ecrit
Internet-Draft
Intended status: Standards Track
Expires: September 2, 2007

B. Rosen
NeuStar
H. Schulzrinne
Columbia U.
J. Polk
Cisco Systems
A. Newton
SunRocket

March 01, 2007

Framework for Emergency Calling in Internet Multimedia
draft-ietf-ecrit-framework-01

Status of this Memo

Copyright Notice

Abstract

Summoning emergency help by the public is a core feature of telephone networks. This document describes a framework of how various IETF protocols and mechanisms are combined to place emergency calls. This

includes how these calls are routed to the correct Public Safety Answering Point (PSAP) based on the physical location of the caller, while providing the call taker the necessary information to dispatch a first responder to that location. This document explains how location mapping, call identification and end system behavior are combined to allow multimedia emergency calls. It describes at a high level how the pieces (recognizing a call as an emergency call, marking it as such, determining the location of the caller, routing the call based on location) go together, and references the Internet standards that define the details of these mechanisms.

Table of Contents

1. Terminology

As a framework document, we do not define any new protocols or articulate new behaviors. Thus we do not use RFC2119 [RFC2119] notation. In this document, we reuse terms, and their definition, from [I-D.ietf-ecrit-requirements]. In addition, the following terms are used:

(Emergency) call taker: see [I-D.ietf-ecrit-requirements]

ESRP (emergency service routing proxy): see [I-D.ietf-ecrit-requirements]

Access Network: The wide area network that supplies IP packet service to an endpoint. In a residential or small business environment, this might be a DSL or cable modem or WiMax service. In a large enterprise environment, this would be the enterprise network. In a mobile environment, this might be a mobile (cellular) data network or a WiFi network.

Location Configuration: The process by which an endpoint learns its physical location.

Location Conveyance: The process of sending location to another element.

Location Determination: The mechanism used to resolve where an endpoint is physically. For example, the endpoint may have a GPS receiver.

Location Information Server An element that stores location information for retrieval by an authorized entity

Location Validation: see [I-D.ietf-ecrit-requirements]

Mapping: see [I-D.ietf-ecrit-requirements]

NENA (National Emergency Number Association: A North American organization of public safety focused individuals defining

emergency calling specifications and procedures.

PSAP (public safety answering point): see [I-D.ietf-ecrit-requirements]

SIP B2BUA see [RFC3261]

SIP proxy: see [RFC3261].

SIP Server see [RFC3261]

SIP UA (user agent): see [RFC3261].

Stationary device (user): An immobile user agent that is connected to the network at a fixed, long-term-stable geographic location. Examples include a home PC or a payphone.

Nomadic device (user): User agent that is connected to the network temporarily, for relatively short durations, but does not move significantly during the lifetime of a network connection or during the emergency call. Examples include a laptop using an 802.11 hotspot or a desk IP phone that is moved from one cubicle to another.

Mobile device (user): User agent that changes geographic location and possibly its network attachment point during an emergency call.

2. Introduction

Summoning police, the fire department or an ambulance in emergencies is one of the fundamental and most-valued functions of the telephone. As telephone functionality moves from circuit-switched telephony to Internet telephony, its users rightfully expect that this core functionality will continue to work at least as well as it has for the older technology. New devices and services are being made available which could be used to make a request for help which are not traditional telephones, and users are increasingly expecting them to be used to place emergency calls. However, many of the technical advantages of Internet multimedia require re-thinking of the traditional emergency calling architecture. This challenge also offers an opportunity to improve the operation of emergency calling technology, while potentially lowering its cost and complexity.

It is beyond the scope of this document to enumerate and discuss all

the differences between traditional (PSTN) and Internet telephony, but the core differences can be summarized as:

- the separation/interleaving of signaling and media data packets;
- the interleaving over the same infrastructure of what is an emergency call with non-emergency traffic, whether that other traffic is another type of call or other Internet-based traffic such as email or web browsing
- the emergence of application-independent carriers;
- the plethora of different media that can be accommodated;
- potential mobility of all end systems, including endpoints nominally thought of as fixed systems and not just those using radio access technology. For example, a wired phone connected to a router using a mobile data network such as EV-DO as an uplink;

This document focuses on how devices using the Internet can place emergency calls and how PSAPs can natively handle Internet multimedia emergency calls, rather than describing how circuit-switched PSAPs can handle VoIP calls. In many cases, PSAPs making the transition from circuit-switched interfaces to packet-switched interfaces may be able to use some of the mechanisms described here, in combination with gateways that translate packet-switched calls into legacy interfaces, e.g., to continue to be able to use existing call taker equipment.

We distinguish an individual request for help, usually accomplished by dialing a short digit sequence like 9-1-1 or 1-1-2 from a call

placed by specially designated persons who have authority to claim priority on available Internet communications facilities. This document only discusses the former - a request for help by an ordinary user answered at an emergency call center (i.e. a PSAP).

Existing emergency call systems are organized locally/nationally; there are currently no international standards. However, the Internet does not respect national boundaries, and thus international standards for equipment and software are required. To further complicate matters, VoIP endpoints can be connected through tunneling mechanisms such as virtual private networks (VPNs). This significantly complicates emergency calling, because the location of the caller and the first element that routes emergency calls can be on different continents, with different conventions and processes for handling of emergency calls.

The IETF has historically refused to create national variants of its standards. Thus, this document attempts to take into account best practices that have evolved for circuit switched PSAPs, but makes no assumptions on particular operating practices currently in use, numbering schemes or organizational structures.

This document discusses the use of the Session Initiation Protocol (SIP) [RFC3261] by PSAPs and calling parties. While other inter-domain call signaling protocols may be used for emergency calling, SIP is ubiquitous and possesses, through its related specifications, more of the needed features for the proper support of this use case. Only protocols such as H.323, XMPP/Jingle, ISUP and SIP are suitable for inter-domain communications, ruling out MG/MGC protocols such as MGCP or H.248/Megaco. The latter protocols can naturally be used by the enterprise or carrier placing the call, but any such call would reach the PSAP through a media gateway controller, similar to how interdomain VoIP calls would be placed. Other signaling protocols may also use protocol translation to communicate with a SIP-enabled PSAP.

