Network Working Group
Request for Comments: 4190
Category: Informational

K. Carlberg
G11
I. Brown
UCL
C. Beard
UMKC

November 2005

Framework for Supporting Emergency Telecommunications Service (ETS) in IP Telephony

Status of This Memo

Copyright Notice

Abstract

This document presents a framework for supporting authorized, emergency-related communication within the context of IP telephony. We present a series of objectives that reflect a general view of how authorized emergency service, in line with the Emergency Telecommunications Service (ETS), should be realized within today's IP architecture and service models. From these objectives, we present a corresponding set of protocols and capabilities, which provide a more specific set of recommendations regarding existing IETF protocols. Finally, we present two scenarios that act as guiding models for the objectives and functions listed in this document. These models, coupled with an example of an existing service in the Public Switched Telephone Network (PSTN), contribute to a constrained solution space.

Table of Contents

1. Introduction

The Internet has become the primary target for worldwide communications in terms of recreation, business, and various imaginative reasons for information distribution. A constant fixture in the evolution of the Internet has been the support of Best Effort as the default service model. Best Effort, in general terms, implies that the network will attempt to forward traffic to the destination as best as it can, with no guarantees being made, nor any resources reserved, to support specific measures of Quality of Service (QoS). An underlying goal is to be "fair" to all the traffic in terms of the resources used to forward it to the destination.

In an attempt to go beyond best effort service, [1] presented an overview of Integrated Services (int-serv) and its inclusion into the Internet architecture. This was followed by [2], which specified the RSVP signaling protocol used to convey QoS requirements. With the addition of [3] and [4], specifying controlled load (bandwidth bounds) and guaranteed service (bandwidth & delay bounds), respectively, a design existed to achieve specific measures of QoS for an end-to-end flow of traffic traversing an IP network. In this case, our reference to a flow is one that is granular in definition and applies to specific application sessions.

From a deployment perspective (as of the date of this document), int-serv has been predominantly constrained to intra-domain paths, at best resembling isolated "island" reservations for specific types of traffic (e.g., audio and video) by stub domains. [5] and [6] will probably contribute to additional deployment of int-serv to Internet Service Providers (ISP) and possibly some inter-domain paths, but it seems unlikely that the original vision of end-to-end int-serv between hosts in source and destination stub domains will become a reality in the near future (the mid- to far-term is a subject for others to contemplate).

In 1998, the IETF produced [7], which presented an architecture for Differentiated Services (diff-serv). This effort focused on a more aggregated perspective and classification of packets than that of [1]. This is accomplished with the recent specification of the diff-serv field in the IP header (in the case of IPv4, it replaced the old ToS field). This new field is used for code points established by IANA, or set aside as experimental. It can be expected that sets of microflows, a granular identification of a set of packets, will correspond to a given code point, thereby achieving an aggregated treatment of data.

One constant in the introduction of new service models has been the designation of Best Effort as the default service model. If traffic is not, or cannot be, associated as diff-serv or int-serv, then it is treated as Best Effort and uses what resources are made available to it.

Beyond the introduction of new services, the continued pace of additional traffic load experienced by ISPs over the years has continued to place a high importance on intra-domain traffic engineering. The explosion of IETF contributions, in the form of drafts and RFCs produced in the area of Multi-Protocol Label Switching (MPLS), exemplifies the interest in versatile and manageable mechanisms for intra-domain traffic engineering. One interesting observation is the work involved in supporting QoS related traffic engineering. Specifically, we refer to MPLS support

of differentiated services [8], and the ongoing work in the inclusion of fast bandwidth recovery of routing failures for MPLS [9].

1.1. Emergency Related Data

The evolution of the IP service model architecture has traditionally centered on the type of application protocols used over a network. By this we mean that the distinction, and possible bounds on QoS, usually centers on the type of application (e.g., audio video tools) that is being referred to.

[10] has defined a priority field for SMTP, but it is only for mapping with X.400 and is not meant for general usage. SIP [11] has an embedded field denoting "priority", but it is only targeted toward the end-user and is not meant to provide an indication to the underlying network or end-to-end applications.

Given the emergence of IP telephony, a natural inclusion of its service is an ability to support existing emergency related services. Typically, one associates emergency calls with "911" telephone service in the U.S., or "999" in the U.K. -- both of which are attributed to national boundaries and accessible by the general public. Outside of this there exist emergency telephone services that involve authorized usage, as described in the following subsection.

1.1.1. Government Emergency Telecommunications Service (GETS)

GETS is an emergency telecommunications service available in the U.S. and is overseen by the National Communications System (NCS) -- an office established by the White House under an executive order [27] and now a part of the Department of Homeland Security. Unlike "911", it is only accessible by authorized individuals. The majority of these individuals are from various government agencies like the Department of Transportation, NASA, the Department of Defense, and the Federal Emergency Management Agency (to name a few). In addition, a select set of individuals from private industry (telecommunications companies, utilities, etc.) that are involved in critical infrastructure recovery operations are also provided access to GETS.

The purpose of GETS is to achieve a high probability that phone service will be available to selected authorized personnel in times of emergencies, such as hurricanes, earthquakes, and other disasters, that may produce a burden in the form of call blocking (i.e., congestion) on the U.S. Public Switched Telephone Network by the general public.

GETS is based in part on the ANSI T1.631 standard, specifying a High Probability of Completion (HPC) for SS7 signaling [12][24].

1.1.2. International Emergency Preparedness Scheme (IEPS)

[25] is a recent ITU standard that describes emergency-related communications over the international telephone service. While systems like GETS are national in scope, IEPS acts as an extension to local or national authorized emergency call establishment and provides a building block for a global service.

As in the case of GETS, IEPS promotes mechanisms like extended queuing, alternate routing, and exemption from restrictive management controls in order to increase the probability that international emergency calls will be established. The specifics of how this is to be accomplished are to be defined in future ITU document(s).

