Internet Engineering Task Force
INTERNET DRAFT
October 18 2001

Ken Carlberg
UCL
Ian Brown
UCL

Framework for Supporting IEPS in IP Telephony
<draft-carlberg-ieps-framework-02.txt>

Status of this Memo

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 International Emergency Preparedness Scheme (IEPS), should be realized within today's IP architecture and service models. From these objectives, we present a corresponding set of functional requirements, 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 PSTN, contribute to a constrained solution space.

1. Introduction

The Internet has become the primary target for worldwide communications. This is 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, infers 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, [2] presented an overview of Integrated Services (int-serv) and its inclusion into the Internet architecture. This was followed by [3], which specified the RSVP signaling protocol used to convey QoS requirements. With the addition of [4] and [5], specifying control 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 applying to specific application sessions.

From a deployment perspective (as of the date of this document), int-serv has been predominantly constrained to stub intra-domain paths, at best resembling isolated "island" reservations for specific types of traffic (e.g., audio and video) by stub domains. [6] and [7] 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 [8], which presented an architecture for Differentiated Services (diff-serv). This effort focused on a more aggregated perspective and classification of packets than that of [2]. 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 for 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 sensitive traffic like Voice over IP (VoIP). Specifically, we refer to the work in progress discussion of a framework to support VoIP using MPLS [9], and the inclusion of fault tolerance [10]. This latter item can be viewed as being similar to "crank-back", a term used to describe the means by which the Public Switched Telephone Network (PSTN) routes around congested switches.

1.2 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).

While protocols like SMTP [11] and SIP [12] have embedded fields denoting "priority", there has not been a previous IETF standards based effort to state or define what this distinction means with respect to the underlying network and how it should be supported. Given the emergence of IP telephony, a natural inclusion of it as part of a telco carriers backbone network, or into the Internet as a whole, implies the 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 exists emergency telephone services that involved authorized usage, as described in the following subsection.

GETS is an emergency telecommunications service available in the U.S. and established by the National Communications System (NCS) -- an office established by the White House under an executive order. 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 but a few). In addition, individuals from private industry

(telecommunications companies, utilities, etc.) that are involved in criticial infrastructure recovery operations are also provided access to GETS.

The purpose of GETS is to increase the probability that phone service will be available to selected government agency 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.

The key aspect is that GETS only supports a probabilistic approach to call completion, as opposed to call preemption. This distinction is important because emergency systems like GETS are not allowed to terminate existing calls in order to allow a GETS call to be established. Thus, the mechanisms and specifications that comprise GETS only focus on increasing the chances that a particular telephone call will be established.

The basis for GETS with respect to Signaling System 7 (SS7) support is found in the T1.631 protocol on High Probability of Completion (HPC) network capability [13]. This document describes the specification of a National Security and Emergency Preparedness (NS/EP) Calling Party Category (CPC) code point used for SS7 ISDN User Part (ISUP) Initial Address Message (IAM). In the presense of this code point, Local Exchange Carriers (LEC) will attempt (if necessary and if the option is supported) to route the call through alternate inter-exchange carriers (IXC) if it cannot complete the call through the default IXC.

The procedure for a user (i.e., a person) establishing a GETS call is as follows:

1) Dial a non-geographical area code number: 710-XXX-XXXX
2) Dial a PIN used to authenticate the call
3) Dial the actual destination number to be reached

In conjunction with the above, the source LEC (where the call originated) attempts to establish the call through an IXC. This is done even if the destination number is within the LEC itself. If the IXC cannot forward the call to the destination LEC, then the source LEC attempts to route the call through an alternate IXC. If alternate IXCs cannot help establish the call, then a busy signal is finally returned to the user. Otherwise, the call is completed and retains the same quality of service as all other telephone calls.