Existing emergency services rely exclusively on voice and conventional text telephony (known as TTY in the United States) media streams. However, more choices of media offer additional ways to communicate and evaluate the situation as well as to assist callers and call takers to handle emergency calls. For example, instant messaging and video could improve the ability to communicate and evaluate the situation and to provide appropriate instruction prior to arrival of emergency crews. Thus, the architecture described here supports the creation of sessions of any media type, negotiated between the caller and PSAP using existing SIP protocol mechanisms [RFC3264]. To ensure that at least one common means of communications, this document recommends certain minimal capabilities

in [I-D.ietf-ecrit-phonebcp] that call taker user agents and PSAP-operated proxies should possess.

This document does not prescribe the detailed network architecture for a PSAP or collection of PSAPs. For example, it does not describe where PSAPs may place firewalls or how many SIP proxies they should use.

This document does not introduce any new SIP header fields, request methods, status codes, message bodies, or event packages. User agents unaware of the recommendations in this draft can place emergency calls, but may not be able to provide the same elevated user interface functionality. The document suggests behavior for proxy servers, in particular outbound proxy servers.

3. Overview of How Emergency Calls are Placed

We distinguish (Section 4) an emergency call from any other call by a unique Service URN[I-D.ietf-ecrit-service-urn], which is placed in the initial call set-up signaling when a home or visited emergancy dialstring is detected. We route emergency calls based on the location ((Section 5)) of the caller. To get this location we either include a form of measuring (e.g. GPS) ((Section 5.3.3)) device location in the endpoint, or the endpoint is configured ((Section 5.5)) with its location from the access network's Location Information Server (LIS) The location is conveyed ((Section 5.6)) in the SIP signaling with the call. We route((Section 6)) the call based on location using the LoST protocol ([I-D.ietf-ecrit-lost]) which maps a location to a set of PSAP URIs. Each URI resolves to a PSAP or an Emergency Services Routing Proxy which serves a group of PSAPs. The call arrives at the PSAP with the location included in the INVITE request.

Configuration Servers

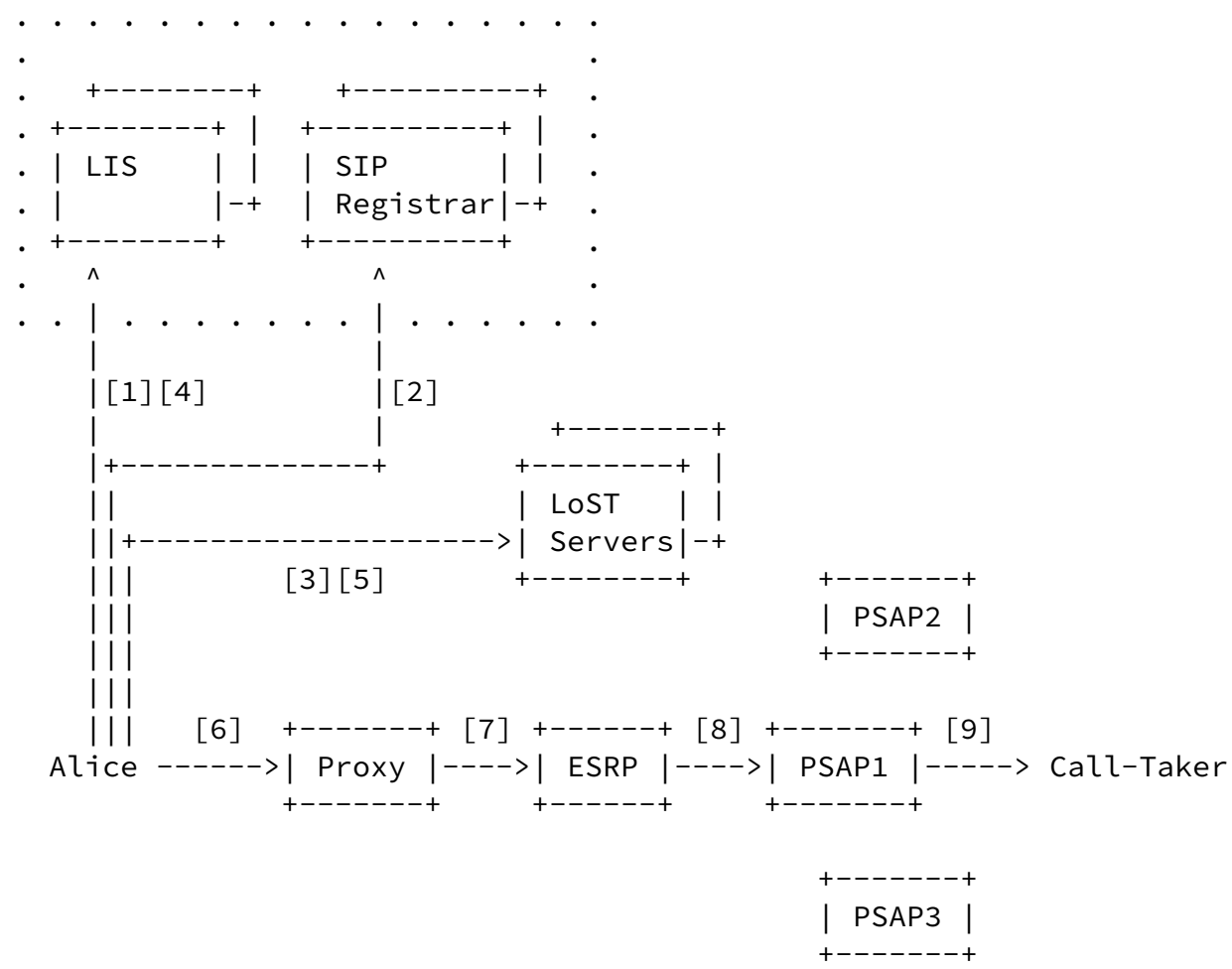


Figure 1: Generic ECRIT Component Topology

Figure 2 shows a generic emergency call establishment. This includes the following:

- Alice - who will make the emergency call.
- Configuration Servers - Servers providing Alice's UA its IP address and other configuration information, perhaps including Location by-value or by-reference. In this flow, we use DHCP as an example location acquisition protocol. Configuration servers also may include a SIP Registrar server, for Alice's UA to register Alice's UA to register with. Most SIP UAs will register with a call server, so it will be a common scenario for UAs that make emergency calls to be registered with such a server in the originating calling network. All these configuration messages are labeled M1 through M3, but could easily require more messages than 4 to complete.
- ESRP - The Emergency Services Routing Proxy Server that is the incoming call proxy in the emergency services domain. The ESRP makes further routing decisions based on PSAP state and location

of the caller to choose the actual PSAP which handles the call. In some jurisdictions, that may involve another LoST dip

- LoST Server - Processes the LoST request for Location to PSAP-URI Mapping function, either for an initial request from a UA, or an in-call routing by the Proxy server in the originating network, or possibly by an ESRP.
- PSAP - Call center where emergency calls are destined for in times of emergencies.

