©À

Transport Working Group
Internet-Draft
Expires: April 2, 2005

F. Baker
J. Polk
Cisco Systems

October 2, 2004

Implementing MLPP for Voice and Video in the Internet Protocol Suite
draft-baker-tsvwg-mlpp-that-works-02

Status of this Memo

Copyright Notice

Abstract

The Defense Information Systems Agency of the United States Department of Defense, with its contractors, has proposed a service architecture for military (NATO and related agencies) telephone systems. This is called the Assured Service, and is defined in two documents: "Architecture for Assured Service Capabilities in Voice over IP" and "Requirements for Assured Service Capabilities in Voice over IP". Responding to these are two documents: "Extending the

Session Initiation Protocol Reason Header to account for Preemption Events", "Communications Resource Priority for the Session Initiation Protocol".

What remains to this specification is to provide a Call Admission Control procedure and a Per Hop Behavior for the data which meet the needs of this architecture. Such a CAC procedure and PHB is appropriate to any service that might use H.323 or SIP to set up real time sessions. These obviously include but are not limited to Voice and Video applications, although at this writing the community is mostly thinking about Voice on IP and many of the examples in the document are taken from that environment.

In a network where a call that is permitted initially and is not denied or rejected at a later time, call and capacity admission procedures performed only at the time of call setup may be sufficient. However in a network where sessions ÇÖ status can be reviewed by the network and preempted or denied due to changes in routing (when the new routes lack capacity to carry calls switched to them) or changes in offered load (where higher precedence calls supercede existing calls), maintaining a continuing model of the status of the various calls is required.

Table of Contents

1. Overview

The Defense Information Systems Agency of the United States Department of Defense, with is contractors, has proposed a service architecture for military (NATO and related agencies) telephone systems. This is called the Assured Service, and is defined in two documents: [I-D.pierce-ieprep-assured-service-arch] and [I-D.pierce-ieprep-assured-service-req]. Responding to these are two documents: [I-D.ietf-sipping-reason-header-for-preemption] and [I-D.ietf-sip-resource-priority].

What remains to this specification is to provide a Call Admission Control procedure and a Per Hop Behavior for the data which meet the needs of this architecture. Such a CAC procedure and PHB is appropriate to any service that might use H.323 or SIP to set up real time sessions. These obviously include but are not limited to Voice and Video applications, although at this writing the community is mostly thinking about Voice on IP and many of the examples in the document are taken from that environment.

In a network where a call that is permitted initially and is not denied or rejected at a later time, call and capacity admission procedures performed only at the time of call setup may be sufficient. However in a network where sessions ÇÖ status can be reviewed by the network and preempted or denied due to changes in routing (when the new routes lack capacity to carry calls switched to them) or changes in offered load (where higher precedence calls supercede existing calls), maintaining a continuing model of the status of the various calls is required.

1.1 Multi-Level Preemption and Precedence

Before doing so, however, let us discuss the problem that MLPP is intended to solve and the architecture of the system. The Assured Service is designed as an IP implementation of an existing ITU-T/ NATO/DoD telephone system architecture known as [ITU.MLPP.1990][ANSI.MLPP.Spec][ANSI.MLPP.Supplement], or MLPP. MLPP is an architecture for a prioritized call handling service such that in times of emergency in the relevant NATO and DoD commands, the relative importance of various kinds of communications is strictly defined, allowing higher precedence communication at the expense of lower precedence communications. These precedences, in descending order, are:

Flash Override Override: used by the Commander in Chief, Secretary of Defense, and Joint Chiefs of Staff, Commanders of combatant commands when declaring the existence of a state of war. Commanders of combatant commands when declaring Defense Condition

One or Defense Emergency or Air Defense Emergency and other national authorities that the President may authorize in conjunction with Worldwide Secure Voice Conferencing System conferences. Flash Override Override cannot be preempted. This precedence level is not enabled on all DoD networks.

Flash Override: used by the Commander in Chief, Secretary of Defense, and Joint Chiefs of Staff, Commanders of combatant commands when declaring the existence of a state of war. Commanders of combatant commands when declaring Defense Condition One or Defense Emergency and other national authorities the President may authorize. Flash Override cannot be preempted in the DSN.

Flash: reserved generally for telephone calls pertaining to command and control of military forces essential to defense and retaliation, critical intelligence essential to national survival, conduct of diplomatic negotiations critical to the arresting or limiting of hostilities, dissemination of critical civil alert information essential to national survival, continuity of federal government functions essential to national survival, fulfillment of critical internal security functions essential to national survival, or catastrophic events of national or international significance.

Immediate: reserved generally for telephone calls pertaining to situations that gravely affect the security of national and allied forces, reconstitution of forces in a post-attack period, intelligence essential to national security, conduct of diplomatic negotiations to reduce or limit the threat of war, implementation of federal government actions essential to national survival, situations that gravely affect the internal security of the nation, Civil Defense actions, disasters or events of extensive seriousness having an immediate and detrimental effect on the welfare of the population, or vital information having an immediate effect on aircraft, spacecraft, or missile operations.

Priority: reserved generally for telephone calls requiring expeditious action by called parties and/or furnishing essential information for the conduct of government operations.

Routine: designation applied to those official government communications that require rapid transmission by telephonic means but do not require preferential handling.

The rule, in MLPP, is that more important calls override less important calls when congestion occurs within a network. Station based preemption is used when a more important call needs to be placed to either party in an existing call. Trunk based preemption

is used when trunk bandwidth needs to be reallocated to facilitate a higher precedence call over a given path in the network. In both station and trunk based preemption scenarios, preempted parties are positively notified, via preemption tone, that their call can no longer be supported. The same preemption tone is used, regardless of whether calls are terminated for the purposes of station of trunk based preemption. The remainder of this discussion focuses on trunk based preemption issues.

MLPP is built as a proactive system in which callers must assign one of the precedence levels listed above at call initiation; this precedence level cannot be changed throughout that call. If an elevated status is not assigned by a user at call initiation time, the call is assumed to be "routine". If there is end to end capacity to place a call, any call may be placed at any time. However, when any trunk (in the circuit world) or interface (in an IP world) reaches a utilization threshold, a choice must be made as to which calls to accept or allow to continue. The system will seize the trunks or bandwidth necessary to place the more important calls in preference to less important calls by preempting an existing call (or calls) of lower precedence to permit a higher precedence call to be placed.

More than one call might properly be preempted if more trunks or bandwidth is necessary for this higher precedence call. A video call (perhaps of 384 KBPS, or 6 trunks) competing with several lower precedence voice calls is a good example of this situation.

1.2 Definition of Call Admission

Traditionally, in the PSTN, "Call Admission Control", or CAC, has had the responsibility of determining whether a caller has permission (an identified subscriber, with identify attested to by appropriate credentials, is authorized) to use an available circuit. MLPP, or any emergency telephone service, creates two feedback paths in the algorithm: if a caller is authorized to use a higher precedence and is asserting that the advanced precedence applies to a given call, he may also be authorized to use other networks, or the PSTN may be obligated to preempt a call if possible and necessary to create appropriate bandwidth, or it may be authorized to use a guard band of bandwidth that other callers are not. At the completion of CAC, however, the caller either has a circuit that he or she is authorized to use, or has no circuit. Since the act of preemption or consideration of alternative bandwidth sources is part and parcel of the problem of providing bandwidth, the authorization step in bandwidth provision also affects the choice of networks that may be authorized to be considered. The three cannot be separated. The CAC procedure finds available bandwidth that the caller is authorized to

use and preemption may in some networks be part of making that happen.

1.3 Assumptions about the Network

IP networks generally fall into two categories: those with constrained bandwidth, and those that are massively overprovisioned. In a network wherein over any interval that can be measured (including sub-second intervals) capacity exceeds offered load by at least 2:1, the jitter and loss incurred in transit are nominal. This is generally a characteristic of properly engineered Ethernet LANs and of optical networks (networks that measure their link speeds in multiples of 51 MBPS); in the latter, circuit-switched networking solutions such as ATM, MPLS, and GMPLS can be used to explicitly place routes, and so improve the odds a bit.