The HPC component of GETS is not ubiquitously supported by the U.S. PSTN. The only expectation is that the 710 area code is recognized

by all carriers. Additional support is conditional and dependent upon the equivalent service level agreements established between the U.S. Government and various telco carriers. Thus, the default end-to-end service for establishing a GETS call can be viewed as best effort and associated with the same priority as calls from the general public. The exception to this rule is when the call is forwarded through carriers that have been contracted through a service level agreement to support HPC, which results in a higher probability that the GETS call will be established.

It should be noted from the above description that GETS is separate and unrelated to other emergency services like "911".

1.2.1 International Emergency Preparedness Scheme (IEPS)

[18] is a recent ITU standard that describes emergency related communications over international telephone service (Note, this document has also been published as a draft-RFC in [28]. 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).

1.3 Scope of this Document

The scope of this document centers on the support of IEPS within the context of IP telephony, though not necessarily Voice over IP. We make a distinction between these two by treating IP telephony as a subset of VoIP, where in the former we assume 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 Redundant Audio Tool (RAT) [14] 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 signal before sending data towards the destination.

Beyond this, part of our motivation in writing this document is to provide a reference point for ISPs and carriers so that they have an understanding of objectives and accompanying functional requirements used to support IEPS related IP telephony traffic. In addition, we also wish to provide a reference point for potential customers (users of IEPS) 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 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 IEPS related IP telephony traffic. These objectives represent a generic set of goals and capabilities attributed to supporting IEPS based IP telephony. Section 3 presents additional value added objectives. These are capabilities that are viewed as useful, but not critical in support of IEPS. Section 4 presents a series of functional requirements that stem from the objectives articulated in section 2. Finally, Section 5 presents two scenarios in IEPS that exist or are being deployed over IP networks. These are not all-inclusive scenarios, nor are they the only ones that can be articulated. However, they do show cases where some of the functional requirements 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 carriers with regard to dedicated links.

2. Objective

The support of IEPS within IP telephony can be realized in the form of several primary objectives. These objectives define the generic functions or capabilities associated with IEPS, and the scope of the support needed to achieve these capabilities. From this generic set of objectives, we present specific functional requirements of existing IP protocols (presented below in section 3).

There are two underlying goals in the selection of these objectives. One goal is to produce a design that maximizes the use of existing IP protocols and minimizes the set of additional specifications needed to support IP-telephony based IEPS. Thus, with the inclusion of these minimal augmentations, the bulk of the work in achieving IEPS

over an IP network that is connected or unconnected to the Internet involves operational issues. Examples of this would be the establishment of Service Level Agreements (SLA) with ISPs, and/or the provisioning of traffic engineered paths for IEPS-related telephony traffic.

A second underlying goal in selecting the following objectives is to take into account experiences from an existing emergency-type communication system (as described in section 1.2.1) as well as the existing restrictions and constraints placed by some countries. In the former case, we do not attempt to mimic the system, but rather extract information as a reference model. With respect to constraints based on laws or agency regulations, this would normally be considered outside of the scope of any IETF document. However, these constraints act as a means of determining the lowest common denominator in specifying technical functional requirements. If such constraints do not exist, then additional functions can be added to the baseline set of functions. This last item will be expanded upon in the description of Objective #3 below.

The following list of objectives are termed primary because they pertain to that which defines the underlying goals of IEPS in relation to IP telephony. However, the primary objectives are not meant to dictate major overhauls of existing IP protocols, nor do they require new protocols to be developed.

Primary Objectives in support of authorized emergency calls:

1) High Probability of Call Completion
2) Interaction with PSTN
3) Distinction of IEPS data traffic
4) Non-preemptive action
5) Non-ubiquitous support
6) Authenticated service

The first objective is the crux of our work because it defines our expectations for both data and call signaling for IP telephony. As stated, our objective is achieving a high probability that emergency related calls (both data and signaling) will be forwarded through an IP network. Specifically, we envision the relevance of this objective during times of congestion, the context of which we describe further below in this section. The critical word in this objective is "probability", as opposed to assurance or guarantee -- the latter two placing a higher burden on the network. It stands to reason, though, that the word "probability" is a less tangible description that cannot be easily quantified. It is relative in relation to other traffic transiting the same network. Objectives 3 through 5 below help us to qualify the term probability in the

context of other objectives.