Generally, Alice's UA either has location configured manually, has an integral location measurement mechanism, or it runs a location configuration protocol to obtain location from the access (broadband) network. For most devices, an LCP will be used, for example a DHCPREQUEST message or another location acquisition mechanism. Alice's UA then will most likely register with a SIP domain. This allows her to be contacted by other SIP entities. Next, her UA will perform an initial LoST Location-to-PSAP SIP(S)-URI query to learn a URI, for use if the Lost Query fails during an emergency call. The LoST query may contain the dialstring for emergency calls appropriate for the location provided.

Some time has hopefully passed since Alice's UA booted. In this example, she dials or initiates an emergency call. This may have been through her keypad with her locally known emergency dialstring. It is important that this dialstring be recognized by her UA wherever Alice is because she may be in enough distress she forgets what the traveled-to emergency dialstring is; as there are more than 60 around the world.

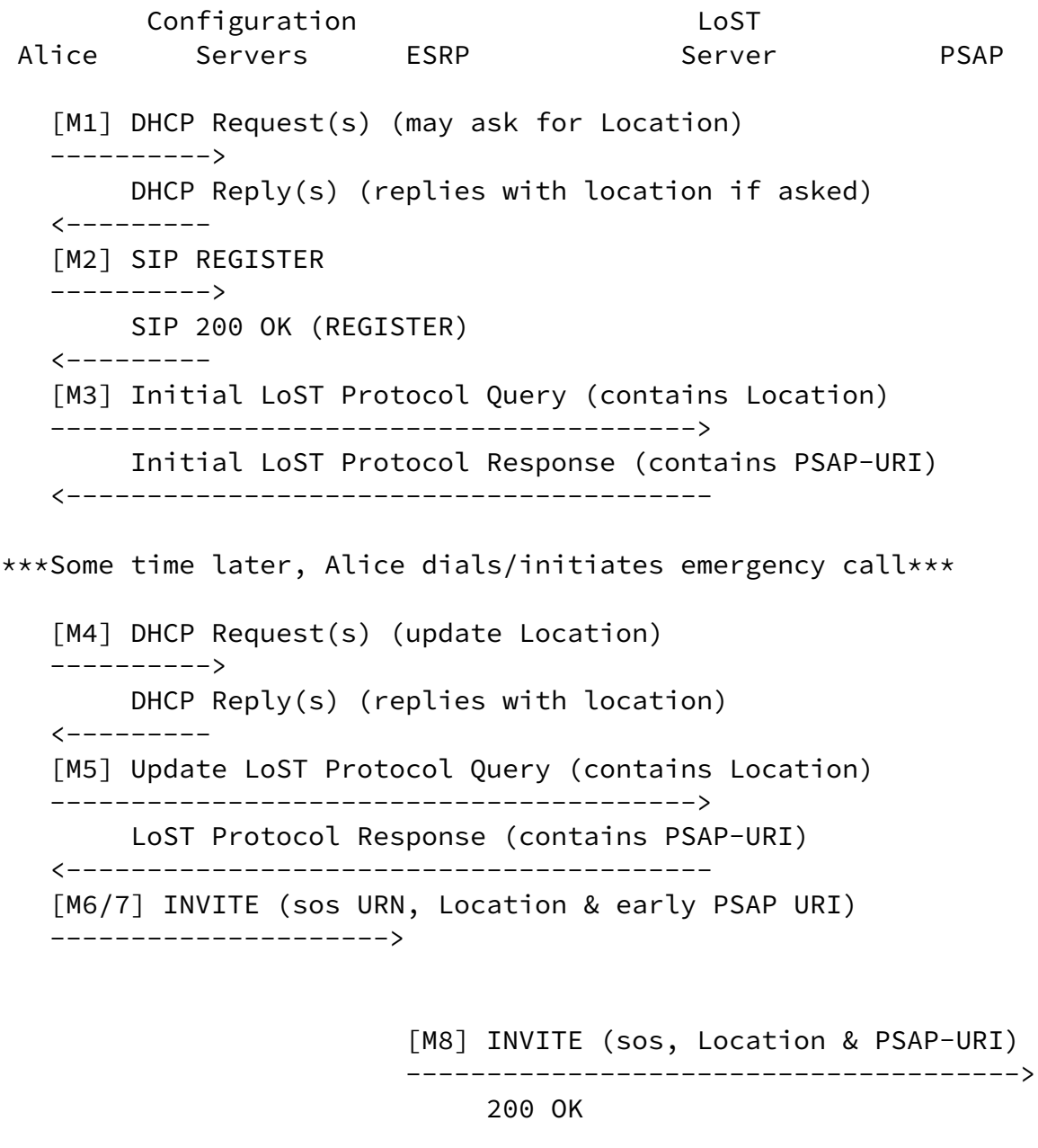
The UA recognizes the dialstring, which means this is an emergency call. The UA attempts to refresh its location, and with that location, the LoST mapping, to get the most accurate information to use for routing the call. If the location request or the LoST request fails (or takes too long) the UA uses it's cached values.

The UA creates an INVITE which includes the location. [I-D.ietf-sip-location-conveyance] defines a SIP Location header that either contain the location-by-reference URI, or a [RFC2396] "cid:" indicating where in the message body the location-by-value is.

The INVITE message routes to the ESRP, which is the first inbound proxy for the emergency services domain. This message, is then routed by the ESRP towards the most current PSAP for Alice's location, which uses PSAP state, location and other state information to choose this PSAP.

A proxy in the PSAP choses an available call taker and extends the call to its UA.

The 200 OK to the INVITE traverses the path in reverse, from call taker UA to PSAP proxy to ESRP to originating network proxy to Alice's UA. The ACK completes the call set-up and the emergency call is established, allowing the PSAP call-taker to talk to Alice about her emergency.