Between those networks, in places commonly called "inter-campus links", "access links" or "access networks", for various reasons including technology and cost, it is common to find links whose offered load can approximate or exceed the available capacity. Such events may be momentary, or may occur for extended periods of time.

In addition, primarily in tactical deployments, it is common to find bandwidth constraints in the local infrastructure of networks. For example, the US Navy's network afloat connects approximately 300 ships, via satellite, to five network operation centers, and those NOCs are in turn interconnected via the DISA backbone. A typical ship may have between two and six radio systems aboard, often at speeds of 64 KBPS or less. In US Army networks, current radio technology likewise limits tactical communications to links below 100 KBPS.

Over this infrastructure, military communications expect to deploy voice communication systems (30-80 KBPS per session), video conferencing using MPEG 2 (3-7 MBPS) and MPEG 4 (80 KBPS to 800 KBPS), in addition to traditional mail, file transfer, and transaction traffic.

1.4 Assumptions about application behavior

Parekh and Gallagher published a series of papers [Parekh1][Parekh2] analyzing what is necessary to ensure a specified service level for a stream of traffic. In a nutshell, they showed that to predict the behavior of a stream of traffic in a network, one must know two things:

- o the rate and arrival distribution with which traffic in a class is introduced to the network, and

- o what network elements will do, in terms of the departure distribution, injected delay jitter and loss characteristics, with the traffic they see.

For example, TCP tunes its effective window (the amount of data it sends per round trip interval) so that the ratio of the window and the round trip interval approximate the available capacity in the network. As long as the round trip delay remains roughly stable and loss is nominal (which are primarily behaviors of the network), TCP is able to maintain a predictable level of throughput. In an environment where loss is random or in which delays wildly vary, TCP behaves in a far less predictable manner.

Voice and video systems do not tune their behavior to that of the network. Rather, they send traffic at a rate specified by the codec depending on what it perceives is required. In an MPEG-4 system, for example, if the camera is pointed at a wall, the codec determines that an 80 KBPS data stream will describe that wall, and issues that amount of traffic. If a person walks in front of the wall or the camera is pointed an a moving object, the codec may easily send 800 KBPS in its effort to accurately describe what it sees. In commercial broadcast sports, which may line up periods in which advertisements are displayed, the effect is that traffic rates suddenly jump across all channels at certain times because the eye-catching ads require much more bandwidth than the camera pointing at the green football field.

As described in [RFC1633], when dealing with a real-time application, there are basically two things one must do to ensure Parekh's first requirement. To ensure that one knows how much offered load the application is presenting, one must police (measure load offered and discard excess) traffic entering the network. If that policing behavior has a debilitating effect on the application, as non-negligible loss has on voice or video, one must admit sessions judiciously according to some policy. A key characteristic of that policy must be that the offered load does not exceed the capacity dedicated to the application.

In the network, the other thing one must do is ensure that the application's needs are met in terms of loss, variation in delay, and end to end delay. One way to do this is to supply sufficient bandwidth that loss and jitter are nominal. Where that cannot be accomplished, one must use queuing technology to deterministically apply bandwidth to accomplish the goal.

1.5 Desired Characteristics in an Internet Environment

The key elements of the MLPP service include the following:

Precedence Level Marking each call: Call initiators choose the appropriate precedence level for each call based on user perceived importance of the call. This level is not to be changed for the duration of the call. The call before, and the call after are independent with regard to this level choice.

Call Admission/Preemption Policy: There is likewise a clear policy regarding calls that may be in progress at the called instrument. During call admission (SIP/H.323), if they are of lower precedence, they must make way according to a prescribed procedure. All callers on the preempted call must be informed that the call has been preempted, and the call must make way for the higher precedence call.

Bandwidth Admission Policy: There is a clear bandwidth admission policy: sessions may be placed which assert any of several levels of precedence, and in the event that there is demand and authorization is granted, other sessions will be preempted to make way for a call of higher precedence.

Authentication and Authorization of calls placed: Unauthorized attempts to place a call at an elevated status are not permitted. In the telephone system, this is managed by controlling the policy applied to an instrument by its switch plus a code produced by the caller identifying himself or herself to the switch. In the Internet, such characteristics must be explicitly signaled.

Voice handling characteristics: A call made, in the telephone system, gets a circuit, and provides the means for the callers to conduct their business without significant impact as long as their call is not preempted. In a VoIP system, one would hope for essentially the same service.

Defined User Interface: If a call is preempted, the caller and the callee are notified via a defined signal, so that they know that their call has been preempted and that at this instant there is no alternative circuit available to them at that precedence level.

A VoIP implementation of the MLPP service must, by definition, provide those characteristics.

1.6 The use of bandwidth as a solution for QoS

There is a discussion in Internet circles concerning the relationship of bandwidth to QoS procedures, which needs to be put to bed before this procedure can be adequately analyzed. The issue is that it is possible and common in certain parts of the Internet to solve the problem with bandwidth. In LAN environments, for example, if there

is significant loss between any two switches or between a switch and a server, the simplest and cheapest solution is to buy the next faster interface - substitute 100 MBPS for 10 MBPS Ethernet, 1 Gigabit for 100 MBPS, or for that matter upgrade to a ten gigabit Ethernet. Similarly, in optical networking environments, the simplest and cheapest solution is often to increase the data rate of the optical path either by selecting a faster optical carrier or deploying an additional lambda. In places where the bandwidth can be overprovisioned to a point where loss or queuing delay are negligible, 10:1 overprovisioning is often the cheapest and surest solution, and by the way offers a growth path for future requirements. However, there are places in communication networks where bandwidth is not free and is therefore not effectively infinite. It is in these places, and only these places, where the question of resource management is relevant.

The places where bandwidth constriction takes place is typically where one pays a significant amount for bandwidth, such as in access paths, or where available technology limits the options. In military networks, Type 1 encryption often presents such a barrier, as do satellite links and various kinds of radio systems.

In short, the fact that we are discussing this class of policy control says that such constrictions in the network exist and must be dealt with. However much we might like to, in those places we are not solving the problem with bandwidth.

2. Solution Proposal

A typical voice or video network, including a backbone domain, is shown in Figure 1.

```
   ...............             ......................
  .                .          .                      .
 .  H  H  H  H      .        .   H  H  H  H           .
.  /----------/      .       .  /----------/           .
.     R     SIP      .       .    R      R              .
.      \             .       .   /        \              .
.       R  H  H  H   . .......  /          \             .
.      /----------/  ..      ../            R     SIP    .
 .              R  ..         /.           /----------/  .
   .....       ..\.    R-----R  .           H  H  H  H   .
         ......  .\   /       \  .                      .
                 . \ /         \  .                    .
                 .  R-----------R  ...................
                 .   \         /   .
                 .    \       /   .
                 .     R-----R   .
                  .             .
                    ............
          SIP   = SIP Proxy
          H     = SIP-enabled Host (Telephone, call gateway or PC)
          R     = Router
          /---/ = Ethernet or Ethernet Switch
```

Figure 1: Typical VoIP or Video/IP Network

Reviewing that figure, it becomes obvious that Voice/IP and Video/IP call flows are very different than call flows in the PSTN. In the PSTN, call control traverses a switch, which in turn controls data handling services like ATM switches or circuit multiplexers. While they may not be physically co-located, the control plane software and the data plane services are closely connected; the switch routes a call using bandwidth that it knows is available. In a voice/video-on-IP network, call control is completely divorced from the data plane: It is possible for a telephone instrument in the United States to have a Swedish telephone number if that is where its SIP proxy happens to be, but on any given call to use only data paths in the Asia/Pacific region, data paths provided by a different company, and often data paths provided by multiple companies/providers.

Call management therefore addresses a variety of questions, all of which must be answered:

o May I make this call from an administrative policy perspective?

- o What IP address correlates with this telephone number or SIP URI?
- o Is the other instrument "on hook"? If it is busy, under what circumstances may I interrupt?
- o Is there bandwidth available to support the call?
- o Does the call actually work?

2.1 Call admission/preemption procedure

Administrative Call Admission is the objective of SIP and H.323. It asks fundamental questions like "what IP address is the callee at?" and "Did you pay your bill?".