The second objective involves the interaction of IP telephony signaling with existing PSTN support for emergency related voice communications. As mentioned above in Section 1.2, standard T1.631 [26] specifies emergency code points for SS7. Specifically, the National Security and Emergency Preparedness (NS/EP) Calling Party Category code point is defined for ISUP IAM messages used by SS7 [26]. Hence, our objective in the interaction between the PSTN and IP telephony with respect to IEPS (and national indicators) is a direct mapping between related code points.

The third objective focuses on the ability to distinguish IEPS data packets from other types of VoIP packets. With such an ability, transit providers can more easily ensure that service level agreements relating to IEPS are adhered. Note that we do not assume that the actions taken to distinguish IEPS type packets is easy. Nor, in this section, do we state the form of this distinction. We simply present the objective of identifying flows that relate to IEPS versus others that traverse a transit network.

At an abstract level, the forth objective pertains to the actions taken when an IP telephony call, via a signaling protocol such as SIP, cannot be forwarded because the network is experiencing a form of congestion. We state this in general terms because of two reasons: a) there may exist applications other than SIP, like H.248, used for call establishment, and b) congestion may come in several forms. For example, congestion may exist at the IP packet layer with respect to queues being filled to their configured limit. Congestion may also arise from resource allocation attributed per call or aggregated sets of calls. In this latter case, while there may exist resources to forward the packets, a signaling server may have reached its limit as to how many telephony calls it will support while retaining toll-quality service per call. Typically, one terms this form of congestion as call blocking. Note that we do not address the case when congestion occurs at the bit level below that of IP, due to the position that it is outside the scope of IP and the IETF.

So, given the existence of congestion in its various forms, our objective is to support IEPS-related IP telephony call signaling and data traffic via non-preemptive actions taken by the network. More specifically, we associate this objective in the context of IP telephony acting as part of the Public Telephone Network (PTN). This, as opposed to the use of IP telephony within a private or stub network. In section 5 below, we expand on this through the description of two distinct scenarios of IP telephony and its operation with IEPS and the PSTN.

It is important to mention that this is a default objective influenced by existing laws & regulations. Those countries not bound by these restrictions can remove this objective and make provisions to enforce preemptive action. In this case, it would probably be advantageous to deploy a signaling system similar to that proposed in [15], wherein multiple levels of priority are defined and preemption via admission control from SIP servers is enforced.

The fifth objective stipulates that we do not advocate the need or expectation for ubiquitous support of IEPS across all administrative domains of the Internet. While it would be desirable to have ubiquitous support, we feel the reliance of such a requirement would doom even the contemplation of supporting IEPS by the IETF and the expected entities (e.g., ISPs and vendors) involved in its deployment.

We use the existing GETS service in the U.S. as an existing example in which emergency related communications does not need to be ubiquitous. As mentioned previously, the measure and amount of support provided by the U.S. PSTN for GETS is not ubiquitous across all U.S. Inter-exchange Carriers (IXC) nor Local Exchange Carriers (LEC). Given the fact that GETS still works within this context, it is our objective to follow this deployment model such that we can accomplish the first objective listed above -- a higher probability of call completion than that of normal IP telephony call traffic.

Our final objective is that only authorized users may use the services outlined in this framework. GETS users are authenticated using a PIN provided to the telecommunications carrier, which signals approval back to the user's local exchange over SS7. In an IP network, the authentication center will need to securely signal back to the IP ingress point that a given user is authorized to send IEPS related flows. Similarly, transit networks with IEPS SLAs must securely interchange authorized IEPS traffic. In both cases, IPSec authentication transforms may be used to protect this traffic. This is entirely separate from end-to-end IPSec protection of user traffic, which will be configured by users. IP-PSTN gateways must also be able to securely signal IEPS authorization for a given flow. As these gateways are likely to act as SIP servers, we further consider the use of SIP's security functions to aid this objective.