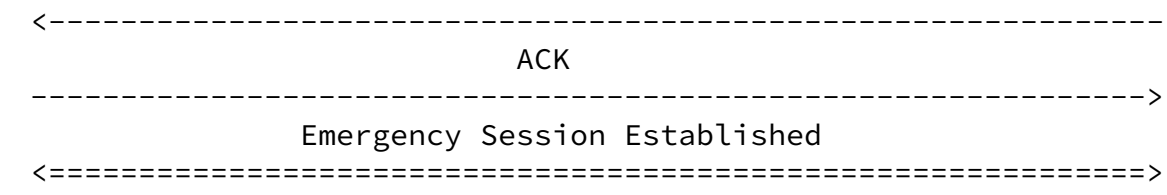


Figure 2: General Flow of an Emergency Call Establishment

This is a very rough example of the operation of an emergency call

establishment. There are no layer 3 routers in the message flow, and whatever security messages exist in the call are not shown either. Each of those aspects will be addressed individually, to keep each discussion in context of that subject, for clarity.

4. Identifying an Emergency Call

Using the PSTN, emergency help can often be summoned by dialing a nationally designated, widely known number, regardless of where the telephone was purchased. The appropriate number is determined by which infrastructure the telephone is connected to. However, this number differs between localities, even though it is often the same for a country or region, such as many countries in the European Union. In some countries, there is a single digit sequence that is used for all types of emergencies. In others, there are several sequences that are specific to the responder, e.g., one for police, another for fire. It is deemed impractical to change the dialed digits to summon help. For end systems, it is desirable to have a universal identifier, independent of location, to allow the automated inclusion of location information and to allow the device and other entities in the call path to perform appropriate processing within the signaling protocol in an emergency call set-up.

As part of the overall emergency calling architecture, we define common emergency call URIs which are defined in [I-D.ietf-ecrit-service-urn]. Users are not expected to "dial" an emergency URN. Rather, the current dialstring should be translated to the appropriate service URN. Such translation could ideally be performed in the endpoint, but could be performed in a signaling intermediary (proxy server). For devices that are mobile or nomadic, an issue arises of whether the home or visited dialing strings should

be used. Many users would prefer that their home dialing sequences work no matter where they are. Local laws and preferences of the emergency response professionals are such that the visited dialing sequences must always work. Having the home dialstring work is optional. The best answer seems to be for both to work.

The mechanism for obtaining the dialing sequences for a given location is provided by LoST [I-D.ietf-ecrit-lost]. Where the endpoint does not support the translation of dialstrings to telephone numbers, the dialing sequence would be represented as a dialstring [I-D.rosen-iptel-dialstring] and the outgoing proxy would recognize the dialstring and translate to the service URN. It should be noted that the endpoint would not normally supply location unless it understood the call to be an emergency call. To determine the local dialstring, the proxy needs the location of the endpoint. This may be difficult in situations where the user can roam or be nomadic.

Endpoint recognition of emergency dialstrings is therefore preferred.

5. Location and Its Role in an Emergency Call

5.1. Introduction

Caller location plays a central role in routing emergency calls. For practical reasons, each PSAP generally handles only calls for a certain geographic area (overload arrangements between PSAPs to handle each others calls notwithstanding). Other calls that reach it by accident must be manually re-routed (transferred) to the appropriate PSAP, increasing call handling delay and the chance for errors. The area covered by each PSAP differs by jurisdiction, where some countries have only a small number of PSAPs, while others decentralize PSAP responsibilities to the level of counties or municipalities.

In most cases, PSAPs cover at least a city or town, but there are some areas where PSAP coverage areas follow old telephone rate center boundaries and may straddle more than one city. Irregular boundaries are common, often for historical reasons. Routing must be done on PSAP service boundaries, not "closest" or "best fit" algorithms.

5.2. Types of Location Information

There are four primary types of location information: civic, postal, geospatial, and cellular cell tower and sector.

Civic: Civic location information describes the location of a person or object by a street address that corresponds to a building or other structure. (This is sometimes also called "civil" location information.) Civic location may include more finer grained location information such as floor, room, cubicle. Civic information comes in two forms:

Jurisdictional - This refers to a civic location using actual political subdivisions, especially for the community name.

Postal - This refers to a civic location used to mail a letter to. The name of the post office sometimes does not correspond to the actual community name and a postal addrress may contain post office boxes or street addresses that do not correspond to an actual building. Postal addresses are generally unsuitable for emergency call routing, but may be the only address available.

Geospatial: Geospatial addresses contain longitude, latitude and altitude information based on an understood datum (starting point) and earth shape model. While there have been many datums developed over time, most modern systems are using or moving towards WGS84.

Cell tower/sector: Cell tower and sectors identify the cell tower and the antenna sector that the mobile device is currently using. Traditionally, the tower location is expressed as a point, and routing decisions are made on that point. Cell/sector information could also be transmitted as an irregularly shaped polygon of geospatial coordinates reflecting the likely geospatial location of the mobile device.

In IETF protocols, civic and geo forms are both supported. The civic forms include both the postal and jurisdictional fields. The cell tower/sector can be represented as a point.

5.3. Location Determination

Location information can be entered by the user or installer of a device ("manual configuration"), can be measured by the end system, can be delivered to the end system by some protocol or can be measured by a third party and inserted into the call signaling. We discuss these in detail below.

In some cases, an entity may have multiple sources of location information, possibly partially contradictory. This is particularly likely if the location information is determined both by the end system and a third party. Handling multiple locations is discussed in [I-D.ietf-geopriv-pdif-lo-profile]. Conflicting location information is particularly harmful if it points to multiple distinct PSAPs. Guidelines for dealing with multiple locations is also given in [I-D.ietf-ecrit-lost].

All location objects MUST be delivered to the PSAP. To facilitate such policy decisions, location information should contain information about the source of data, such as GPS, manually entered or based on access network topology. In addition, the generator of the location information should be included. The ability of the UA to understand how it learned its location, and include this information element in the location object that is sent to the PSAP, provides the call-taker with many pieces of information to make decisions upon, and guidance for what to ask the caller and what to tell the responders.

The call should indicate which location information has been used for routing, so that the same location information is used for all call routing decisions. Otherwise, two proxies might pick different

location information from the call request, resulting in different routing decisions for different transactions. The location conveyance mechanism [I-D.ietf-ecrit-lost] contains a parameter which can be used for this purpose

End systems and network elements can derive location information from a variety of sources. It is not the goal of this document to exhaustively enumerate them, but we provide a few common examples in the sections below.

5.3.1. User-Entered Location Information

Location information can be maintained by the end user or the installer of an endpoint in the endpoint itself, or in a database.