For specialized policy like call preemption, two capabilities are necessary from an administrative perspective: [I-D.ietf-sip-resource-priority] provides a way to communicate policy-related information regarding the precedence of the call; and [I-D.ietf-sipping-reason-header-for-preemption] provides a reason code when a call fails or is refused, indicating the cause of the event. If it is a failure, it may make sense to redial the call. If it is a policy-driven preemption, even if the call is redialed it may not be possible to place the call.

The Communications Resource Priority Header (or RP Header) serves the call set-up process with the precedence level chosen by the initiator of the call. The syntax is in the form:

```
Resource Priority : namespace.priority level
```

The "namespace" part of the syntax ensures the domain of significance to the originator of the call, and this travels end-to-end to the destination (called) device (phone). If the receiving phone does not support the namespace, it can easily ignore (what [I-D.ietf-sip-resource-priority] calls "loose mode") or errors (what [I-D.ietf-sip-resource-priority] calls "strict mode") the set-up request. This ability to denote the domain of origin allows SLAs to be in place to limit the ability of an unknown requestor to gain preferential treatment into an MLPP domain.

For the DSN infrastructure, this header would look like this:

```
Resource Priority : dsn.routine
```

for a routine precedence level call. The precedence level chosen in this header would be compared to the requestor's authorization

profile to user that precedence level. This would typically occur in the SIP first hop Proxy, which can challenge many aspects of the call set-up request including the requestor choice of precedence levels (verifying they aren't using a level they are not authorized to use.)

The DSN has 5 precedence levels of MLPP in descending order:

dsn.flash-override

dsn.flash

dsn.immediate

dsn.priority

dsn.routine

The US Defense Red Switched Network (DRSN), as another example that is to be IANA registered in [I-D.ietf-sip-resource-priority], has 6 levels of precedence. The DRSN simply adds one higher precedence level than flash-override:

drsn.flash-override-override

to be used by the President and a select few others. Note that the namespace changed for this level. The lower 5 levels within the DRSN would also have this as their namespace for all DRSN originated call set-up requests.

This informs both the use of DSCPs by the callee (who needs to use the same DSCP as the caller to obtain the same data path service) and to facilitate policy-based preemption of calls in progress when appropriate.

Once a call is established in an MLPP domain, the Reason Header for Preemption, described in [I-D.ietf-sipping-reason-header-for-preemption], ensures that all SIP nodes are synchronized to a preemption event occurring either at the endpoint or in a router that experiences congestion. In SIP, the normal indication for the end of a session is for one end system to send a BYE Method request as specified in [RFC3261]. This, too, is the proper means for signaling a termination of a call due to a preemption event, as it essentially performs a normal termination with additional information informing the peer of the reason for the abrupt end - it indicates that a preemption occurred. This will be used to inform all relevant SIP entities, and whether this was a endpoint generated preemption event, or that the preemption event occurred within a router along the communications path (described in

Section 2.3.1).

Figure X is a simple example of a SIP call set-up that includes the layer 7 precedence of a call between Alice and Bob. After Alice successfully sets up a call to Bob at the "Routine" precedence level, Carol calls Bob at a higher precedence level (Immediate). At the SIP layer (this has nothing to do with RSVP yet, that example involving SIP and RSVP signaling will be in the appendix), once Bob's user agent (phone) receives the INVITE message from Carol, his UA needs to make a choice between retaining the call to Alice and sending Carol a "busy" indication, or preempting the call to Alice in favor of accepting the call from Carol. That choice in MLPP networks is a comparison of Resource Priority headers. Alice, who controlled the precedence level of the call to Bob, sent the precedence level of her call to him at "Routine" (the lowest level within the network). Carol, who controls the priority of the call signal to Bob, sent her priority level to "Immediate" (higher than "Routine"). Bob's UA needs to (under MLPP policy) preempt the call from Alice (and provide her with a preemption indication in the call termination message). Bob needs to successfully answer the call set-up from Carol.

```
      UA Alice                     UA Bob                       UA Carol

         |    INVITE (RP: Routine)    |                            |
         |--------------------------->|                            |
         |           200 OK           |                            |
         |<---------------------------|                            |
         |            ACK             |                            |
         |--------------------------->|                            |
         |            RTP             |                            |
         |<===========================>|                            |
         |                            |                            |
         |                            |   INVITE (RP: Immediate)   |
         |                            |<---------------------------|
         |      ************************************************    |
         |      *Resource Priority value comparison by Bob's UA*   |
         |      ************************************************    |
         |                            |                            |
         | BYE (Reason:UA_preemption) |                            |
         |<---------------------------|                            |
         |                            |           200 OK           |
         |                            |--------------------------->|
         |       200 OK (BYE)         |                            |
         |--------------------------->|                            |
         |                            |            ACK             |
         |                            |<---------------------------|
         |                            |            RTP             |
         |                            |<===========================>|
         |                            |                            |
```

Figure 2: Priority Call Establishment and Termination at SIP Layer

Nothing in this example involved mechanisms other than SIP. It is also assumed each user agent recognized the Resource-Priority header's namespace value. Therefore, it is assumed that the domain allowed Alice, Bob and Carol to communicate. Authentication and Authorization are discussed later in this document.

2.2 Voice handling characteristics

The Quality of Service architecture used in the data path is that of [RFC2475]. Differentiated Services uses a flag in the IP header called the DSCP [RFC2474] to identify a data stream, and then applies a procedure called a Per Hop Behavior, or PHB, to it. This is largely as described in the [RFC2998].

In the data path, the Expedited Forwarding PHB [RFC3246][RFC3247] describes the fundamental needs of voice and video traffic. This PHB entails ensuring that sufficient bandwidth is dedicated to real-time

traffic to ensure minimal variation in delay and a minimal loss rate, as codecs are hampered by excessive loss [G711.1][G711.2][G711.3] . In parts of the network where bandwidth is heavily overprovisioned, there may be no remaining concern. In places in the network where bandwidth is more constrained, this may require the use of a priority queue. If a priority queue is used, the potential for abuse exists, meaning that it is also necessary to police traffic placed into the queue to detect and manage abuse. A fundamental question is "where does this policing need to take place?". The obvious places would be the first hop routers and any place where converging data streams might congest a link.

For policy reasons, DISA would like to mark traffic with various code points appropriate to the service precedence of the call. In normal service, if the traffic is all in the same queue and EF service requirements are met (applied capacity exceeds offered load, variation in delay is minimal, and loss is negligible), details of traffic marking should be irrelevant, as long as they get the packets into the right service class. The major issue, then is primarily one of appropriate policing of traffic, especially around route changes.

The real time voice/video application should be generating traffic at a rate appropriate to its content and codec, which is either a constant bit rate stream or a stream whose rate is variable within a specified range. The first hop router should be policing traffic originated by the application, as is performed in traditional virtual circuit networks like Frame Relay and ATM. Between these two, the application traffic should be guaranteed to be within acceptable limits. As such, given bandwidth-aware call admission control, there should be minimal actual loss. The cases where loss would occur include cases where routing has recently changed and CAC has not caught up, or cases where statistical thresholds are in use in CAC and the data streams happen to coincide at their peak rates.

If it is demonstrated that routing transients and variable rate beat frequencies present a sufficient problem, it is possible to provide a policing mechanism that isolates intentional loss among an ordered set of classes. While the ability to do so, by various algorithms, has been demonstrated, the technical requirement has not. If dropping random packets from all calls is not appropriate, concentrating random loss in a subset of the calls makes the problem for those calls worse; a superior approach would reject or preempt an entire call.

Parekh's second condition has been met: we must know what the network will do with the traffic. If the offered load exceeds the available bandwidth, the network will remark and drop the excess traffic. The key questions become "How does one limit offered load to a rate less

than or equal to available bandwidth?" and "how much traffic does one admit with each appropriate marking?"

2.3 Bandwidth admission procedure

Since the available voice and video codecs require a nominal loss rate to deliver acceptable performance, Parekh's first requirement is that offered load be within the available capacity. There are several possible approaches.

An approach that is commonly used in H.323 networks is to limit the number of calls simultaneously accepted by the gatekeeper. SIP networks do something similar when they place a SIP proxy near a single ingress/egress to the network. This is able to impose an upper bound on the total number of calls in the network or the total number of calls crossing the significant link. However, the gatekeeper has no knowledge of routing, so the engineering must be very conservative, and usually requires a single ingress/egress - a single point of failure. While this may serve as a short term work-around, it is not a general solution that is readily deployed. This limits the options in network design.