1.2. Scope of This Document

The scope of this document centers on the near and mid-term support of ETS within the context of IP telephony versus Voice over IP. We make a distinction between these two by treating IP telephony as a subset of VoIP, where in the former case, we assume that some form of application layer signaling is used to explicitly establish and maintain voice data traffic. This explicit signaling capability provides the hooks from which VoIP traffic can be bridged to the PSTN.

An example of this distinction is when the Robust Audio Tool (RAT) [13] begins sending VoIP packets to a unicast (or multicast) destination. RAT does not use explicit signaling like SIP to establish an end-to-end call between two users. It simply sends data packets to the target destination. On the other hand, "SIP phones" are host devices that use a signaling protocol to establish a call before sending data towards the destination.

One other aspect we should probably assume exists with IP Telephony is an association of a target level of QoS per session or flow. [28] makes an argument that there is a maximum packet loss and delay for VoIP traffic, and that both are interdependent. For delays of ~200ms, a corresponding drop rate of 5% is deemed acceptable. When delay is lower, a 15-20% drop rate can be experienced and still be considered acceptable. [29] discusses the same topic and makes an argument that packet size plays a significant role in what users tolerate as "intelligible" VoIP. The larger the packet, correlating to a longer sampling rate, the lower the acceptable rate of loss. Note that [28, 29] provide only two of several perspectives in examining VoIP. A more in-depth discussion on this topic is outside

the scope of this document, though it should be noted that the choice of codec can significantly alter the above results.

Regardless of a single and definitive characteristic for stressed conditions, it would seem that interactive voice has a lower threshold of some combinations of loss/delay/jitter than elastic applications such as email or web browsers. This places a higher burden on the problem of supporting VoIP over the Internet. This problem is further compounded when toll-quality service is expected because it assumes a default service model that is better than best effort. This, in turn, can increase the probability that a form of call-blocking can occur with VoIP or IP telephony traffic.

Beyond this, part of our motivation in writing this document is to provide a framework for ISPs and telephony carriers to understand the objectives used to support ETS-related IP telephony traffic. In addition, we also wish to provide a reference point for potential customers in order to constrain their expectations. In particular, we wish to avoid any temptation of trying to replicate the exact capabilities of existing emergency voice service that are currently available in the PSTN to that of IP and the Internet. If nothing else, intrinsic differences between the two communications architectures precludes this from happening. Note, this does not prevent us from borrowing design concepts or objectives from existing systems.

Section 2 presents several primary objectives that articulate what is considered important in supporting ETS-related IP telephony traffic. These objectives represent a generic set of goals and desired capabilities. Section 3 presents additional value-added objectives, which are viewed as useful, but not critical. Section 4 presents protocols and capabilities that relate or can play a role in support of the objectives articulated in Section 2. Finally, Section 5 presents two scenarios that currently exist or are being deployed in the near term over IP networks. These are not all-inclusive scenarios, nor are they the only ones that can be articulated ([34] provides a more extensive discussion on the topology scenarios related to IP telephony). However, these scenarios do show cases where some of the protocols discussed in Section 4 apply, and where some do not.

Finally, we need to state that this document focuses its attention on the IP layer and above. Specific operational procedures pertaining to Network Operation Centers (NOC) or Network Information Centers (NIC) are outside the scope of this document. This includes the "bits" below IP, other specific technologies, and service-level agreements between ISPs and telephony carriers with regard to dedicated links.

2. Objective

The objective of this document is to present a framework that describes how various protocols and capabilities (or mechanisms) can facilitate and support the traffic from ETS users. In several cases, we provide a bit of background in each area so that the reader is given some context and a more in-depth understanding. We also provide some discussion on aspects about a given protocol or capability that could be explored and potentially advanced to support ETS. This exploration is not to be confused with specific solutions since we do not articulate exactly what must be done (e.g., a new header field, or a new code point).

3. Considerations

When producing a solution, or examining existing protocols and mechanisms, there are some things that should be considered. One is that inter-domain ETS communications should not rely on ubiquitous or even widespread support along the path between the end points. Potentially, at the network layer there may exist islands of support realized in the form of overlay networks. There may also be cases where solutions may be constrained on an end-to-end basis (i.e., at the transport or application layer). It is this diversity and possibly partial support that needs to be taken into account by those designing and deploying ETS-related solutions.

Another aspect to consider is that there are existing architectures and protocols from other standards bodies that support emergency-related communications. The effort in interoperating with these systems, presumably through gateways or similar types of nodes with IETF protocols, would foster a need to distinguish ETS flows from other flows. One reason would be the scenario of triggering ETS service from an IP network.

Finally, we take into consideration the requirements of [35, 36] in discussing the protocols and mechanisms below in Section 4. In doing this, we do not make a one-to-one mapping of protocol discussion a requirement. Rather, we make sure the discussion of Section 4 does not violate any of the requirements in [35, 36].

4. Protocols and Capabilities

In this section, we take the objectives presented above and present a set of protocols and capabilities that can be used to achieve them. Given that the objectives are predominantly atomic in nature, the measures used to address them are to be viewed separately with no specific dependency upon each other as a whole. Various protocols and capabilities may be complimentary to each other, but there is no

need for all to exist, given different scenarios of operation; and ETS support is not expected to be an ubiquitously available service. We divide this section into 5 areas:

1) Signaling
2) Policy
3) Traffic Engineering
4) Security
5) Routing

4.1. Signaling and State Information

Signaling is used to convey various information to either intermediate nodes or end nodes. It can be out-of-band of a data flow, and thus in a separate flow of its own, such as SIP messages. It can be in-band and part of the state information in a datagram containing the voice data. This latter example could be realized in the form of diff-serv code points in the IP packet.

In the following subsections, we discuss the current state of some protocols and their use in providing support for ETS. We also discuss potential augmentations to different types of signaling and state information to help support the distinction of emergency-related communications in general.