Location information added by end users is almost always inferior to measured or wire database information, as users may mistype civic location information, may not know the meaning of geospatial coordinates or may use address information that does not correspond to a recognized civic address. A user-entered location can fail to be changed when the location of a device changes during or after movement. For example, a user could move their residence to another dwelling, not update their device/equipment with this new location, and place an emergency call with old location information.

All that said, there are always a small number of cases where the mechanisms used by the access network to determine location fail to accurately reflect the actual location of the endpoint. For example, the user may deploy his own WAN behind an access network, effectively remoting an endpoint some distance from the access network's notion of its location. There must be some mechanism provided to provision a location for an endpoint by the user or by the access network on behalf of a user. The use of the mechanism introduces the possibility of users falsely declaring themselves to be somewhere they are not. As an aside, normally, if an emergency caller insists he is at a location different from what any automatic location determination system reports he is, responders will always be sent to the user's self-declared location. However this is a matter of local policy and is outside the scope of this document.

5.3.2. Access Network "Wire Database" Location Information

Location information can be maintained by the access network, relating some form of identifier for the end subscriber or device to a location database ("wire database"). In enterprise LANs, wiremap databases map Ethernet switch ports to building layouts at known locations. In DSL installations, the local telephone carrier maintains a mapping of wire-pairs to subscriber addresses.

Even for IEEE 802.11 wireless access points, wire databases may provide sufficient location resolution; the location of the access point may be sufficient location information for each of the clients served by that access point. This may be the connectivity type for

both residential users of DSL and Cable Modem installations, as well as the only infrastructure at a WiFi hotspot, such as a coffee shop. Each of these cases will have a known civic address of the dwelling/ business, likely providing sufficient location resolution.

Wire databases to the home are likely to be the most promising solution for residential users where a service provider knows the customer's service address. The service provider can then perform address verification, similar to the current system in some jurisdictions.

5.3.3. End-System Measured Location Information

Global Positioning System (GPS) sensors may be embedded directly in the end device. GPS produces relatively high precision location fixes in open-sky conditions, but the technology still faces several challenges in terms of performance (time-to-fix and time-to-first-fix), as well as obtaining successful location fixes within shielded structures, or underneath the ground (tunnels, basements, etc.). It also requires all devices to be equipped with the appropriate GPS capability. GPS technology is improving, and is increasingly successful in more difficult conditions such as dense urban canyons and inside commercial structures. It is currently accurate to tens of meters using some kind of "assist", which may be operated by the access network (A-GPS) or by a government (WAAS). Newer multi-frequency systems will improve accuracy without assist.

GPS equipped devices vary depending on which element initiates requests, which element actually determines final location, assist mechanisms, etc. Some common implementations include:

1. GPS S/A (standalone), device initiated
2. GPS S/A, network initiated
3. AGPS-device initiated, network determined
4. AGPS-device initiated, network augmented
5. AGPS-network initiated, network determined
6. AGPS-network initiated, network augmented

5.3.4. Third-party Measured Location Information

Wireless triangulation: Elements in the network infrastructure triangulate end systems based on signal strength, angle of arrival or time of arrival. Common mechanisms deployed include.

1. Time Difference Of Arrival - TDOA
2. Uplink Time Difference Of Arrival - U-TDOA
3. Angle of Arrival - AOA
4. RF-Fingerprinting
5. Advanced Forward Link Trilateration - AFLT
6. Enhanced Forward Link Trilateration - EFLT

Sometimes triangulation and measured mechanisms are combined, for example A-GPS with AFLT

Location beacons: A short range wireless beacon, e.g., using Bluetooth or infrared, announces its location to mobile devices in the vicinity.

5.4. Location and References to Location

Location information may be expressed as the actual civic or geo value but can be transmitted as by-value (wholly contained within the signaling message) or by-reference (a URI pointing to the value residing on a remote node waiting to be dereferenced). There are pros and cons to each form:

location-by-value:

pro- Value available to each device along the path immediately for further processing.

con- Size, especially if constrained to a UDP transport. Value fixed at the time the value is acquired from the access network. Value can be changed by endpoint, which may be considered untrustworthy for this critical usage.

location-by-reference

pro- Small size. Value can be fixed at time of dereference. Value cannot be changed by endpoint

con- URI resolution requires location source be available and accessible by dereferencer. Dereferencing takes time. Dereferencing may fail.

5.5. End System Location Configuration

Unless a user agent has access to provisioned or locally measured location information, it must obtain it from the access network. There are several Location Configuration Protocols that can be used for this purpose.

DHCP can deliver civic [RFC4676] or geospatial [RFC3825] information. User agents would need to support both formats. Note that a user agent can use DHCP, via the DHCP REQUEST or INFORM messages, even if it uses other means to acquire its IP address.

Insert reference to L7 acquisition protocol document> is another choice.
Link-Layer Discovery Protocol [LLDP]), with proposed extensions [LLDP-MED], may also be used to deliver location information.
SUPL OASIS <insert reference> is yet another choice.

Other LCPs may be devised by other standards bodies. Each LCP has limitations in the kinds of networks that can reasonably support it. For this reason, it is not possible to choose a single mandatory to deploy LCP. For endpoints with common network connections (such as an Ethernet jack or a WiFi connection), unless every network supported every protocol, or alternatively, every device supported every protocol, serious incompatibilities would ensue. [I-D.ietf-ecrit-lost] contains a (short) list of protocols such devices must support.

Where an access network can control the specification of EVERY endpoint that could make an emergency call that is directly connected to the network, or indirectly connected (for example, a device on a LAN behind a network attachment unit), it may specify any protocol it wishes for each endpoint. This is a very unusual case; nearly every access network can be used to support an Ethernet based LAN behind it

For example, existing mobile networks are being used to support routers and LANs behind a wireless data network WAN connection, with Ethernet connected phones connected to that. It is possible that the access network supports a protocol not on the phonebcp list, and every handset supported in that network could use that protocol for emergency calls. However, unless another element which the access network provider controls the specification of can acquire location using that protocol and then that element can support one of the phonebcp's list of protocols, the Ethernt connected phone won't be able to acquire location. In this case, if the access network provider supplies a router which includes a DHCP server, it can acquire location using the access network specific protocol, and then use the location information to supply it to its clients (e.g. the Ethernet connected phone) via DHCP.

For most networks, it will not be practical to control the specification of every device, or arrange interworking with network

specific LCPs. For this reason, most devices will need to support ALL of the LCPs in [I-D.ietf-ecrit-lost], and access networks will have to support at least one of these LCPs.