3. Value Added Objective

This objective is viewed as being helpful in achieving a high probability of call completion. Its realization within an IP network would be in the form of new protocols or enhancements to existing ones. Thus, objective listed in this section are treated as value

added -- an expectation that their existence would be beneficial, and yet not viewed as critical to support IEPS related IP telephony traffic.

3.1 Alternate Path Routing

This objective 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 requirements 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 call completion of IEPS traffic by making use of noncongested alternate paths. We use the term "minimal form" to concede the fact that care must be taken in how the system provides alternate paths so it does not significantly contribute to the congestion that is to be avoided (e.g., via excess control/discovery messages).

At the time that this document was written, we can identify two work-in-progress areas in the IETF that can be helpful in providing alternate paths for call signaling. The first is [21], which is focused on network layer routing and describes 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 [19] has been published as an RFC which describes the discovery and exchange of IP telephony gateway routing tables between providers. The TRIP protocol [22], a supplemental work in progress, 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. Upon standardization of TRIP, additional attributes can be registered with IANA. Initially, we would recommend two additional attributes that would relate to emergency related flows. These being:

EmergencyMultiExitDisc

The EmergencyMultiExitDisc attribute is similar to the MultiExitDisc in that it is an inter-domain attribute used to express a preference of one or more links over others between domains. Unlike the MultiExitDisc, this attribute specifically identifies links that are preferred for emergency related calls that span domains.

EmergencyLocalPreference

The EmergencyLocalPreference attribute is similar to the LocalPreference in that it is an intra-domain attribute used to inform other LSs of the local LSs preference for a given route. The difference between the two types attributes is that the preferred route specifically relates to emergency-type calls (e.g., 911). This attribute has no significance between domains. Local policy determines if there is an association between the EmergencyLocalPreference and the EmergencyMultiExitDisc attribute.

3.2 End-to-End Fault Tolerance

This topic involves the 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% or less number of packets, 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 arguement that the act of protection against loss has contributed to additional pressures leading to congestion, which in turn helps trigger packet loss. In addition, in 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 [24, 25] 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 [14]. 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.

4. Functional Requirements

In this section, we take the objectives presented above and specify a corresponding set of functional requirements to achieve them. Given that the objectives are predominantly atomic in nature, the corresponding functional requirements are to be viewed separately with no specific dependency upon each other as a whole. They may be complimentary with each other, but there is no need for all to exist given different scenarios of operation, and that IEPS support is not viewed as a ubiquitously available service. We divide the functional requirements into 4 areas:

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

4.1 Signaling

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 also be in-band and part of the datagram containing the voice data. This latter example could also be in the form of diff-serv code points in the IP packet, and/or in an extension to the RTP header denoting an additional classification of the payload -- the latter predominantly being used on an end-to-end basis.

In the following subsections, we discuss augmentations to three specific types of signaling to help support the distinction of emergency related communications in general, and IEPS specifically. We also discuss Media Gateway Control (MEGACO).

4.1.1 SIP

With respect to application level signaling for IP telephony, we focus our attention to 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 currently defined values are: "emergency", "urgent", "normal", "non-urgent", "other-priority".

It is understood that the IETF prefers that no changes or additions be made to these existing values. Hence, we shall follow the approach taken in [15] and propose the specification of a new field in the "Request-Header-Field" titled "Emergency-State". This new field provides an additional level in distinguishing types of emergencies. Currently, we would propose defining two values for this field:

1) "authorized-emergency"
2) "general-emergency"

The former would correlate to calls that have been initiated by an authorized individual. Specifically, this single SIP value would correlate to other authorized PSTN based code points like NS/EP and IEPS. The second defined value would correlate to the more commonly known type of local emergency calls initiated by the general public (e.g., "911" in the U.S., "999", in the UK, and "112" in Germany). The objective is to define a single generic value that correlates to several similar but different types of emergency calls.