4.1.1. SIP

With respect to application-level signaling for IP telephony, we focus our attention on the Session Initiation Protocol (SIP). Currently, SIP has an existing "priority" field in the Request-Header-Field that distinguishes different types of sessions. The five values currently defined are: "emergency", "urgent", "normal", "non-urgent", "other-priority". These values are meant to convey importance to the end-user and have no additional semantics associated with them.

[14] is an RFC that defines the requirements for a new header field for SIP in reference to resource priority. The requirements are meant to lead to a means of providing an additional measure of distinction that can influence the behavior of gateways and SIP proxies.

4.1.2. Diff-Serv

In accordance with [15], the differentiated services code point (DSCP) field is divided into three sets of values. The first set is assigned by IANA. Within this set, there are currently, three types of Per Hop Behaviors that have been specified: Default (correlating

to best effort forwarding), Assured Forwarding, and Expedited Forwarding. The second set of DSCP values are set aside for local or experimental use. The third set of DSCP values are also set aside for local or experimental use, but may later be reassigned to IANA if the first set has been completely assigned.

One approach discussed on the IEPREP mailing list is the specification of a new Per-Hop Behaviour (PHB) for emergency-related flows. The rationale behind this idea is that it would provide a baseline by which specific code points may be defined for various emergency-related traffic: authorized emergency sessions (e.g., ETS), general public emergency calls (e.g., "911"), Multi-Level Precedence and Preemption (MLPP) [19], etc. However, in order to define a new set of code points, a forwarding characteristic must also be defined. In other words, one cannot simply identify a set of bits without defining their intended meaning (e.g., the drop precedence approach of Assured Forwarding). The one caveat to this statement are the set of DSCP bits set aside for experimental purposes. But as the name implies, experimental is for internal examination and use and not for standardization.

Note:

It is important to note that at the time this document was written, the IETF had been taking a conservative approach in specifying new PHBs. This is because the number of code points that can be defined is relatively small and is understandably considered a scarce resource. Therefore, the possibility of a new PHB being defined for emergency-related traffic is, at best, a long term project that may or may not be accepted by the IETF.

In the near term, we would initially suggest using the Assured Forwarding (AF) PHB [18] for distinguishing emergency traffic from other types of flows. At a minimum, AF could be used for the different SIP call signaling messages. If the Expedited Forwarding (EF) PHB [40] was also supported by the domain, then it would be used for IP telephony data packets. Otherwise, another AF class would be used for those data flows.

4.1.3. Variations Related to Diff-Serv and Queuing

Scheduling mechanisms like Weighted Fair Queueing and Class Based Queueing are used to designate a percentage of the output link bandwidth that would be used for each class if all queues were backlogged. Its purpose, therefore, is to manage the rates and delays experienced by each class. But emergency traffic may not necessarily require QoS perform any better or differently than non-

emergency traffic. It may just need higher probability of being forwarded to the next hop, which could be accomplished simply by dropping precedences within a class.

To implement preferential dropping between classes of traffic, one of which is emergency traffic, one would probably need to use a more advanced form of Active Queue Management (AQM). Current implementations use an overall queue fill measurement to make decisions; this might cause emergency classified packets to be dropped. One new form of AQM could be a Multiple Average-Multiple Threshold approach, instead of the Single Average-Multiple Threshold approach used today. This allows creation of drop probabilities based on counting the number of packets in the queue for each drop precedence individually.

So, it could be possible to use the current set of AF PHBs if each class were reasonably homogenous in the traffic mix. But one might still have a need to differentiate three drop precedences within non-emergency traffic. If so, more drop precedences could be implemented. Also, if one wanted discrimination within emergency traffic, as with MLPP's five levels of precedence, more drop precedences might also be considered. The five levels would also correlate to a recent effort in Study Group 11 of the ITU to define 5 levels for Emergency Telecommunications Service.

4.1.4. RTP

The Real-Time Transport Protocol (RTP) provides end-to-end delivery services for data with real-time characteristics. The type of data is generally in the form of audio or video type applications, and is frequently interactive in nature. RTP is typically run over UDP and has been designed with a fixed header that identifies a specific type of payload representing a specific form of application media. The designers of RTP also assumed an underlying network providing best effort service. As such, RTP does not provide any mechanism to ensure timely delivery or provide other QoS guarantees. However, the emergence of applications like IP telephony, as well as new service models, present new environments where RTP traffic may be forwarded over networks that support better than best effort service. Hence, the original scope and target environment for RTP has expanded to include networks providing services other than best effort.

In 4.1.2, we discussed one means of marking a data packet for emergencies under the context of the diff-serv architecture. However, we also pointed out that diff-serv markings for specific PHBs are not globally unique, and may be arbitrarily removed or even changed by intermediary nodes or domains. Hence, with respect to

emergency related data packets, we are still missing an in-band marking in a data packet that stays constant on an end-to-end basis.

There are three choices in defining a persistent marking of data packets and thus avoiding the transitory marking of diff-serv code points. One can propose a new PHB dedicated for emergency type traffic as discussed in 4.1.2. One can propose a specification of a new shim layer protocol at some location above IP. Or, one can add a new specification to an existing application layer protocol. The first two cases are probably the "cleanest" architecturally, but they are long term efforts that may not come to pass because of a limited number of diff-serv code points and the contention that yet another shim layer will make the IP stack too large. The third case, placing a marking in an application layer packet, also has drawbacks; the key weakness being the specification of a marking on a per-application basis.