Location for non-mobile devices is normally expected to be acquired at network attachment time and retained by the device. It should be refreshed when the cached value becomes invalid (for example, if DHCP is the acquisition protocol, refresh of location may occur when the

IP address lease is renewed). At the time of an emergency call, the location should be refreshed, with the retained location used if the location acquisition does not immediately return a value. Mobile devices may determine location at network attachment time and periodically thereafter as a backup in case location determination at the time of call does not work. Mobile device location may be refreshed when a TTL expires, the device moves beyond some boundaries (as provided by [I-D.ietf-ecrit-lost]), etc. Normally, mobile devices will acquire its location at call time for use in an emergency call routing, but see Section 5.7

5.6. Conveyance of Location

When an emergency call is placed, the endpoint (normally) puts location information in the signaling with the call. We refer to that as "conveyance" to distinguish it from "configuration". Configuration gets location from access network to endpoint, conveyance sends location from endpoint to elements that route the call based on that location object and the PSAP. Using SIP, the location information is conveyed following the procedures in [I-D.ietf-sip-location-conveyance]. The form of the location information obtained by the acquisition protocol may not be the same as the conveyance protocol uses (PIDF-LO [RFC4119]). Mapping by the endpoint may be required. Calling networks which support devices which do not support location may have to add location to emergency calls. Some calling networks have relationships with the access network that may allow it to accurately determine location of the endpoint, although NATs and other middleboxes usually make it impossible to determine a reference identifier the access network could use to determine the location.

For emergency call purposes, conversion of location information from civic to geo or vice versa prior to conveyance is not desirable. The

location should be sent in the form it was determined. The PSAP may convert, if it needs to, and if conversion resulted from an earlier conversion, unacceptable errors may be introduced.

5.7. Location Updates

Location information may not be available at call setup time for mobile devices. For example, if a GPS-enabled cell phone is turned on and then immediately places an emergency call, it can take significant additional time before the cell phone acquires a GPS fix and its location. Thus, while it is desirous to base emergency routing on precise caller location information, it is not possible in all circumstances to do so. In some cases, the initial call setup will proceed based on, for example, cell and sector information and then add location information during the call, rather than delaying

the initial call setup by an unacceptable amount of time.

In addition, the location of a mobile caller, e.g., in a vehicle or aircraft, can change significantly during the emergency call. The PSAP must be able to get updated location information while it is processing the call.

Location updates where the location is conveyed by value may be conveyed either in a re-INVITE or UPDATE [RFC3311] request message (where UPDATE is preferred) or the PSAP may subscribe to the location information of the caller, using SIP presence mechanisms (RFC 3265 [RFC3265] RFC 3856 [RFC3856]). Authorization for subscriptions is for future study. When location is conveyed by reference, additional dereference operations yield updated location.

5.8. Location Validation

Location must be validated prior to a device placing an actual emergency call. Validation in this context means both that there is a mapping from the address to a PSAP and that the PSAP understands how to direct responders to the location. This is not as easy as it sounds. There are, for example, many cases of two names for the same street, or two streets with the same name in a city. In some countries, the current system provides validation. For example, in the United States, the Master Street Address Guide (MSAG) records all valid street addresses and is used to ensure that the service

addresses in phone billing records correspond to valid emergency service street addresses. Validation is normally a concern for civic addresses, although there could be a concern that a given geo is within at least one PSAP service boundary; that is, a "valid" geo is one for which there is a mapping.

The LoST resolver[I-D.ietf-ecrit-lost] includes a validation function. Validation should ideally be performed when a location is entered into a Location Information Server (which is normally a provisioning mechanism in the access carrier's operation and support system). It should be confirmed periodically, because the mapping database undergoes slow change; new streets are added or removed, community names change, postal codes change, etc. Endpoints may wish to validate locations they receive from the access network, and will need to validate manually entered locations. Proxies which insert location may wish to validate locations they recieve from a LIS. Test functions (Section 13) should also re-validate.

5.9. Default Location

Occasionally, a failure may occur where the access network cannot determine the actual location of the caller. In these cases, it must

supply a default location. The default location should be as accurate as the network can determine. For example, in a cable network, a default location for each Cable Modem Termination System (CMTS), with a representative location for all cable modems served by that CMTS could be provided if the network is unable to resolve the subscriber to any unit less than the CMTS. Default locations must be marked as such (how?) so that the PSAP knows that the location is not accurate.

6. Routing the Call to the PSAP

Emergency calls are routed based on one or more of the following criteria expressed in the call setup request (INVITE):

Location: Since each PSAP serves a limited geographic region and transferring existing calls delays the emergency response, calls need to be routed to the most appropriate PSAP. In this architecture, emergency call setup requests contain location

information, expressed in civic or geospatial coordinates, that allows such routing. If there is no or imprecise (e.g., cell tower and sector) information at call setup time, an on-going emergency call may also be transferred to another PSAP based on location information that becomes available in mid-call.

Type of emergency service: In some jurisdictions, emergency calls for fire, police, ambulance or mountain rescue are directed to just those emergency-specific PSAPs. We support this mechanism by optionally labeling calls with a service identifier [I-D.ietf-ecrit-service-urn].

Media capabilities of caller: In some cases, emergency call centers for specific caller media preferences, such as typed text or video, are separate from voice systems. Also, even if media capability does not affect the selection of the PSAP, there may be call takers within the PSAP that are specifically trained, e.g., in interactive text or sign language communications. Again, we use the callee capabilities [RFC3840] mechanism to label and route such calls.

Routing for calls by location and by service is the primary function LoST [I-D.ietf-ecrit-lost] provides. LoST accepts a query with location (by-value) in either civic or geo form, plus a service identifier, and returns an xml data structure containing a URI (or set of URIs) to route the call to. Normal SIP [RFC3261] routing functions are used to resolve the URI to a next hop destination.

The endpoint can complete the LoST mapping from its location at boot time, and periodically thereafter. It should attempt to obtain a "fresh" location, and from that a current mapping when it places an

emergency call, and if accessing either its location acquisition function or mapping function fails, it should use this cached value. The call would follow its normal outbound call processing. Networks that support devices that do not implement LoST mapping themseleves would have the outbound proxy do the mapping. The proxy must have the location of the endpoint, which is often difficult for the calling network to accurately determine. The endpoint may have its location, but would not normally include it on the call signaling. There is no mechanism provided in [I-D.ietf-sip-location-conveyance] to allow a proxy to require the endpoint supply location, because that would open the endpoint to an attack by any proxy on the path to get it to reveal location. The Proxy CAN redirect a call to the

service URN which, if the device recognized the significance, would include location in the redirected call. All networks should detect emergency calls and supply default location and/or routing if it is not already performed.