It is important to note that this is the one functional requirement that is considered mandatory with respect to supporting IEPS within IP telephony. We take this position because regardless of the extent by which the underlying network supports IEPS-based traffic, there is a need to distinguish IEPS sessions (i.e., authorized-emergency calls) from others.

The existence of this new value in the SIP priority field allows an IP telephony domain to map an IEPS call to the existing NS/EP code point from an SS7 telephony domain. This will help facilitate a seamless interaction between the PSTN and the an IP network acting as either an internal backbone or as a peering ISP.

Author's Note: The work put forth by James Polk in [15] is quite similar to our own in that both articulate a need to specify a more granular and specific means of identifying different types of emergencies. Beyond the different values specified for MLPP, the main difference between the two efforts involves the use of preemption for [15], as opposed to our need to simply increase the probability of call completion.

4.1.2 Diff-Serv

In accordance to [16], 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 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 local or experimental use, but may later be reassigned to IANNA in case the first set has been completely assigned.

One recomendation involves the specification of a new type of Per-Hop Behavior (PHB) we term Emergency Related Forwarding (ERF). This would provide a specific means of distinguishing emergency related traffic (signaling and user data) from other traffic. The existence of this PHB then provides a baseline by which specific code points may be defined related to various emergency related traffic: authorized emergency sessions (e.g., IEPS), general public emergency calls (e.g., "911"), MLPP. Aggregates would still exist with respect to the bundling of applications per code point. Further, one would associate a forwarding paradigm aimed at a low loss rate reflective of the code point selected. Hence, SIP or H.323 messages marked with "authorized-emergency" or "emergency" may be assigned a code point reflecting a lower loss rate than other type of traffic (even the emergency-related data flow itself). The jitter associated with application layer signaling of IP telephony would be inversely important with respect to loss rate, and thus would be reflective of the code points defined for the proposed PHB.

Another recomendation would be to define a new or fifth class for the existing AF PHB. Unlike the other currently defined classes, this new one would be based on five levels of drop precedence. This increase in the number of levels would conveniently correlate to the worst case scenario posed by MLPP, which has five types of priorities. In addition, it would provide a higher level of variance that could be used to supercede the existing 3 levels used in the other classes. Hence, if other non-emergency aggregate traffic where assigned to the class, the highest drop precedence they are assigned

to is (3) -- corresponding to the other four currently defined classes. Emergency traffic would be set to (4) or (5), depending on the SLA tht has been defined.

It is important to note that as of the time that this document was written, the IETF is taking a conservative approach in specifying new PHBs. This is because the number of code points that can be defined is relatively small, and thus 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 meantime, we would initially recommend using the Assured Forwarding (AF) PHB [20] for distinguishing emergency traffic from other types of flows. Specifically, we would suggest the use the low drop precedence of one of the four defined classes of AF codepoints. It is critical to note that one cannot specify an exact code point used for emergency related data flows because the relevance of a code point is local to the given diff-serv domain (i.e., they are not globally unique per micro-flow or aggregate of flows). In addition, we can expect that the existence of a codepoint for emergency related flows is based on the service level agreements established with a given diff-serv domain.

4.1.3 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 are 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 -- typically 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, presents 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.

We have three choices in defining a persistent marking of data packets and thus avoid the transitory marking of diff-serv code points. We can propose a new PHB dedicated for emergency type traffic as discussed in 4.1.2. We can propose a specification of a new shim layer protocol at some location above IP. Or, we can add a new specification to an existing upper layer protocol. The first two cases are probably the "cleanest" architecturally, but they are long term efforts that will take time to support in terms of design and implementation. It also may be argued that yet another shim layer will make the IP stack too large. The third case, placing a marking in an application layer packet, has the potential to be more appealing depending on where the augmentation is targeted.