The [RFC1633] provides for signalled admission for the use of capacity. This is currently implemented using the Resource Reservation Protocol [RFC2205][RFC2209] (RSVP). The use of Capacity Admission with SIP is described in [RFC3312] ; at this writing, Capacity Admission is not integrated with H.323.

2.3.1 Recommended procedure: explicit call admission - RSVP Admission using Policy

RSVP is a resource reservation setup protocol providing the one-way (at a time) setup of resource reservations for multicast and unicast flows. Each reservation is set up in one direction (meaning one reservation from each end system; in a multicast environment, N senders set up N reservations). These reservations complete a communication path with a deterministic bandwidth allocation through each router along that path between end systems. These reservations setup a known quality of service for end-to-end communications and maintain a "soft-state" within a node. The meaning of the term "soft state" is that in the event of a network outage or change of routing, these reservations are cleared without manual intervention, but must be periodically refreshed. In RSVP, the refresh period is by default 30 seconds, but may be as long as appropriate.

RSVP is a locally-oriented process, not a globally- or domain-oriented one like a routing protocol or like H.323 Call Counting. Although it uses the local routing databases to determine

the routing path, it is only concerned with the quality of service for a particular or aggregate flow through a device. RSVP is not aware of anything other than the local goal of QoS and its RSVP-enabled adjacencies, operating below the network layer. The process by itself neither requires nor has any end-to-end network knowledge or state. Thus, RSVP can be enabled in a network without the need to have every node participate.

```
              HOST                            ROUTER
 ____________________________       ____________________________
|  _______                    |     |                            |
| |       |   _______         |     |            _______         |
| |Appli- |  |       |        |RSVP |           |       |        |
| | cation|  | RSVP <----------------------------> RSVP  <---------->
| |       <-->       |        |     | _______   |       |        |
| |       |  |process|  _____ |     ||Routing|  |process|  _____ |
| |_._____|  |       -->Polcy||     ||       <-->       -->Polcy||
|   |        |__.__._| |Cntrl||     ||process|  |__.__._| |Cntrl||
|   |data       |  |   |_____||     ||__.____|     |  |   |_____||
|===|===========|==|==========|     |===|==========|==|==========|
|   |   --------|  |    _____ |     |   |  --------|  |    _____ |
|   |  |        |  ---->Admis||     |   |  |       |  ---->Admis||
|  _V__V_    ___V____  |Cntrl||     |  _V__V_    __V_____ |Cntrl||
| |      |  |        | |_____||     | |      |  |        ||_____||
| |Class-|  | Packet |        |     | |Class-|  | Packet |       |
| | ifier|==>Schedulr|================> ifier|==>Schedulr|===========>
| |______|  |________|        |data | |______|  |________|       |data
|                             |     |                            |
|_____________________________|     |____________________________|
```

Figure 3: RSVP in Hosts and Routers

Figure 3 shows the internal process of RSVP in both hosts (end systems) and routers, as shown in [RFC2209].

RSVP uses the phrase "traffic control" to describe the mechanisms of how a data flow receives quality of service. There are 3 different mechanisms to traffic control (shown in Figure 2 in both hosts and routers). They are:

A packet classifier mechanism: which resolves the QoS class for each packet; this can determine the route as well.

An admission control mechanism: this consists of two decision modules: the admission control module and the policy control module. Determining whether there is satisfactory resources for the requested QoS is the function of admission control. Determining if the user has the authorization to request such

resources is the function of policy control. If the parameters carried within this flow fail either of these two modules, RSVP errors the request.

A packet scheduler mechanism: at each outbound interface, the scheduler attains the guaranteed QoS for that flow

2.3.2 RSVP Scaling Issues

As originally written, RSVP had scaling limitations due to its data plane behavior. This has, in time, largely been corrected. In edge networks, RSVP is used to signal for individual microflows, admitting the bandwidth. However, Differentiated Services is used for the data plane behavior. Admission and policing may be performed anywhere, but need only be performed in the first hop router (which, if the end system sending the traffic is a DTE, constitutes a DCE for the remaining network) and in routers that have interfaces threatened by congestion. In Figure 1, these would normally be the links that cross network boundaries, and may also include any type 1 encrypted interface, as these are generally limited in bandwidth by the encryption.

2.3.3 RSVP Operation in backbones and VPNs

In backbone networks, networks that are normally awash in bandwidth, RSVP and its affected data flows may be carried in a variety of ways. If the backbone is a maze of tunnels between its edges - true of MPLS networks and of networks that carry traffic from an encryptor to a decryptor, and also of VPNs - applicable technologies include [RFC2207], [RFC2746], and [RFC2983]. An IP tunnel is simplistically a IP packet enveloped inside another IP packet as a payload. When IPv6 is transported over an IPv4 network, encapsulating the entire v6 packet inside a v4 packet is an effective means to accomplish this task. In this type of tunnel, the IPv6 packet is not read by any of the routers while inside the IPv4 envelope. If the inner packet is RSVP enabled, there must be a active configuration to ensure that all relevant backbone nodes read the RSVP fields; [RFC2746] describes this.

This is similar to how IPsec tunnels work. Encapsulating an RSVP packet inside an encrypted packet for security purposes without copying or conveying the RSVP indicators in the outside IP packet header would make RSVP inoperable while in this form of a tunnel. [RFC2207] describes how to modify an IPsec packet header to allow for RSVP awareness by nodes that need to provide QoS for the flow or flows inside a tunnel.

Other networks may simply choose to aggregate the reservations across themselves as described in [RFC3175]. The problem with an individual reservation architecture is that each flow requires a non-trivial amount of message exchange, computation, and memory resources in each router between each endpoint. Aggregation of flows reduces the number of completely individual reservations into groups of individual flows that can act as one for part or all of the journey between end systems. Aggregates are not intended to be from the first router to the last router within a flow, but to cover common paths of a large number of individual flows.

Examples of aggregated data flows include streams of IP data that traverse common ingress and egress points in a network, and also include tunnels of various kinds. MPLS LSPs, IPSEC Security Associations between VPN edge routers, similar tunnels between HAIPE encryptors and decryptors, IP/IP tunnels, and GRE tunnels all fall into this general category. The distinguishing factor is that the system injecting an aggregate into the aggregated network sums the PATH and RESV statistical information on the un-aggregated side and produces a reservation for the tunnel on the aggregated side. If the bandwidth for the tunnel cannot be expanded, RSVP leaves the existing reservation in place and returns an error to the aggregator, which can then apply a policy such as MLPP to determine which session to refuse. In the data plane, the DSCP for the traffic must be copied from the inner to the outer header, to preserve the PHB's effect.

One concern with this approach is that this leaks information into the aggregated zone concerning the number of active calls or the bandwidth they consume. In fact, it does not, as the data itself is identifiable by aggregator address, deaggregator address, and DSCP. As such, even if it is not advertised, such information is measurable.

2.3.4 Interaction with the Differentiated Services Architecture

In the PATH message, the DCLASS object described in [RFC2996] is used to carry the determined DSCP for the precedence level of that call in the stream. This is reflected back in the RESV message. The DSCP will be determined from the authorized SIP message exchange between end systems by using the R-P header. The DCLASS object permits both bandwidth admission within a class and the building up of the various rates or token buckets.

2.3.5 Admission policy

RSVP's basic admission policy, as defined, is to grant any user bandwidth if there is bandwidth available within the current configuration. In other words, if a new request arrives and the

difference between the configured upper bound and the currently reserved bandwidth is sufficiently large, RSVP grants use of that bandwidth. This basic policy may be augmented in various ways, such as using a local or remote policy engine to apply AAA procedures and further qualify the reservation.

2.3.5.1 Admission for variable rate codecs

For certain applications, such as broadcast video using MPEG-1 or voice without activity detection and using a constant bit rate codec such as G.711, this basic policy is adequate apart from AAA. For variable rate codecs, such as MPEG-4 or a voice codec with Voice Activity Detection, however, this may be deemed too conservative. In such cases, two basic types of statistical policy have been studied and reported on in the literature: simple overprovisioning, and approximation to ambient load.

