocols and procedures are required for

authentication, authorization, and capacity admission, we note that clear standards do not at this time exist for bandwidth brokers, NSIS has not at this time been finalized and in any event is limited to unicast sessions, and that RSVP has been standardized and has the relevant services. We therefore recommend the use of RSVP at the UNI. Procedures at the NNI are business matters to be discussed between the relevant networks, and are recommended but not required.

4. IANA Considerations

This note, fundamentally, requests IANA the assign a DSCP value to a second EF traffic class consistent with [RFC3246] and [RFC3247] and implementing the Telephony Service Class described in [RFC4594] at lower speeds and [I-D.ietf-tsvwg-diffserv-class-aggr] at higher speeds. This new traffic class requires the use of capacity admission such as RSVP services together with AAA services at the User/Network Interface (UNI); the use of such services at the NNI is at the option of the interconnected networks. The recommended value for the code point 101100, paralleling the EF code point, which is 101110, and both of which are allocated from Pool 1 as described in [RFC2474].

The code point should be referred to as EF-ADMIT.

5. Security Considerations

A major requirement of this service is effective use of a signaling protocol such as RSVP, with the capabilities to identify its user either as an individual or as a member of some corporate entity, and assert a policy such as "routine" or "priority".

This capability, one has to believe, will be abused by script kiddies and others if the proof of identity is not adequately strong or if policies are written or implemented improperly by the carriers. This goes without saying, but this section is here for it to be said...

6. Acknowledgements

7. References

7.1. Normative References

[I-D.ietf-tsvwg-diffserv-class-aggr]
           Chan, K., "Aggregation of DiffServ Service Classes",

draft-ietf-tsvwg-diffserv-class-aggr-00 (work in progress), June 2006.

[RFC2474] Nichols, K., Blake, S., Baker, F., and D. Black, "Definition of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Headers", RFC 2474, December 1998.

[RFC2475] Blake, S., Black, D., Carlson, M., Davies, E., Wang, Z., and W. Weiss, "An Architecture for Differentiated Services", RFC 2475, December 1998.

[RFC2998] Bernet, Y., Ford, P., Yavatkar, R., Baker, F., Zhang, L., Speer, M., Braden, R., Davie, B., Wroclawski, J., and E. Felstaine, "A Framework for Integrated Services Operation over Diffserv Networks", RFC 2998, November 2000.

[RFC3246] Davie, B., Charny, A., Bennet, J., Benson, K., Le Boudec, J., Courtney, W., Davari, S., Firoiu, V., and D. Stiliadis, "An Expedited Forwarding PHB (Per-Hop Behavior)", RFC 3246, March 2002.

[RFC3247] Charny, A., Bennet, J., Benson, K., Boudec, J., Chiu, A., Courtney, W., Davari, S., Firoiu, V., Kalmanek, C., and K. Ramakrishnan, "Supplemental Information for the New Definition of the EF PHB (Expedited Forwarding Per-Hop Behavior)", RFC 3247, March 2002.

[RFC3260] Grossman, D., "New Terminology and Clarifications for Diffserv", RFC 3260, April 2002.

[RFC4542] Baker, F. and J. Polk, "Implementing an Emergency Telecommunications Service (ETS) for Real-Time Services in the Internet Protocol Suite", RFC 4542, May 2006.

7.2. Informative References

[I-D.briscoe-tsvwg-cl-architecture]
Briscoe, B., "An edge-to-edge Deployment Model for Pre-Congestion Notification: Admission Control over a DiffServ Region", draft-briscoe-tsvwg-cl-architecture-03 (work in progress), June 2006.

[RFC1633] Braden, B., Clark, D., and S. Shenker, "Integrated Services in the Internet Architecture: an Overview", RFC 1633, June 1994.

[RFC2205] Braden, B., Zhang, L., Berson, S., Herzog, S., and S.

Jamin, "Resource ReSerVation Protocol (RSVP) -- Version 1 Functional Specification", RFC 2205, September 1997.

[RFC2208] Mankin, A., Baker, F., Braden, B., Bradner, S., O'Dell, M., Romanow, A., Weinrib, A., and L. Zhang, "Resource ReSerVation Protocol (RSVP) Version 1 Applicability Statement Some Guidelines on Deployment", RFC 2208, September 1997.

[RFC2904] Vollbrecht, J., Calhoun, P., Farrell, S., Gommans, L., Gross, G., de Bruijn, B., de Laat, C., Holdrege, M., and D. Spence, "AAA Authorization Framework", RFC 2904, August 2000.

[RFC3312] Camarillo, G., Marshall, W., and J. Rosenberg, "Integration of Resource Management and Session Initiation Protocol (SIP)", RFC 3312, October 2002.

[RFC4190] Carlberg, K., Brown, I., and C. Beard, "Framework for Supporting Emergency Telecommunications Service (ETS) in IP Telephony", RFC 4190, November 2005.

[RFC4504] Sinnreich, H., Lass, S., and C. Stredicke, "SIP Telephony Device Requirements and Configuration", RFC 4504, May 2006.

[RFC4594] Babiarz, J., Chan, K., and F. Baker, "Configuration Guidelines for DiffServ Service Classes", RFC 4594, August 2006.

Authors' Addresses

Fred Baker
Cisco Systems
Santa Barbara, California 93117
USA

Phone: +1-408-526-4257
Email: fred@cisco.com

James Polk
Cisco Systems
Richardson, Texas 75082
USA

Phone: +1-817-271-3552
Email: jmpolk@cisco.com

Martin Dolly
AT&T Labs
Middletown Township, New Jersey 07748
USA

Phone: +1-732-420-4574
Email: mdolly@att.com

Full Copyright Statement

Intellectual Property

Acknowledgment

Funding for the RFC Editor function is provided by the IETF Administrative Support Activity (IASA).