An approach in realizing this third case involves an augmentation to RTP so that it can carry a marking that distinguishes emergency-related traffic from that which is not. Specifically, one would define a new extention that contains a "classifier" field indicating the condition associated with the packet (e.g., authorized-emergency, emergency, normal) [29].

An issue in defining a new extension to RTP is that its presence may adversely affect header compression for those implementations that are not expecting added optional octets in RTP packets. In addition, the issue of security and authentication of such a marking remains an important issue and is subject to the constraints discussed below in section 4.4, and in more detail in [27].

4.1.4 MEGACO/H.248

The Media Gateway Control protocol (MEGACO) [23] defines the interaction between a media gateway and a media gateway controller. [23] is viewed as common text with ITU-T Recommendation H.248 and is a result of applying the changes of RFC 2886 (Megaco Errata) to the text of RFC 2885 (Megaco Protocol version 0.8).

In [23], 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 MEGAGO priority. Any values set should be a function of any SLAs that have been established regarding the handling of emergency traffic. In addition, given that priority values denote precedence

(according to the Megaco protocol), then by default the IEPS flows should probably receive the same priority as other non-emergency calls. This approach follows the objective of not relying on preemption as the default treatment of emergency-related.

4.2 Policy

One of the objectives listed in section 3 above is to treat IEPS-signaling, and related data traffic, as non-preemptive in nature. Further, that 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 function to be supported; in our case, a default policy for non-preemptive call establishment of IEPS-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 [17]. There is no requirement as to how policy is accomplished. Instead, if a domain follows actions outside of the default non-preemptive action of IEPS-related communication, then we stipulate a functional requirement that some type of policy mechanism is in place to satisfy the local policies of an SLA established for IEPS 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 IEPS based traffic, it can be expected that some form of traffic engineering is applied to the operation of the network. We make no requirements as to which type of traffic engineering mechanism is used, but that such a system exists and can distinguish and support IEPS signaling and data traffic.

A potentially complimentary work in progress can be found in [9],

which articulates a framework for Voice over MPLS. We cite the draft only as a point of reference, with the idea that it may be augmented to reflect labeled path(s) dedicated to different values in the SIP priority field -- such as those pertaining to emergencies. But of more significance, [9] presents a specific framework for traffic engineering support of toll quality (i.e., a particular grade of service) IP telephony.

Note: As a point of reference, existing SLAs established by the NCS for GETS service tend to focus on a maximum allocation of 1% of calls allowed to be established through a given LEC using HPC. Once this limit is reached, all other GETS calls experience the same probably of call completion as the general public. It is expected, and encouraged, that IEPS related SLAs will have a limit with respect to the amount of traffic distinguished as being emergency related, and initiated by an authorized user.

4.4 Security

As IEPS support moves from intra-domain PSTN and IP networks to diffuse inter-domain pure IP, authenticated service becomes more complex to provide. Where an IEPS call is carried from PSTN to PSTN via one carrier's backbone IP network, very little IP-specific security support is required. The user authenticates herself as usual to the network using a PIN. The gateway from her PSTN connection into the backbone IP network must be able to signal that the flow has IEPS priority. Conversely, the gateway back into the PSTN must similarly signal the call's higher priority. A secure link between the gateways may be set up using IPSec or SIP security functionality. If the endpoint is an IP device on the carrier's network, the link may be set up securely from the ingress gateway to the end device.

As flows traverse more than one IP network, domains whose peering agreements include IEPS support must have means to securely signal a given flow's IEPS status. They may choose to use physical link security and/or IPSec authentication, combined with traffic conditioning measures to limit the amount of IEPS traffic that may pass between the two domains. The inter-domain agreement may require the originating network to take responsibility for ensuring only authorized traffic is marked with IEPS priority; the downstream domain may still perform redundant conditioning to prevent the propagation of theft and denial of service attacks. Security may be provided between ingress and egress gateways or IP endpoints using IPSec or SIP security functions.