Simple overprovisioning sets the bandwidth admission limit higher than the desired load, on the assumption that a session that admits a certain bandwidth will in fact use a fraction of the bandwidth. For example, if MPEG-4 data streams are known to use data rates between 80 and 800 KBPS and there is no obvious reason that sessions would synchronize (such as having commercial breaks on 15 minute boundaries), one could imagine estimating that the average session consumes 400 KBPS and treating an admission of 800 KBPS as actually consuming half the amount.

One can also approximate to average load, which is perhaps a more reliable procedure. In this case, one maintains a variable which measures actual traffic through the admitted data's queue, approximating it using an exponentially weighted moving average. When a new reservation request arrives, if the requested rate is less than the difference between the configured upper bound and the current value of the moving average, the reservation is accepted and the moving average is immediately increased by the amount of the reservation to ensure that the bandwidth is not promised out to several users simultaneously. In time, the moving average will decay from this guard position to an estimate of true load, which may offer a chance to another session to be reserved that would otherwise have been refused.

Statistical reservation schemes such as these are overwhelmingly dependent on the correctness of their configuration and its appropriateness for the codecs in use. But they offer the opportunity to take advantage of statistical multiplexing gains that might otherwise be missed.

2.3.5.2 Interaction with complex admission policies, AAA, and preemption of bandwidth

Policy is carried and applied as described in [RFC2753]. Figure 4 below is the basic conceptual model for policy decisions and enforcement in an Int-Serv model. This model was created to provide ability to monitor and control reservation flows based on user identify, specific traffic and security requirements and conditions which might change for various reasons, including as a reaction to a disaster or emergency event involving the network or its users.

```
  Network Node       Policy server
 _______________
|   ______     |
|  |      |    |        _____
|  | PEP  |    |       |     |------------->
|  |______|<---|------>| PDP |May use LDAP,SNMP,COPS... for accessing
|     ^        |       |     | policy database, authentication, etc.
|     |        |       |_____|------------->
|   __v___     |
|  |      |    |       PDP = Policy Decision Point
|  | LPDP |    |       PEP = Policy Enforcement Point
|  |______|    |      LPDP = Local Policy Decision Point
|_______________|
```

Figure 4: Conceptual Model for Policy Control of Routers

The Network Node represents a router in the network. The Policy Server represents the point of admission and policy control by the network operator. Policy Enforcement Point (PEP)(the router) is where the policy action is carried out. Policy decisions can be either locally present in the form of a Local Policy Decision Point (LPDP), or in a separate server on the network called the Policy Decision Point. The easier the instruction set of rules, the more likely this set can reside in the LDPD for speed of access reasons. The more complex the rule set, the more likely this is active on a remote server. The PDP will use other protocols (LDAP, SNMP, etc) to request information (e.g. user authentication and authorization for precedence level usage) to be used in creating the rule sets of network components. This remote PDP should also be considered where non-reactive policies are distributed out to the LPDPs.

Taking the above model as a framework, [RFC2750] extends RSVP's concept of a simple reservation to include policy controls, including the concepts of Preemption [RFC3181] and Identity [RFC3182], specifically speaking to the usage of policies which preempt calls under the control of either a local or remote policy manager. The policy manager assigns a precedence level to the admitted data flow.

If it admits a data flow that exceeds the available capacity of a system, the expectation is that the RSVP affected RSVP process will tear down a session among the lowest precedence sessions it has admitted. The RESV Error resulting from that will go to the receiver of the data flow, and be reported to the application (SIP or H.323). That application is responsible to disconnect its call, with a reason code of "bandwidth preemption".

2.4 Authentication and authorization of calls placed

It will be necessary, of course, to ensure that any policy is applied to an authenticated user; it is the capabilities assigned to an authenticated user that may be considered to have been authorized for use in the network. For bandwidth admission, this will require the utilization of [RFC2747][RFC3097]. In SIP and H.323, AAA procedures will also be needed.

2.5 Defined User Interface

The user interface - the chimes and tones heard by the user - should ideally remain the same as in the MLPP PSTN for those indications that are still applicable to an IP network. There should be some new effort generated to update the list of announcements sent to the user which don't necessarily apply. For example, in an end-to-end IP call, there is no known benefit to informing the user which Ethernet switch or router caused the call to fail - as is the equivalent case if a TDM Switch were the cause. All indications to the user, of course, depend on positive signals, not unreliable measures based on changing measurements.

3. IANA Considerations

This document makes no request of IANA.

Note to RFC Editor: this section may be removed on publication as an RFC.

4. Security Considerations

This document outlines a networking capability composed entirely of existing specifications. It has significant security issues, in the sense that a failure of the various authentication or authorization procedures can cause a fundamental breakdown in communications. However, the issues are internal to the various component protocols, and are covered by their various security procedures.

5. Acknowledgements

This document was developed with the knowledge and input of many people, far too numerous to be mentioned by name. Key contributors of thoughts include, however, Francois Le Faucheur, Haluk Keskiner, Rohan Mahy, Scott Bradner, Scott Morrison, and Subha Dhesikan. Pete Babendreier's review was especially useful.

6. References

6.1 Normative References

[ANSI.MLPP.Spec]
American National Standards Institute, "Telecommunications - Integrated Services Digital Network (ISDN) - Multi-Level Precedence and Preemption (MLPP) Service Capability", ANSI T1.619-1992 (R1999), 1992.

[ANSI.MLPP.Supplement]
American National Standards Institute, "MLPP Service Domain Cause Value Changes", ANSI ANSI T1.619a-1994 (R1999), 1990.

[I-D.ietf-sip-resource-priority]
Schulzrinne, H. and J. Polk, "Communications Resource Priority for the Session Initiation Protocol (SIP)", draft-ietf-sip-resource-priority-04 (work in progress), September 2004.

[I-D.ietf-sipping-reason-header-for-preemption]
Polk, J., "Extending the Session Initiation Protocol Reason Header for Preemption Events", draft-ietf-sipping-reason-header-for-preemption-02 (work in progress), August 2004.

[I-D.pierce-ieprep-assured-service-arch]
Pierce, M. and D. Choi, "Architecture for Assured Service Capabilities in Voice over IP", draft-pierce-ieprep-assured-service-arch-02 (work in progress), January 2004.

[I-D.pierce-ieprep-assured-service-req]
Pierce, M. and D. Choi, "Requirements for Assured Service Capabilities in Voice over IP", draft-pierce-ieprep-assured-service-req-02 (work in progress), January 2004.

[ITU.MLPP.1990]
International Telecommunications Union, "Multilevel Precedence and Preemption Service (MLPP)", ITU-T Recommendation I.255.3, 1990.

[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.

[RFC2207] Berger, L. and T. O'Malley, "RSVP Extensions for IPSEC Data Flows", RFC 2207, September 1997.

[RFC2209] Braden, B. and L. Zhang, "Resource ReSerVation Protocol (RSVP) -- Version 1 Message Processing Rules", RFC 2209, September 1997.

[RFC2327] Handley, M. and V. Jacobson, "SDP: Session Description Protocol", RFC 2327, April 1998.

[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.

[RFC2746] Terzis, A., Krawczyk, J., Wroclawski, J. and L. Zhang, "RSVP Operation Over IP Tunnels", RFC 2746, January 2000.

[RFC2747] Baker, F., Lindell, B. and M. Talwar, "RSVP Cryptographic Authentication", RFC 2747, January 2000.

[RFC2750] Herzog, S., "RSVP Extensions for Policy Control", RFC 2750, January 2000.

[RFC2753] Yavatkar, R., Pendarakis, D. and R. Guerin, "A Framework for Policy-based Admission Control", RFC 2753, January 2000.

[RFC2983] Black, D., "Differentiated Services and Tunnels", RFC 2983, October 2000.

[RFC2996] Bernet, Y., "Format of the RSVP DCLASS Object", RFC 2996, November 2000.

[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.

[RFC3097] Braden, R. and L. Zhang, "RSVP Cryptographic

Authentication -- Updated Message Type Value", RFC 3097, April 2001.

[RFC3175] Baker, F., Iturralde, C., Le Faucheur, F. and B. Davie, "Aggregation of RSVP for IPv4 and IPv6 Reservations", RFC 3175, September 2001.