With the URI obtained from mapping, whether by the endpoint or the proxy, the proxy routes the call. Normal SIP[RFC3261] mechanisms are used to route calls to the URI obtained from the LoST dip.

Often, the SIP routing of an emergency call will first route to an incoming call proxy in the domain operated by the emergency service. That proxy is called an "Emergency Services Routing Proxy" (ESRP). The ESRP, which is a normal SIP proxy server, may use a variety of PSAP state information, the location of the caller, and other criteria to onward route the call to the PSAP.

7. Signaling of Emergency Calls

As discussed above, location is carried in all emergency calls in the call signaling. Since emergency calls carry privacy-sensitive information, they are subject to the requirements for geospatial protocols [RFC3693]. In particular, signaling information should be carried in TLS, i.e., in 'sips' mode. While requiring TLS is actually the way the standards are written, it is unacceptable to have an emergency call fail to complete because a TLS connection was not created, for any reason. In many cases, persistent TLS connections can be maintained between elements to minimize the time needed to establish them.

The use of SIP Identity [RFC4474] to protect the headers of the message could improve end-to-end integrity of the information.

Details of how location is carried in call signaling can be found in [I-D.ietf-sip-location-conveyance].

8. Caller Preferences

SIP Caller Preferences [RFC3841] may be used to signal how the PSAP should handle the call. For example, a language preference expressed in an Accept-Language header may used as a hint to cause the PSAP to

route the call to a call taker who speaks the requested language.

9. Including a Valid Call-Back Identifier

The call-taker must be able to reach the emergency caller if the original call is disconnected. In traditional emergency calls, wireline and wireless emergency calls include a callback identifier for this purpose. In SIP systems, the caller should include a Contact header field indicating its device URI, if available, or possibly a GRUU[I-D.ietf-sip-gruu] if calls need to be routed via a proxy. This identifier would be used to initiate call-backs immediately by the call-taker if, for example, the call is prematurely dropped.

In addition, a call-back identifier should be included either as the URI in the From header field [RFC3261] preferably verified by SIP Identity[RFC4474]. This identifier would be used to initiate a call-back at a later time and may reach the caller, not necessarily on the same device (and at the same location) as the original emergency call. Both the Contact and From specific requirements are detailed in [I-D.ietf-ecrit-phonebcp]

Finally, there may be two other call identifiers included in an emergency call. An identifier may be included which can be used to identify the caller, as opposed to the device or the subscriber of a specific calling service. This identifier may be used to retrieve information about the caller that is independent of calling service. For example, Alice may have home, office and mobile telephony services, but she is the same Alice in all of them. Information about Alice may be kept by an entity independent of any telephony service provider. The caller identity is a URI and is placed in a SIP Call-Info header [RFC3261] using the token "?" following the recommendations in [I-D.ietf-ecrit-phonebcp].

The communications service provider may also include an identifier that may be used to retrieve information specific to the call held by the service provider. This identifier, also a URI may be placed in the Call-Info header using the token "?" per [I-D.ietf-ecrit-phonebcp].

10. Mid-Call Services and Behavior

A PSAP may need to REFER[RFC3515] a call to a bridge for conferencing. The caller should also be prepared to have the call transferred (usually attended, but possibly blind) as per[I-D.ietf-sipping-service-examples].

While in a call, a number of of other call features, such as call waiting, must be disabled. This is also discussed in [I-D.ietf-ecrit-phonebcp].

11. Call Termination

It is undesirable for the caller to terminate an emergency call. Strategies for devices to handle caller attempts to terminate may be found in [I-D.ietf-ecrit-phonebcp]. PSAP call termination is accomplished with normal SIP call termination procedures.

12. Media

PSAPs should accept media streams on RTP [RFC3550]. Traditionally, voice has been the only media stream accepted by PSAPs. In some countries, text, in the form of BAUDOT codes or similar tone encoded signaling within a voiceband is accepted ("TTY") for persons who have hearing disabilities. With the Internet comes a wider array of potential media which a PSAP should accept. Using SIP signaling includes the capability to negotiate media. Normal SIP offer/answer [RFC3264] negotiations should be used to agree on the media streams to be used. PSAPs should accept real-time text [RFC4103]. All PSAPs should accept G.711 A law (and mu Law in North America) encoded voice as described in [RFC3551]. Newer text forms are rapidly appearing, with Instant Messaging now very common, PSAPs should accept IM with at least [RFC3428] as well as [RFC3920].

13. Testing

Since the emergency calling architecture consists of a number of pieces operated by independent entities, it is important to be able to test whether an emergency call is likely to succeed without actually occupying the human resources at a PSAP. Both signaling and media paths need to be tested since NATs and firewalls may allow the session setup request to reach the PSAP, while preventing the exchange of media.

[I-D.ietf-ecrit-phonebcp] includes a description of an automated test

procedure that validates routing, signaling and media path continuity. This test would be used at boot time, and whenever the device location changes enough that a new PSAP mapping is returned from LoST. A manual operation for the test should also be possible.

14. Example Call Flows

TBD

15. Alternatives Considered

This is a non-normative appendix. During discussions of emergency calling, a number of suggestions are commonly made. Below, we discuss some of the reasons why these alternatives do not satisfy the requirements of emergency calling.

15.1. tel URIs

Instead of providing URIs to call routing proxies or end systems, it has been suggested that end systems be configured with a "tel" URI [RFC3966]. Such a "tel" URI would have to be routed to a geographically appropriate telephony gateway, as it is unlikely that every building, enterprise or residence will have its own gateway. VoIP devices can be used in networks that are completely unaware of VoIP services, with VoIP service providers that are physically far removed from the caller's network location. Thus, the use of a tel URI simply moves the problem to the outbound proxy, which has to use the caller's location to determine the appropriate telephony gateway.

In addition, emergency telephone numbers are far from universal, with some such numbers used for non-emergency purposes elsewhere. Thus, an outbound proxy would have to ascertain the location of the caller to guess whether the "tel" URI identifies an emergency call or some other number.