Discussions have been held in the Audio/Visual Transport (AVT) working group on augmenting RTP so that it can carry a marking that distinguishes emergency-related traffic from that which is not. Specifically, these discussions centered on defining a new extension that contains a "classifier" field indicating the condition associated with the packet (e.g., authorized-emergency, emergency, normal) [26]. The rationale behind this idea was that focusing on RTP would allow one to rely on a point of aggregation that would apply to all payloads that it encapsulates. However, the AVT group has expressed a rough consensus that placing an additional classifier state in the RTP header to denote the importance of one flow over another is not an approach they wish to advance. Objections ranging from relying on SIP to convey the importance of a flow, to the possibility of adversely affecting header compression, were expressed. There was also the general feeling that the extension header for RTP that acts as a signal should not be used.

4.1.5. GCP/H.248

The Gateway Control Protocol (GCP) [21] defines the interaction between a media gateway and a media gateway controller. [21] is viewed as an updated version of common text with ITU-T Recommendation H.248 [41] and is a result of applying the changes of RFC 2886 (Megaco Errata) [43] to the text of RFC 2885 (Megaco Protocol version 0.8) [42].

In [21], the protocol specifies a Priority and Emergency field for a context attribute and descriptor. The Emergency is an optional boolean (True or False) condition. The Priority value, which ranges from 0 through 15, specifies the precedence handling for a context.

The protocol does not specify individual values for priority. We also do not recommend the definition of a well known value for the GCP priority as this is out of scope of this document. Any values set should be a function of any SLAs that have been established regarding the handling of emergency traffic.

4.2. Policy

One of the objectives listed in Section 3 above is to treat ETS signaling, and related data traffic, as non-preemptive in nature. Further, this treatment is to be the default mode of operation or service. This is in recognition that existing regulations or laws of certain countries governing the establishment of SLAs may not allow preemptive actions (e.g., dropping existing telephony flows). On the other hand, the laws and regulations of other countries influencing the specification of SLA(s) may allow preemption, or even require its existence. Given this disparity, we rely on local policy to determine the degree by which emergency-related traffic affects existing traffic load of a given network or ISP. Important note: we reiterate our earlier comment that laws and regulations are generally outside the scope of the IETF and its specification of designs and protocols. However, these constraints can be used as a guide in producing a baseline capability to be supported; in our case, a default policy for non-preemptive call establishment of ETS signaling and data.

Policy can be in the form of static information embedded in various components (e.g., SIP servers or bandwidth brokers), or it can be realized and supported via COPS with respect to allocation of a domain’s resources [16]. There is no requirement as to how policy is accomplished. Instead, if a domain follows actions outside of the default non-preemptive action of ETS-related communication, then we stipulate that some type of policy mechanism be in place to satisfy the local policies of an SLA established for ETS-type traffic.

4.3. Traffic Engineering

In those cases where a network operates under the constraints of SLAs, one or more of which pertains to ETS-based traffic, it can be expected that some form of traffic engineering is applied to the operation of the network. We make no recommendations as to which type of traffic engineering mechanism is used, but that such a system exists in some form and can distinguish and support ETS signaling and/or data traffic. We recommend a review of [32] by clients and prospective providers of ETS service that gives an overview and a set of principles of Internet traffic engineering.

MPLS is generally the first protocol that comes to mind when the subject of traffic engineering is brought up. This notion is heightened concerning the subject of IP telephony because of MPLS's ability to permit a quasi-circuit switching capability to be superimposed on the current Internet routing model [30].

However, having cited MPLS, we need to stress that it is an intradomain protocol, and so may or may not exist within a given ISP. Other forms of traffic engineering, such as weighted OSPF, may be the mechanism of choice by an ISP.

As a counter example of using a specific protocol to achieve traffic engineering, [37] presents an example of one ISP relying on a high amount of overprovisioning within its core to satisfy potentially dramatic spikes or bursts of traffic load. In this approach, any configuring of queues for specific customers (neighbors) to support the target QoS is done on the egress edge of the transit network.

Note: As a point of reference, existing SLAs established by the NCS for GETS service tend to focus on a loosely defined maximum allocation of, for example, 1% to 10% of calls allowed to be established through a given LEC using HPC. It is expected, and encouraged, that ETS related SLAs of ISPs will be limited with respect to the amount of traffic distinguished as being emergency related and initiated by an authorized user.

4.4. Security

This section provides a brief overview of the security issues raised by ETS support.

4.4.1. Denial of Service

Any network mechanism that enables a higher level of priority for a specific set of flows could be abused to enhance the effectiveness of denial of service (DoS) attacks. Priority would magnify the effects of attack traffic on bandwidth availability in lower-capacity links, and increase the likelihood of it reaching its target(s). An attack could also tie up resources such as circuits in a PSTN gateway.

Any provider deploying a priority mechanism (such as the QoS systems described in Section 4.1) must therefore carefully apply the associated access controls and security mechanisms. For example, the priority level for traffic originating from an unauthorized part of a network or ingress point should be reset to normal. Users must also be authenticated before being allowed to use a priority service (see Section 4.4.2). However, this authentication process should be lightweight to minimise opportunities for denial of service attacks

on the authentication service itself, and ideally should include its own anti-DoS mechanisms. Other security mechanisms may impose an overhead that should be carefully considered to avoid creating other opportunities for DoS attacks.

As mentioned in Section 4.3, SLAs for ETS facilities often contain maximum limits on the level of ETS traffic that should be prioritised in a particular network (say 1% of the maximum network capacity). This should also be the case in IP networks to again reduce the level of resources that a denial of service attack can consume.

As of this writing, a typical inter-provider IP link uses 1 Gbps Ethernet, OC-48 SONET/SDH, or some similar or faster technology. Also, as of this writing, it is not practical to deploy per-IP packet cryptographic authentication on such inter-provider links, although such authentication might well be needed to provide assurance of IP-layer label integrity in the inter-provider scenario.

While Moore's Law will speed up cryptographic authentication, it is unclear whether that is helpful because the speed of the typical inter-domain link is also increasing rapidly.