When a call originates from an IP device, the ingress network may

authorize IEPS traffic over that link as part of its user authentication procedures without necessarily communicating with a central IEPS authentication center as happens with POTS-originated calls. These authentication procedures may occur at the link or network layers, but are entirely at the discretion of the ingress network. That network must decide how often it should update its list of authorized IEPS users based on the bounds it is prepared to accept on traffic from recently-revoked users.

5. Key Scenarios

There are various scenarios in which IP telephony can be realized, each of which can infer 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 & 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.

There are two scenarios that we use as a model for determining our objectives and subsequent functional requirements. These are:

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 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 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.

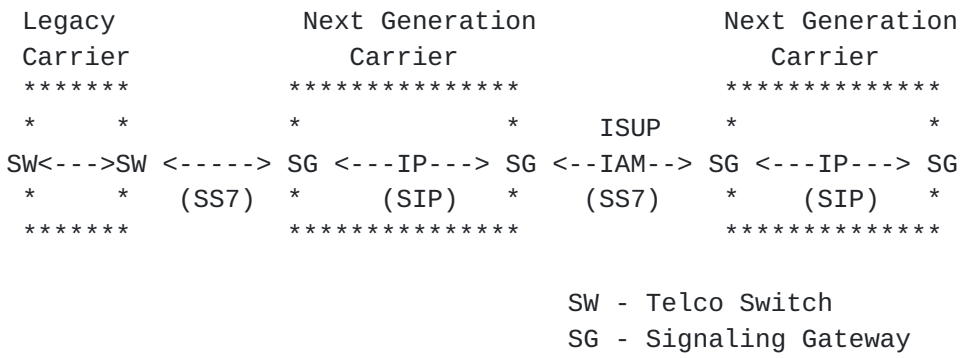

Figure 1

The significant aspect of this scenario is that all the resources of each IP "island" fall within a given administrative authority. Hence, there is no problem of retaining toll quality Grade of Service as the voice traffic (data and signaling) exits the IP network because of the existing SS7 provisioned service between carriers. Thus, the need for support of mechanisms like diff-serv, and an expansion of the defined set of Per-Hop Behaviors is reduced (if not eliminated) 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.

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.

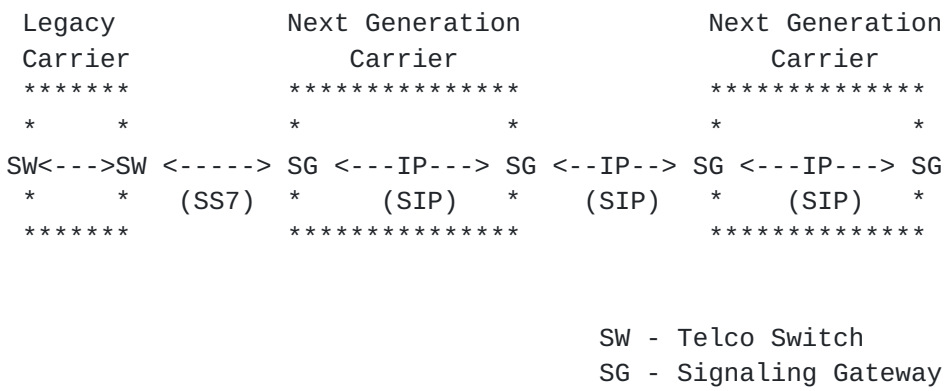

Figure 2

Given multiple IP domains, and the presumption that SLAs relating to IEPS 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 IEPS-type traffic is indeed valid for the given source.

8. Security Considerations

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

9. References

1 Bradner, S., "The Internet Standards Process -- Revision 3", BCP 9, RFC 2026, October 1996.

2 Braden, R., et. al., "Integrated Services in the Internet Architecture: An Overview", Informational, RFC 1633, June 1994.