[RFC3181] Herzog, S., "Signaled Preemption Priority Policy Element", RFC 3181, October 2001.

[RFC3182] Yadav, S., Yavatkar, R., Pabbati, R., Ford, P., Moore, T., Herzog, S. and R. Hess, "Identity Representation for RSVP", RFC 3182, October 2001.

[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.

[RFC3261] Rosenberg, J., Schulzrinne, H., Camarillo, G., Johnston, A., Peterson, J., Sparks, R., Handley, M. and E. Schooler, "SIP: Session Initiation Protocol", RFC 3261, June 2002.

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

[RFC3326] Schulzrinne, H., Oran, D. and G. Camarillo, "The Reason Header Field for the Session Initiation Protocol (SIP)", RFC 3326, December 2002.

6.2 Informative References

[G711.1] Viola Networks, "Netally VoIP Evaluator", January 2003, <http://www.sygnusdata.co.uk/white_papers/viola/netally_voip_sample_report_preliminary.pdf>.

[G711.2] ETSI Tiphon, "ETSI Tiphon Temporary Document 64", July 1999, <http://docbox.etsi.org/tiphon/tiphon/archives/1999/05-9907-Amsterdam/14TD113.pdf>.

[G711.3] Nortel Networks, "Packet Loss and Packet Loss Concealment", 2000, <http://www.nortelnetworks.com/products/01/succession/es/collateral/tb_pktloss.pdf>.

[G711.4] Clark, A., "Modeling the Effects of Burt Packet Loss and recency on Subjective Voice Quality", 2000, <http://www.telchemy.com/references/tech_papers/iptel2001.pdf>.

[G711.5] Cisco Systems, "Understanding Codecs: Complexity, Hardware Support, MOS, and Negotiation", 2003, <http://www.cisco.com/en/US/tech/tk652/tk701/technologies_tech_note09186a00800b6710.shtml#mos>.

[I-D.ietf-avt-ilbc-codec] Andersen, S., "Internet Low Bit Rate Codec", draft-ietf-avt-ilbc-codec-05 (work in progress), June 2004.

[ILBC] Chen, M. and M. Murthi, "On The Performance Of ILBC Over Networks With Bursty Packet Loss", July 2003.

[Parekh1] Parekh, A. and R. Gallager, "A Generalized Processor Sharing Approach to Flow Control in Integrated Services Networks: The Multiple Node Case", INFOCOM 1993: 521-530, 1993.

[Parekh2] Parekh, A. and R. Gallager, "A Generalized Processor Sharing Approach to Flow Control in Integrated Services Networks: The Single Node Case", INFOCOM 1992: 915-924, 1992.

Authors' Addresses

Fred Baker
Cisco Systems
1121 Via Del Rey
Santa Barbara, California 93117
USA

Phone: +1-408-526-4257
Fax: +1-413-473-2403
EMail: fred@cisco.com

James Polk
Cisco Systems
2200 East President George Bush Turnpike
Richardson, Texas 75082
USA

Phone: +1-469-255-5208
EMail: jmpolk@cisco.com

Appendix A. 2-Call Preemption Example using RSVP

This appendix will present a more complete view of the interaction between SIP, SDP and RSVP. The bulk of the material is referenced from [RFC2327], [RFC3312], [I-D.ietf-sipping-reason-header-for-preemption], [I-D.ietf-sip-resource-priority]. There will be some discussion on basic RSVP operations regarding reservation paths, this will be mostly from [RFC2205].

SIP signaling occurs at layer 7, riding on a UDP/IP or TCP/IP (including TLS/TCP/IP) transport that is bound by routing protocols such as BGP and OSPF to determine the route the packets traverse through a network between source and destination devices. RSVP is riding on top of IP as well, which means RSVP is at the mercy of the IP routing protocols to determine a path through the network between endpoints. RSVP is not a routing protocol. In this appendix there will be a escalation of building blocks getting to how the many layers are involved in SIP with QoS Preconditions requiring successful RSVP signaling between endpoints prior to SIP successfully acknowledging the set-up of the session (for voice or video or both). Then we will present what occurs when a network overload occurs (congestion), causing a SIP session to be preempted.

There are 3 diagrams in this appendix to show multiple views of the same example of connectivity for discussion throughout this appendix. The first diagram (Figure 5) is of many routers between many endpoints (SIP user agents, or UAs). There are 4 UAs of interest, those are for users Alice, Bob, Carol and Dave. When a user (the human) of a UA gets involved and must do something to a UA to progress a SIP process, this will be explicitly mentioned to avoid confusion; otherwise, when Alice is referred to - this means Alice's UA (her phone) in the text here.

RSVP reserves bandwidth in one direction only (the direction of the RESV message), as has been discussed, IP forwarding of packets are dictated by the routing protocol for that portion of the infrastructure from the point of view of where the packet is to go next.

The RESV message traverses the routers in the reverse path taken by the PATH message. The PATH message establishes a record of the route taken through a network portion to the destination endpoint, but it does not reserve resources (bandwidth). The RESV message back to the original requestor of the RSVP flow requests for the bandwidth resources. This means the endpoint that initiates the RESV message controls the parameters of the reservation. This document specifies in the body text that the SIP initiator (the UAC) establishes the

parameters of the session in an INVITE message, and that the INVITE recipient (the UAS) must follow the parameters established in that INVITE message. One exception to this is which codec to use if the UAC offered more than one to the UAS. This exception will be shown when the INVITE message is discussed in detail later in the appendix. If there was only one codec in the SDP of the INVITE message, the parameters of the reservation will follow what the UAC requested (specifically to include the Resource-Priority header namespace and priority value).

Here is the first figure with the 4 UAs and a meshed routed infrastructure between each. For simplicity of this explanation, this appendix will only discuss the reservations from Alice to Bob (one direction) and from Carol to Dave (one direction). An interactive voice service will require two one-way reservations that end in each UA. This gives the appearance of a two-way reservation, when indeed it is not.

```
        Alice -----R1----R2----R3----R4------ Bob
                   | \  /  \  /  \  / |
                   |  \/    \/    \/  |
                   |  /\    /\    /\  |
                   | /  \  /  \  /  \ |
        Carol -----R5----R6----R7----R8------ Dave
```

Figure 5: Complex Routing and Reservation Topology

The PATH message from Alice to Bob (establishing the route for the RESV message) will be through routers:

Alice -> R1 -> R2 -> R3 -> R4 -> Bob

The RESV message (and therefore the reservation of resources) from Bob to Alice will be through routers:

Bob -> R4 -> R3 -> R2 -> R1 -> Alice

The PATH message from Carol to Dave (establishing the route for the RESV message) will be through routers:

The reservation from Carol to Dave be through routers:

Carol -> R6 -> R2 -> R3 -> R7 -> R11 -> Dave

The RESV message (and therefore the reservation of resources) from Dave to Carol will be through routers:

Dave -> R11 -> R7 -> R3 -> R2 -> R6 -> Carol

The reservations from Alice to Bob traverse a common router link: between R3 and R2 and thus a common interface at R2. Here is where there will be congestion in this example, on the link between R2 and R3. Since the flow of data (in this case voice media packets) travels the direction of the PATH message, and RSVP establishes reservation of resources at the egress interface of a router, the interface in Figure 6 shows Int7 to be what will first know about a congestion condition.

```
          Alice                                Bob
             \                                 /
              \                               /
               +--------+          +--------+
               |        |          |        |
               |   R2   |          |   R3   |
               |       Int7-------Int5      |
               |        |          |        |
               +--------+          +--------+
              /                               \
             /                                 \
          Carol                                Dave
```

Figure 6: Reduced Reservation Topology

From Figure 6, the messaging between the UAs and the RSVP messages between the relevant routers can be shown to understand the binding that was established in [RFC3312] "SIP Preconditions for QoS".

We will assume all devices have powered up, and received whatever registration or remote policy downloads were necessary for proper operation. The routing protocol of choice has performed its routing table update throughout this part of the network. Now we are left to focus only on end-to-end communications and how that affects the infrastructure between endpoints.