4.4.2. User Authorization

To prevent theft of service and reduce the opportunities for denial of service attacks, it is essential that service providers properly verify the authorization of a specific traffic flow before providing it with ETS facilities.

Where an ETS call is carried from PSTN to PSTN via one telephony carrier's backbone IP network, very little IP-specific user authorization support is required. The user authenticates itself to the PSTN as usual -- for example, using a PIN in the US GETS. The gateway from the PSTN connection into the backbone IP network must be able to signal that the flow has an ETS label. Conversely, the gateway back into the PSTN must similarly signal the call's label. A secure link between the gateways may be set up using IPSec or SIP security functionality to protect the integrity of the signaling information against attackers who have gained access to the backbone network, and to prevent such attackers from placing ETS calls using the egress PSTN gateway. If the destination of a call is an IP device, the signaling should be protected directly between the IP ingress gateway and the end device.

When ETS priority is being provided to a flow within one domain, that network must use the security features of the priority mechanism being deployed to ensure that the flow has originated from an authorized user or process.

The access network may authorize ETS traffic over a link as part of its user authentication procedures. These procedures may occur at the link, network, or higher layers, but are at the discretion of a single domain network. That network must decide how often it should update its list of authorized ETS users based on the bounds it is prepared to accept on traffic from recently-revoked users.

If ETS support moves from intra-domain PSTN and IP networks to inter-domain end-to-end IP, verifying the authorization of a given flow becomes more complex. The user's access network must verify a user's ETS authorization if network-layer priority is to be provided at that point.

Administrative domains that agree to exchange ETS traffic must have the means to securely signal to each other a given flow's ETS status. They may use physical link security combined with traffic conditioning measures to limit the amount of ETS traffic that may pass between the two domains. This agreement must require the originating network to take responsibility for ensuring that only authorized traffic is marked with ETS priority, but the recipient network cannot rely on this happening with 100% reliability. Both domains should perform conditioning to prevent the propagation of theft and denial of service attacks. Note that administrative domains that agree to exchange ETS traffic must deploy facilities that perform these conditioning and security services at every point at which they interconnect with one another.

Processes using application-layer protocols, such as SIP, should use the security functionality in those protocols to verify the authorization of a session before allowing it to use ETS mechanisms.

4.4.3. Confidentiality and Integrity

When ETS communications are being used to respond to a deliberate attack, it is important that they cannot be altered or intercepted to worsen the situation -- for example, by changing the orders to first responders such as firefighters, or by using knowledge of the emergency response to cause further damage.

The integrity and confidentiality of such communications should therefore be protected as far as possible using end-to-end security protocols such as IPSec or the security functionality in SIP and SRTP [39]. Where communications involve other types of networks such as the PSTN, the IP side should be protected and any security functionality available in the other network should be used.

4.5. Alternate Path Routing

This subject involves the ability to discover and use a different path to route IP telephony traffic around congestion points, and thus avoid them. Ideally, the discovery process would be accomplished in an expedient manner (possibly even a priori to the need of its existence). At this level, we make no assumptions as to how the alternate path is accomplished, or even at which layer it is achieved -- e.g., the network versus the application layer. But this kind of capability, at least in a minimal form, would help contribute to increasing the probability of ETS call completion by making use of noncongested alternate paths. We use the term "minimal form" to emphasize the fact that care must be taken in how the system provides alternate paths so that it does not significantly contribute to the congestion that is to be avoided (e.g., via excess control/discovery messages).

Routing protocols at the IP network layer, such as BGP and OSPF, contain mechanisms for determining link failure between routing peers. The discovery of this failure automatically causes information to be propagated to other routers. The form of this information, the extent of its propagation, and the convergence time in determining new routes is dependent on the routing protocol in use. In the example of OSPF’s Equal Cost Multiple Path (ECMP), the impact of link failure is minimized because of pre-existing alternate paths to a destination.

At the time this document was written, we can identify two additional areas in the IETF that can be helpful in providing alternate paths for the specific case of call signaling. The first is [9], which is focused on network layer routing and describes a framework for enhancements to the LDP specification of MPLS to help achieve fault tolerance. This, in itself, does not provide alternate path routing, but rather helps minimize loss in intradomain connectivity when MPLS is used within a domain.

The second effort comes from the IP Telephony working group and involves Telephony Routing over IP (TRIP). To date, a framework document [17] has been published as an RFC that describes the discovery and exchange of IP telephony gateway routing tables between providers. The TRIP protocol [20] specifies application level telephony routing regardless of the signaling protocol being used (e.g., SIP or H.323). TRIP is modeled after BGP-4 and advertises reachability and attributes of destinations. In its current form, several attributes have already been defined, such as LocalPreference and MultiExitDisc. Additional attributes can be registered with IANA.

Inter-domain routing is not an area that should be considered in terms of additional alternate path routing support for ETS. The Border Gateway Protocol is currently strained in meeting its existing requirements, and thus adding additional features that would generate an increase in advertised routes will not be well received by the IETF. Refer to [38] for a commentary on Inter-Domain routing.

4.6. End-to-End Fault Tolerance

This topic involves work that has been done in trying to compensate for lossy networks providing best effort service. In particular, we focus on the use of a) Forward Error Correction (FEC), and b) redundant transmissions that can be used to compensate for lost data packets. (Note that our aim is fault tolerance, as opposed to an expectation of always achieving it.)

In the former case, additional FEC data packets are constructed from a set of original data packets and inserted into the end-to-end stream. Depending on the algorithm used, these FEC packets can reconstruct one or more of the original set that were lost by the network. An example may be in the form of a 10:3 ratio, in which 10 original packets are used to generate three additional FEC packets. Thus, if the network loses 30% of packets or less, then the FEC scheme will be able to compensate for that loss. The drawback to this approach is that, to compensate for the loss, a steady state increase in offered load has been injected into the network. This makes an argument that the act of protection against loss has contributed to additional pressures leading to congestion, which in turn helps trigger packet loss. In addition, by using a ratio of 10:3, the source (or some proxy) must "hold" all 10 packets in order to construct the three FEC packets. This contributes to the end-to-end delay of the packets, as well as minor bursts of load, in addition to changes in jitter.