3 Braden, R., et. al., "Resource Reservation Protocol (RSVP) _ Version 1, Functional Specification", Proposed Standard, RFC 2205, Sept. 1997.

4 Shenker, S., et. al., "Specification of Guaranteed Quality of Service", Proposed Standard, RFC 2212, Sept 1997.

5 Wroclawski, J., "Specification for Controlled-Load Network Service Element", Proposed Standard, RFC 2211, Sept 1997.

6 Gai, S., et. al., "RSVP Proxy", Internet Draft, Work in

Progress, July 2000.

7 Wang, L, et. al., "RSVP Refresh Overhead Reduction by State Compression", Internet Draft, Work In Progress, March 2000.

8 Blake, S., et. al., "An Architecture for Differentiated Service", Proposed Standard, RFC 2475, Dec. 1998.

9 Kankkunen, A., et. al., "VoIP over MPLS Framework", Internet Draft, Work In Progress, July 2000.

10 Sharma, V., et. al., "Framework for MPLS-based Recovery", Internet Draft, Work In Progress, September 2000.

11 Postel, J., "Simple Mail Transfer Protocol", Standard, RFC 821, August 1982.

12 Handley, M., et. al., "SIP: Session Initiation Protocol", Proposed Standard, RFC 2543, March 1999.

13 ANSI, "Signaling System No. 7(SS7) _ High Probability of Completion (HPC) Network Capability_, ANSI T1.631, 1993.

14 Reliable Audio Tool (RAT): http://www-mice.cs.ucl.ac.uk/multimedia/software/rat

15 Polk, J., "SIP Extension for MLPP", Internet Draft, Work In Progress, March, 2001.

16 Nichols, K., et. al.,"Definition of the Differentiated Services Field (DS Field) in the Ipv4 and Ipv6 Headers", Proposed Standard, RFC 2474, December 1998.

17 Durham, D., "The COPS (Common Open Policy Service) Protocol", Proposed Standard, RFC 2748, Jan 2000.

18 ITU, "International Emergency Preparedness Scheme", ITU Recommendation, E.106, March 2000.

19 Rosenburg, J., Schulzrinne, H., "A Framework for Telephony Routing Over IP", Informational, RFC 2871, June 2000

20 Heinanen. et. al, "Assured Forwarding PHB Group", Proposed Standard, RFC 2597, June 1999

21 Farrel, A, et. al, "Fault Tolerance for LDP and CR-LDP", Internet Draft, Work In Progress, February 2001.

22 Rosenburg, J, et. al, "Telephony Routing over IP (TRIP)", Internet Draft, Work In Progress, April 2001.

23 Cuervo, F., et. al, "Megaco Protocol Version 1.0", Standards Track, RFC 3015, November 2000

24 Perkins, C., et al., "RTP Payload for Redundant Audio Data", Standards Track, RFC 2198, September, 1997

25 Rosenburg, J., Schulzrinne, H., "An RTP Payload Format for Generic Forward Error Correction", Standards Track, RFC 2733, December, 1999.

26 ANSI, "Signaling System No. 7, ISDN User Part", ANSI T1.113-2000, 2000.

27 Brown, I., "Securing IEPS over IP", White Paper, http://iepscheme.net/docs/secure_IEPS.doc

28 Folts, H., "Description of an International Emergency Preference Scheme (IEPS) ITU-T Recommendation E.106 (Formerly CCITT Recomendation), Internet Draft, Work In Progress, February 2001

29 Carlberg, K., "The Classifier Extension Header for RTP", Internet Draft, Work In Progress, October 2001.

10. Acknowledgments

The authors would like to acknowledge the helpful comments, opinions, and clarifications of Stu Goldman and James Polk, as well as those comments received from the IEPS mailing list.

11. Author's Addresses

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

Ian Brown
University College London
Department of Computer Science

Gower Street
London, WC1E 6BT
United Kingdom

Full Copyright Statement