The next diagram (Figure 7) (nearly identical to Figure 1 from [RFC3312])shows the minimum SIP messaging (at layer 7) between Alice and Bob for a good quality voice call. The SIP messages are numbered to identify special qualities are each. During the SIP signaling, RSVP will be initiated. That messaging will also be discussed below.

```
      UA Alice                                        UA Bob
          |                                              |
          |                                              |
          |-------------(1) INVITE SDP1---------------->|
          |                                              |   Note 1
          |<------(2) 183 Session Progress SDP2--------|     |
       ***|**********************************************|***<-+
       *  |----------------(3) PRACK------------------>|  *
       *  |                                              |  * When
       *  |<-----------(4) 200 OK (PRACK)--------------|  * RSVP
       *  |                                              |  * is
       *  |                                              |  * signaled
    +->***|**********************************************|***
    |     |-------------(5) UPDATE SDP3---------------->|
   Note 2 |                                              |
          |<--------(6) 200 OK (UPDATE) SDP4-----------|
          |                                              |
          |<-------------(7) 180 Ringing---------------|
          |                                              |
          |-----------------(8) PRACK----------------->|
          |                                              |
          |<------------(9) 200 OK (PRACK)-------------|
          |                                              |
          |                                              |
          |<-----------(10) 200 OK (INVITE)------------|
          |                                              |
          |------------------(11) ACK----------------->|
          |                                              |
          |         RTP (within the reservation)         |
          |<==========================================>|
          |                                              |
```

Figure 7: SIP Reservation Establishment Using Preconditions

The session initiation starts with Alice wanting to communicate with Bob. Alice decides on an MLPP precedence level for their call (the default is the "routine" level, which is for normal everyday calls, but a priority level has to be chosen for each call). Alice puts into her UA Bob's address and precedence level and (effectively) hits the send button. This is reflected in SIP with an INVITE Method Request message [M1]. Below is what SIP folks call a well-formed SIP message (meaning it has all the headers that are mandatory to function properly). We will pick on the USMC for the addressing of this message exchange.

```
[M1 - INVITE from Alice to Bob, RP=Routine, QOS=e2e and mandatory]

INVITE sip:bob@usmc.example.mil SIP/2.0
Via: SIP/2.0/TCP pc33.usmc.example.mil:5060
  ;branch=z9hG4bK74bf9
Max-Forwards: 70
From: Alice <sip:alice@usmc.example.mil>;tag=9fxced76sl
To: Bob <sip:bob@usmc.example.mil>
Call-ID: 3848276298220188511@pc33.usmc.example.mil
CSeq: 31862 INVITE
Requires: 100rel
Resource-Priority: dsn.routine
Contact: <sip:alice@usmc.example.mil>
Content-Type: application/sdp
Content-Length: 191

v=0
o=alice 2890844526 2890844526 IN IP4 usmc.example.mil
c=IN IP4 10.1.3.33
t=0 0
m=audio 49172 RTP/AVP 0 4 8
a=rtpmap:0 PCMU/8000
a=curr:qos e2e none
a=des:qos mandatory e2e sendrecv
```

From the INVITE above, Alice is inviting Bob to a session. The upper half of the lines (before the empty line in the middle) are SIP headers and header values, the lower half of the lines above are Session Description Protocol (SDP) lines. SIP headers (after the first line) are not to be in any particular order, with one exception: the Via header. It is a SIP hop (through a SIP Proxy) route path that has a new Via header line added by each SIP proxy this message traverses. This is similar in function to an RSVP PATH message (building a reverse path back to the originator of the message). At any point in the message's path, a SIP element knows the path to the originator of the message. There will be not SIP Proxies in this example, because for Preconditions, Proxies only make more messages that look identical (with the exception of the Via and Max-Forwards headers), and that is not worth the space here to replicate what has been done in SIP RFCs already.

SIP headers that are used for Preconditions are the:

Requires header - which mandates a reliable provisional response message to the conditions requesting in this INVITE (knowing they are special).

This will result in the 183 "Session Progress" message from Bob's UA

as a reliable confirmation that preconditions are required for this call.

- Resource-Priority header - which denotes the domain namespace and precedence level of the call on an end-to-end basis.

And that's it for SIP. Preconditions is requested, required and signaled for in the SDP portion of the message. SDP is carried in what's called a SIP message body (much like the text in an email message is carried). SDP has special properties [see [RFC2327] for more on SDP, or the MMUSIC WG for ongoing efforts regarding SDP]. SDP lines are in a specific order for parsing reasons by endsystems. Dialog (Call) generating SDP message bodies all must have an "m" line (or media description line). Following the "m" line is zero or more "a" lines (or Attribute lines). The m-line in Alice's INVITE calls for a voice session (this is where video is identified also) using one of 3 different codecs that Alice supports (0 = G.711, 4 = G.723 and 8 = G.729) that Bob gets to choose from for this session. Bob can choose any of the 3. The first a=rtpmap line is specific to the type of codec these 3 are (PCMU). The next two a-lines are the only identifiers that RSVP is to be used for this call. The second a-line:

a=curr:qos e2e none

identifies the "current" status of qos at Alice's UA. Note: everything in SDP is with respect to the sender of the SDP message body (Alice will never tell Bob how his SDP is, she will only tell Bob about her SDP).

"e2e" means RSVP is required from Alice's UA to Bob's UA; meaning an RSVP failure in either direction will fail the call attempt.

"none" means there is no reservation at Alice's UA (to Bob) at this time.

The final a-line (a=des):

a=des:qos mandatory e2e sendrecv

identifies the "desired" level of qos

"mandatory" means this request for qos MUST be successful or the call fails.

"e2e" means RSVP is required from Alice's UA to Bob's UA

"sendrecv" means the reservation is in both directions.

As discussed, RSVP does not reserve bandwidth in both directions, and that it is up to the endpoints to have 2 one-way reservations if that particular application (here voice) requires it. Voice between Alice and Bob requires 2 one-way reservations. The UAs will be the focal points for both reservations in both directions.

Message 2 is the 183 "Session Progress" message sent by Bob to Alice that indicates to Alice that Bob understands that preconditions are required for this call.

```
[M2 - 183 "Session Progress"]

SIP/2.0 183 Session Progress
Via: SIP/2.0/TCP swc50.atlanta.com:5060
  ;branch=z9hG4bK74bf9 ;received=10.1.3.33
From: Alice <sip:alice@atlanta.com>;tag=9fxced76sl
To: Bob <sip:bob@biloxi.com>;tag=8321234356
Call-ID: 3848276298220188511@pc33atlanta.com
CSeq: 31862 INVITE
RSeq: 813520
Contact: <sip:bob@biloxi.com>
Content-Type: application/sdp
Content-Length: 210

v=0
o=bob 2890844527 2890844527 IN IP4 biloxi.com
c=IN IP4 172.16.1.36
t=0 0
m=audio 3456 RTP/AVP 0
a=rtpmap:0 PCMU/8000
a=curr:qos e2e none
a=des:qos mandatory e2e sendrecv
a=conf:qos e2e recv
```

The only interesting header in the SIP portion of this message is the RSeq header, which is the "Reliable Sequence" header. The value is incremented for every Reliable message that's sent in this call set-up (to make sure none are lost, or to ignore duplicates).

Bob's SDP indicates several a-line statuses and picks a codec for the call. The codec picked is in the m=audio line (the "0" at the end of this line means G.711 will be the codec).

The a=curr line gives Alice Bob's status with regard to RSVP (currently "none").

The a=des line also states the desire for mandatory qos e2e in both

directions.

The a=conf line is new. This line means Bob wants confirmation that Alice has 2 one-way reservations before Bob's UA proceeds with the SIP session set-up.

This is where "Note-1" applies in Figure 7. At the point that Bob's UA transmits this 183 message, Bob's UA (the one that picked the codec, so it knows the amount of bandwidth to reserve) transmits an RSVP PATH message to Alice's UA. This PATH message will take the route previously discussed in Figure 5:

```
Bob -> R4 -> R3 -> R2 -> R1 -> Alice
```

This is the path of the PATH message, and the reverse will be the path of the reservation set up RESV message, or:

```
Alice -> R1 -> R2 -> R3 -> R4 -> Bob
```

Immediately after Alice transmits the RESV message towards Bob, Alice sends her own PATH message to initiate the other one-way reservation. Bob, receiving that PATH message, will reply with a RESV.