The other form of fault tolerance we discuss involves the use of redundant transmissions. By this we mean the case in which an original data packet is followed by one or more redundant packets. At first glance, this would appear to be even less friendly to the network than that of adding FEC packets. However, the encodings of the redundant packets can be of a different type (or even transcoded into a lower quality) that produce redundant data packets that are significantly smaller than the original packet.

Two RFCs [22, 23] have been produced that define RTP payloads for FEC and redundant audio data. An implementation example of a redundant audio application can be found in [13]. We note that both FEC and redundant transmissions can be viewed as rather specific, and to a degree tangential, solutions regarding packet loss and emergency

communications. Hence, these topics are placed under the category of value-added objectives.

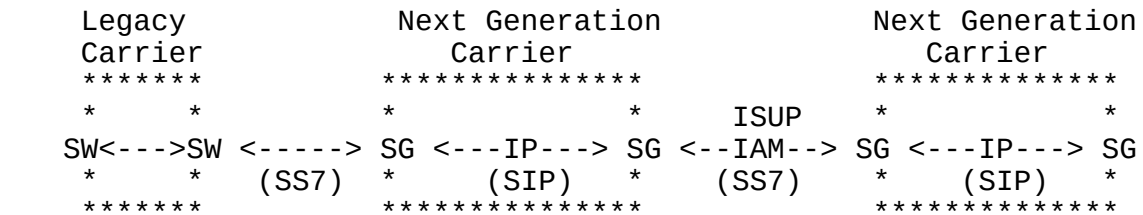
5. Key Scenarios

There are various scenarios in which IP telephony can be realized, each of which can imply a unique set of functional requirements that may include just a subset of those listed above. We acknowledge that a scenario may exist whose functional requirements are not listed above. Our intention is not to consider every possible scenario by which support for emergency related IP telephony can be realized. Rather, we narrow our scope using a single guideline; we assume there is a signaling and data interaction between the PSTN and the IP network with respect to supporting emergency-related telephony traffic. We stress that this does not preclude an IP-only end-to-end model, but rather the inclusion of the PSTN expands the problem space and includes the current dominant form of voice communication.

Note: as stated in Section 1.2, [32] provides a more extensive set of scenarios in which IP telephony can be deployed. Our selected set below is only meant to provide a couple of examples of how the protocols and capabilities presented in Section 3 can play a role.

5.1. Single IP Administrative Domain

This scenario is a direct reflection of the evolution of the PSTN. Specifically, we refer to the case in which data networks have emerged in various degrees as a backbone infrastructure connecting PSTN switches at its edges. This scenario represents a single isolated IP administrative domain that has no directly adjacent IP domains connected to it. We show an example of this scenario below in Figure 1. In this example, we show two types of telephony carriers. One is the legacy carrier, whose infrastructure retains the classic switching architecture attributed to the PSTN. The other is the next generation carrier, which uses a data network (e.g., IP) as its core infrastructure, and Signaling Gateways at its edges. These gateways "speak" SS7 externally with peering carriers, and another protocol (e.g., SIP) internally, which rides on top of the IP infrastructure.



SW - Telco Switch, SG - Signaling Gateway

Figure 1

The significant aspect of this scenario is that all the resources f each IP "island" falls within a given administrative authority. Hence, there is not a problem in retaining PSTN type QoS for voice traffic (data and signaling) exiting the IP network. Thus, the need for support of mechanisms like diff-serv in the presence of overprovisioning, and an expansion of the defined set of Per-Hop Behaviors, is reduced under this scenario.

Another function that has little or no importance within the closed IP environment of Figure 1 is that of IP security. The fact that each administrative domain peers with each other as part of the PSTN, means that existing security, in the form of Personal Identification Number (PIN) authentication (under the context of telephony infrastructure protection), is the default scope of security. We do not claim that the reliance on a PIN-based security system is highly secure or even desirable. But, we use this system as a default mechanism in order to avoid placing additional requirements on existing authorized emergency telephony systems.

5.2. Multiple IP Administrative Domains

We view the scenario of multiple IP administrative domains as a superset of the previous scenario. Specifically, we retain the notion that the IP telephony system peers with the existing PSTN. In addition, segments (i.e., portions of the Internet) may exchange signaling with other IP administrative domains via non-PSTN signaling protocols like SIP.

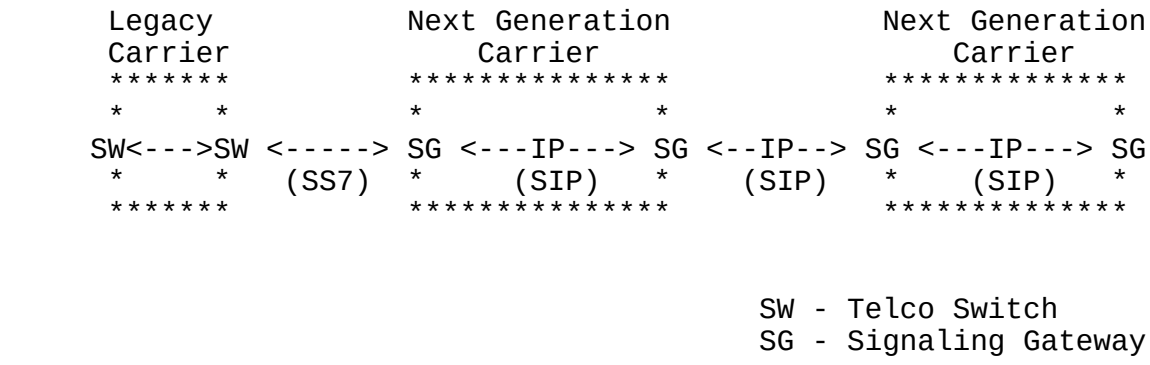


Figure 2

Given multiple IP domains, and the presumption that SLAs relating to ETS traffic may exist between them, the need for something like diff-serv grows with respect to being able to distinguish the emergency related traffic from other types of traffic. In addition, IP security becomes more important between domains in order to ensure that the act of distinguishing ETS-type traffic is indeed valid for the given source.

We conclude this section by mentioning a complementary work in progress in providing ISUP transparency across SS7-SIP interworking [33]. The objective of this effort is to access services in the SIP network and yet maintain transparency of end-to-end PSTN services.

Not all services are mapped (as per the design goals of [33]), so we anticipate the need for an additional document to specify the mapping between new SIP labels and existing PSTN code points like NS/EP and MLPP.

6. Security Considerations

Information on this topic is presented in sections 2 and 4.

7. Informative References

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

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

[3] Shenker, S., Partridge, C., and R. Guerin, "Specification of Guaranteed Quality of Service", RFC 2212, September 1997.

[4] Wroclawski, J., "Specification of the Controlled-Load Network Element Service", RFC 2211, September 1997.

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

[6] Berger, L., Gan, D., Swallow, G., Pan, P., Tommasi, F., and S. Molendini, "RSVP Refresh Overhead Reduction Extensions", RFC 2961, April 2001.

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

[8] Le Faucheur, F., Wu, L., Davie, B., Davari, S., Vaananen, P., Krishnan, R., Cheval, P., and J. Heinanen, "Multi-Protocol Label Switching (MPLS) Support of Differentiated Services", RFC 3270, May 2002.

[9] Sharma, V. and F. Hellstrand, "Framework for Multi-Protocol Label Switching (MPLS)-based Recovery", RFC 3469, February 2003.

[10] Kille, S., "MIXER (Mime Internet X.400 Enhanced Relay): Mapping between X.400 and RFC 822/MIME", RFC 2156, January 1998.

[11] 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.

[12] ANSI, "Signaling System No. 7(SS7), High Probability of Completion (HPC) Network Capability", ANSI T1.631-1993, (R1999).

[13] Robust Audio Tool (RAT): http://www-mice.cs.ucl.ac.uk/multimedia/software/rat

[14] Schulzrinne, H., "Requirements for Resource Priority Mechanisms for the Session Initiation Protocol (SIP)", RFC 3487, February 2003.

[15] 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.

[16] Durham, D., Boyle, J., Cohen, R., Herzog, S., Rajan, R., and A. Sastry, "The COPS (Common Open Policy Service) Protocol", RFC 2748, January 2000.

[17] Rosenberg, J. and H. Schulzrinne, "A Framework for Telephony Routing over IP", RFC 2871, June 2000.

[18] Heinanen, J., Baker, F., Weiss, W., and J. Wroclawski, "Assured Forwarding PHB Group", RFC 2597, June 1999.

[19] ITU, "Multi-Level Precedence and Preemption Service, ITU, Recommendation, I.255.3, July, 1990.

[20] Rosenberg, J., Salama, H., and M. Squire, "Telephony Routing over IP (TRIP)", RFC 3219, January 2002.

[21] Groves, C., Pantaleo, M., Anderson, T., and T. Taylor, "Gateway Control Protocol Version 1", RFC 3525, June 2003.

[22] Perkins, C., Kouvelas, I., Hodson, O., Hardman, V., Handley, M., Bolot, J., Vega-Garcia, A., and S. Fosse-Parisis, "RTP Payload for Redundant Audio Data", RFC 2198, September 1997.

[23] Rosenberg, J. and H. Schulzrinne, "An RTP Payload Format for Generic Forward Error Correction", RFC 2733, December 1999.

[24] ANSI, "Signaling System No. 7, ISDN User Part", ANSI T1.113-2000, 2000.

[25] "Description of an International Emergency Preference Scheme (IEPS)", ITU-T Recommendation E.106 March, 2002

[26] Carlberg, K., "The Classifier Extension Header for RTP", Work In Progress, October 2001.

[27] National Communications System: http://www.ncs.gov

[28] Bansal, R., Ravikanth, R., "Performance Measures for Voice on IP", http://www.ietf.org/proceedings/97aug/slides/tsv/ippm-voiceip/, IETF Presentation: IPPM-Voiceip, Aug, 1997

[29] Hardman, V., et al, "Reliable Audio for Use over the Internet", Proceedings, INET'95, Aug, 1995.

[30] Awduche, D., Malcolm, J., Agogbua, J., O'Dell, M., and J. McManus, "Requirements for Traffic Engineering Over MPLS", RFC 2702, September 1999.

[31] "Service Class Designations for H.323 Calls", ITU Recommendation H.460.4, November, 2002.

[32] Awduche, D., Chiu, A., Elwalid, A., Widjaja, I., and X. Xiao, "Overview and Principles of Internet Traffic Engineering", RFC 3272, May 2002.

[33] Vemuri, A. and J. Peterson, "Session Initiation Protocol for Telephones (SIP-T): Context and Architectures", BCP 63, RFC 3372, September 2002.

[34] Polk, J., "Internet Emergency Preparedness (IEPREP) Telephony Topology Terminology", RFC 3523, April 2003.

[35] Carlberg, K. and R. Atkinson, "General Requirements for Emergency Telecommunication Service (ETS)", RFC 3689, February 2004.

[36] Carlberg, K. and R. Atkinson, "IP Telephony Requirements for Emergency Telecommunication Service (ETS)", RFC 3690, February 2004.

[37] Meyers, D., "Some Thoughts on CoS and Backbone Networks" http://www.ietf.org/proceedings/02nov/slides/ieprep-4.pdf IETF Presentation: IEPREP, Dec, 2002.