All this is independent of SIP. But during this time of reservation establishment, a Provisional Acknowledgement (PRACK) [M3] is sent from Alice to Bob to confirm the request for confirmation of 2 one-way reservations at Alice's UA. This message is acknowledged with a normal 200 OK message [M4]. This is shown in Figure 7.

As soon as the RSVP is successfully completed at Alice's UA (knowing it was the last in the two way cycle or reservation establishment), at the SIP layer an UPDATE message [M5] is sent to Bob's UA to inform his UA that current status of RSVP (or qos) is "e2e" and "sendrecv".

[M5 - UPDATE to Bob that Alice has qos e2e and sendrecv]

```
UPDATE sip:bob@biloxi.com SIP/2.0
From: Alice <sip:alice@atlanta.com>;tag=9fxced76sl
To: Bob <sip:bob@biloxi.com>
Contact: <sip:alice@atlanta.com>
CSeq: 10197 UPDATE
Content-Type: application/sdp
Content-Length: 191

v=0
o=alice 2890844528 2890844528 IN IP4 atlanta.com
c=IN IP4 10.1.3.33
t=0 0
m=audio 49172 RTP/AVP 0
a=rtpmap:0 PCMU/8000
a=curr:qos e2e send
a=des:qos mandatory e2e sendrecv
```

This is shown by the matching table that can be build from the a=curr line and a=des line. If the two lines match, then no further signaling need take place with regard to "qos". [M6] is the 200 OK acknowledgement of this synchronization between the two UAs.

[M6 - 200 OK to the UPDATE from Bob indicating synchronization]

```
SIP/2.0 200 OK sip:bob@biloxi.com
From: Alice <sip:alice@atlanta.com>;tag=9fxced76sl
To: Bob <sip:bob@biloxi.com>
Contact: <sip:alice@atlanta.com>
CSeq: 10197 UPDATE
Content-Type: application/sdp
Content-Length: 195

v=0
o=alice 2890844529 2890844529 IN IP4 atlanta.com
c=IN IP4 10.1.3.33
t=0 0
m=audio 49172 RTP/AVP 0
a=rtpmap:0 PCMU/8000
a=curr:qos e2e sendrecv
a=des:qos mandatory e2e sendrecv
```

At this point, the reservation is operational and both UA's know it, and Bob's UA now rings ([M7] is the SIP indication to Alice this is taking place) telling Bob the user that Alice is calling her. Nothing up until now has involved Bob the user. Bob picks up the phone (generating [M10], from which Alice's UA responds with the

final ACK) and RTP is now operating within the reservations between the two UAs.

Now we get to Carol calling Dave. Figure 6 shows a common router interface for the reservation between Alice to Bob, and one that will also be the route for one of the reservations between Carol to Dave. This interface will experience congestion in our example here.

Carol is now calling Dave at a Resource-Priority level of "Immediate" - which is higher in priority than Alice to Bob's "routine". In this continuing example, Router 2's Interface-7 is congested and cannot accept any more RSVP traffic. Perhaps the offered load is at interface capacity. Perhaps Interface-7 is configured with a fixed amount of bandwidth is can allocate for RSVP traffic and has reached its maximum with one of the reservations going away through normal termination or forced termination (preemption).

Interface-7 is not so full of offered load that it cannot transmit signaling packets, such as Carol's SIP messaging to set up a call to Dave. This should be by design - that not all RSVP traffic can starve an interface from signaling packets. Carol sends her own INVITE with the following characteristics important here:

[M1 - INVITE from Carol to Dave, RP=Immediate, QOS=e2e and mandatory]

This packet does *not* affect the reservations between Alice and Bob (SIP and RSVP are at different layers, and all routers area passing signaling packets without problems). Dave sends his M2:

[M2 - 183 "Session Progress"]

with the SDP chart of:

```
a=curr:qos e2e none

a=des:qos mandatory e2e sendrecv

a=conf:qos e2e recv
```

indicating he understands RSVP reservations are required e2e for this call to be considered successful. Dave sends his PATH message. The PATH message does *not* affect Alice's reservation, it merely establishes a path for the RESV reservation set-up message to take.

To keep this example simple, the PATH message from Dave to Carol took this route (which we make different from the route in the reverse direction):

Dave -> R8 -> R7 -> R6 -> R5 -> Carol

causing the reservation to be this route:

Carol -> R5 -> R6 -> R7 -> R8 -> Dave

The reservation above in this direction (Dave to will not traverse any of the same routers as the Alice to Bob reservations. When Carol transmits her RESV message towards Dave, she immediately transmits her PATH message to set up the complementary reservation.

The PATH message from Carol to Dave be through routers:

Carol -> R5 -> R2 -> R3 -> R8 -> Dave

Thus, the RESV message will be through routers:

Dave -> R8 -> R3 -> R2 -> R5 -> Carol

This RESV message will traverse the same routers R3 and R2 as the Alice to Bob reservation. This RESV message, when received at Int-7 of R2, will create a congestion situation such that R2 will need to make a decision on whether:

- o to keep the Alice to Bob reservation and error the new RESV from Dave, or

- o to error the reservation from Alice to Bob in order to make room for the Carol to Dave reservation

Alice's reservation was set up in SIP at the "routine" precedence level. This will equate to a comparable RSVP priority number (RSVP has 65,535 priority values, or 2*32 bits per [RFC3181]). Dave's RESV equates to a precedence value of "immediate", which is a higher priority. Thus, R2 will preempt the reservation from Alice to Bob, and allow the reservation request from Dave to Carol. The proper RSVP error is the ResvErr that indicates preemption. This message travels downstream towards the originator or the RESV message (Bob). This clears the reservation in all routers downstream of R2 (meaning R3 and R4). Once Bob receives the ResvErr message indicating preemption has occur on this reservation, Bob's UA transmits a SIP preemption indication back towards Alice's UA. This accomplishes two things: first it informs all SIP Servers that were in the session set-up path that wanted to remain "dialog stateful" per [RFC3261]], and informs Alice's UA that this was a purposeful termination, and to play a preemption tone. The proper indication in SIP of this termination due to preemption is a BYE Method message that includes a Reason Header indicating why this occurred (in this case,

"RSVP_Preemption". Here is that message from Bob to Alice that terminates the call in SIP.

```
BYE sip:alice@usmc.example.mil SIP/2.0
Via: SIP/2.0/TCP swp34.usmc.example.mil
  ;branch=z9hG4bK776asegma
To: Alice <sip:alice@usmc.example.mil>
From: Bob < sip:bob@usmc.example.mil>;tag=192820774
Reason: cause=2 ;text=RSVP preemption
Call-ID: a84b4c76e66710@swp34.usmc.example.mil
CSeq: 6187 BYE
Contact: <sip:bob@usmc.example.mil>
```

When Alice's UA receives this message, her UA terminates the call, sends a 200 OK to Bob to confirm reception of the BYE message, and plays a preemption tone to Alice the user.

The RESV message from Dave successfully traverses R2 and Carol's UA receives it. Just as with the Alice to Bob call set-up, Carol sends an UPDATE message to Dave confirming she has qos "e2e" in "sendrecv" directions. Bob acknowledges this with a 200 OK that gives his current status (qos "e2e" and "sendrecv"), and the call set-up in SIP continues to completion.

In summary, Alice set up a call to Bob with RSVP at a priority level of Routine. When Carol called Dave at a high priority, their call will preempt any lower priority calls where these is a contention for resources. In this case, it occurred and affected the call between Alice and Bob. A router at this congestion point preempted Alice's call to Bob in order to place the higher priority call between Carol and Dave. Alice and Bob were both informed of the preemption event. Both Alice and Bob's UAs played preemption indications. What was not mentioned in this appendix was that this document RECOMMENDS R2 (in this example) generating a syslog message to the domain administrator to properly manage and track such events within this domain. This will ensure the domain administrators have recorded knowledge of where such events occur, and what the conditions were that caused them.

Intellectual Property Statement

Disclaimer of Validity

Copyright Statement

Acknowledgment

Funding for the RFC Editor function is currently provided by the Internet Society.