[38] Huston, G., "Commentary on Inter-Domain Routing in the Internet", RFC 3221, December 2001.

[39] Baugher, M., McGrew, D., Naslund, M., Carrara, E., and K. Norrman, "The Secure Real-time Transport Protocol (SRTP)", RFC 3711, March 2004.

[40] Davie, B., Charny, A., Bennet, J.C., 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.

[41] ITU, "Gateway Control Protocol", Version 3, ITU, September, 2005.

[42] Cuervo, F., Greene, N., Huitema, C., Rayhan, A., Rosen, B., and J. Segers, "Megaco Protocol version 0.8", RFC 2885, August 2000.

[43] Taylor, T., "Megaco Errata", RFC 2886, August 2000.

Appendix A: Government Telephone Preference Scheme (GTPS)

This framework document uses the T1.631 and ITU IEPS standard as a target model for defining a framework for supporting authorized emergency-related communication within the context of IP telephony. We also use GETS as a helpful model from which to draw experience. We take this position because of the various areas that must be considered; from the application layer to the (inter)network layer, in addition to policy, security (authorized access), and traffic engineering.

The U.K. has a different type of authorized use of telephony services, referred to as the Government Telephone Preference Scheme (GTPS). At present, GTPS only applies to a subset of the local loop lines within the UK. The lines are divided into Categories 1, 2, and 3. The first two categories involve authorized personnel involved in emergencies such as natural disasters. Category 3 identifies the general public. Priority marks, via C7/NUP, are used to bypass call-gapping for a given Category. The authority to activate GTPS has been extended to either a central or delegated authority.

A.1. GTPS and the Framework Document

The design of the current GTPS, with its designation of preference based on physical static devices, precludes the need for several aspects presented in this document. However, one component that can have a direct correlation is the labeling capability of the proposed Resource Priority extension to SIP. A new label mechanism for SIP could allow a transparent interoperation between IP telephony and the U.K. PSTN that supports GTPS.

Appendix B: Related Standards Work

The process of defining various labels to distinguish calls has been, and continues to be, pursued in other standards groups. As mentioned in Section 1.1.1, the ANSI T1S1 group has previously defined a label in the SS7 ISUP Initial Address Message. This single label or value is referred to as the National Security and Emergency Preparedness (NS/EP) indicator and is part of the T1.631 standard. The following subsections presents a snapshot of parallel, on-going efforts in various standards groups.

It is important to note that the recent activity in other groups have gravitated to defining 5 labels or levels of priority. The impact of this approach is minimal in relation to this ETS framework document because it simply generates a need to define a set of corresponding labels for the resource priority header of SIP.

B.1. Study Group 16 (ITU)

Study Group 16 (SG16) of the ITU is responsible for studies relating to multimedia service definition and multimedia systems, including protocols and signal processing.

A contribution [31] has been accepted by this group that adds a Priority Class parameter to the call establishment messages of H.323. This class is further divided into two parts; one for Priority Value and the other is a Priority Extension for indicating subclasses. It is this former part that roughly corresponds to the labels transported via the Resource Priority field for SIP [14].

The draft recommendation advocates defining PriorityClass information that would be carried in the GenericData parameter in the H323-UU-PDU or RAS messages. The GenericData parameter contains PriorityClassGenericData. The PriorityClassInfo of the PriorityClassGenericData contains the Priority and Priority Extension fields.

At present, 4 levels have been defined for the Priority Value part of the Priority Class parameter: Normal, High, Emergency-Public, Emergency-Authorized. An additional 8-bit priority extension has been defined to provide for subclasses of service at each priority.

The suggested ASN.1 definition of the service class is the following:

```
CALL-PRIORITY {itu-t(0) recommendation(0) h(8) 460 4 version1(0)}
DEFINITIONS AUTOMATIC TAGS::=

BEGIN
IMPORTS
   ClearToken,
   CryptoToken
    FROM H235-SECURITY-MESSAGES;

CallPriorityInfo::= SEQUENCE
{
  priorityValue  CHOICE
   {
     emergencyAuthorized     NULL,
     emergencyPublic         NULL,
     high                    NULL,
     normal                  NULL,
     ...
   },

  priorityExtension   INTEGER (0..255)  OPTIONAL,
```

```
        tokens              SEQUENCE OF ClearToken       OPTIONAL,
        cryptoTokens        SEQUENCE OF CryptoToken    OPTIONAL,
        rejectReason        CHOICE
        {
            priorityUnavailable         NULL,
            priorityUnauthorized        NULL,
            priorityValueUnknown        NULL,
            ...
        } OPTIONAL,        -- Only used in CallPriorityConfirm
        ...
      }
```

The advantage of using the GenericData parameter is that an existing parameter is used, as opposed to defining a new parameter and causing subsequent changes in existing H.323/H.225 documents.

Acknowledgements

The authors would like to acknowledge the helpful comments, opinions, and clarifications of Stu Goldman, James Polk, Dennis Berg, Ran Atkinson as well as those comments received from the IEPS and IEPREP mailing lists. Additional thanks to Peter Walker of Oftel for private discussions on the operation of GTPS, and Gary Thom on clarifications of the SG16 draft contribution.

Authors' Addresses

Ken Carlberg
University College London
Department of Computer Science
Gower Street
London, WC1E 6BT
United Kingdom

EMail: k.carlberg@cs.ucl.ac.uk

Ian Brown
University College London
Department of Computer Science
Gower Street
London, WC1E 6BT
United Kingdom

EMail: I.Brown@cs.ucl.ac.uk

Cory Beard
University of Missouri-Kansas City
Division of Computer Science
Electrical Engineering
5100 Rockhill Road
Kansas City, MO 64110-2499
USA

EMail: BeardC@umkc.edu

Full Copyright Statement

Intellectual Property

Acknowledgement

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