Thus, "tel" URIs are not likely to be appropriate or sufficient for identifying emergency calls and do not, by themselves, solve the call routing problem.

16. Security Considerations

Connecting ANY service to the Internet creates threads to the service which did not exist before. The emergency call service is especially critical compared to other services lately connected to the Internet. It must work reliably even in case of a major disaster when thousands

of citizens call for help simultaneously. Not only does the service need to be protected but also the liberties of the citizens who might need to use the service must be considered.

The emergency service is an obvious target for a deliberate attack, and specifially a denial of service attack. Mechanisms must be provided to help the emergency networks survive such attacks while continuing to provide service to genuine callers.

Failure of any security mechanism should normally not prevent an emergency call to be established. Unlike most systems, suspicious calls (that is, those where normal security mechanisms are not attempted or they fail to produce expected valid credentials) are normally not dropped, but are processed with the call taker made aware that the information given (location, for example), may not be accurate. As the discussion in Section 5 shows, providing accurate location in the presence of a very wide variety of circumstances is challenging. Exceptions may result in some of the security mechanisms not being able to be deployed, and yet the information may be valid.

When the emergency service is under deliberate attack, the policies on call acceptance may be changed. More stringent compliance to security recommendations may be enforced, or at least calls with full security mechanisms in place may be processed before calls without them.

The decision whether other security mechanisms should be tried or the call be dropped depends on the policy of the citizen, the policy of the call router and the policy of the PSAP and out of the scope of this document.

16.1. Caller Authentication

Fraudulent calls to PSAPs is a significant concern. Current systems

rely on inherent security mechanisms in the PSTN to make sure the identity of the owner of the telephone is known. As Internet technologies are increasingly used to place calls, it is becoming easier to hide the identity of a caller. Use of the SIP Identity mechanism [RFC4474] i is recommended. If SIP Identity cannot be provided, carriers should make use of P-Asserted-Identity, [RFC3325]

In keeping with established customs in circuit-switched emergency calling, authentication cannot be made a prerequisite for routing or accepting an emergency call. However, a call taker may be more suspicious of a caller and request additional information if the call authenticity cannot be verified.

16.2. Location Privacy

Location is sensitive information, it must be protected against disclosure to unauthorized persons. In most jurisdictions placing an emergency call implies disclosure of location to all the entities needing location to properly route and respond to the call. Nevertheless, even in an emergency, callers have an expectation that their location will not be divulged outside of that implied release.

During acquisition of the location information, an eavesdropper or impersonator may obtain location. When DHCP is used, authentication [RFC3118] should be used to protect the location option. Use of TLS in other LCPs should be used. Similarly, TLS should be used with SIP signaling when location is conveyed. However, failure to establish a security association should never be used to drop an emergency call. Rather, the operation should be attempted without the security mechanism.

16.3. PSAP Impersonation

See Section 16.4.

With LoST-based call routing (Section 6), an attacker could modify the mapping entries for one or more locations, re-routing calls destined for them. The security mechanisms for provisioning the data in the LoST database must be robust.

LoST is a distributed database, with many replicas of authoritative

data. An attacker may impersonate a valid LoST server and supply fraudulent data. An attacker may also perpetrate a denial of service attack on LoST servers. These issues are addressed in [I-D.ietf-ecrit-lost].

Finally, the URI LoST returns would normally contain a domain name. The domain can be hijacked by several known attacks. TLS should be used to place calls, with the domain name verified. Using DNSSEC [RFC4033] on the DNS entries is recommended. As above, failure of the security mechanism must not impede the processing of an emergency call; the operation should proceed without security rather than abandoning the call.

16.4. Preventing Call Misdirection

We need to prevent an emergency call reaching a destination other than a PSAP. For example, a rogue UA able to intercept SIP requests might be able to impersonate a PSAP.

In the absence of a globally recognized certificate that ensures that

the owner is a legitimate PSAP, we rely on a chain of trust enforced by the 'sips' URI schema. The 'sips' URI schema forces each SIP hop to route the call only to destinations supporting TLS transport. Each ESRP verifies that the next-hop destination chosen as described in Section 6 corresponds to the server certificate offered by that destination.

16.5. Call Signaling Integrity

Preventing a malicious outsider from manipulating call information in SIP requests can be assured by using "sips" (that is, TLS, hop-by-hop from caller to emergency call taker.

16.6. Media Integrity and Confidentiality

Media integrity and confidentiality can be assured by the use of SRTP[RFC3711].

17. Acknowledgements

This draft was created from a draft-schulzrinne-sipping-emergency-arch-02 together with sections from draft-polk-newton-ecrit-arch-considerations-02.

Design Team members participating in this draft creation include Hannes Tschofenig, Ted Hardie, Martin Dolly, Marc Linsner, Roger Marshall, Stu Goldman, Shida Schubert and Tom Taylor.

18. References

18.1. Normative References

[I-D.ietf-ecrit-lost]
Hardie, T., "LoST: A Location-to-Service Translation Protocol", draft-ietf-ecrit-lost-04 (work in progress), February 2007.

[I-D.ietf-ecrit-phonebcp]
Rosen, B. and J. Polk, "Best Current Practice for Communications Services in support of Emergency Calling", draft-ietf-ecrit-phonebcp-00 (work in progress), October 2006.

[I-D.ietf-ecrit-requirements]
Schulzrinne, H. and R. Marshall, "Requirements for Emergency Context Resolution with Internet Technologies",

draft-ietf-ecrit-requirements-12 (work in progress), August 2006.

[I-D.ietf-ecrit-service-urn]
Schulzrinne, H., "A Uniform Resource Name (URN) for Services", draft-ietf-ecrit-service-urn-05 (work in progress), August 2006.

[I-D.ietf-geopriv-pdif-lo-profile]
Tschofenig, H., "GEOPRIV PIDF-LO Usage Clarification, Considerations and Recommendations", draft-ietf-geopriv-pdif-lo-profile-05 (work in progress), October 2006.

[I-D.ietf-sip-gruu]
Rosenberg, J., "Obtaining and Using Globally Routable User Agent (UA) URIs (GRUU) in the Session Initiation Protocol (SIP)", draft-ietf-sip-gruu-11 (work in progress), October 2006.

[I-D.ietf-sip-location-conveyance]
Polk, J. and B. Rosen, "Session Initiation Protocol Location Conveyance", draft-ietf-sip-location-conveyance-07 (work in progress), February 2007.

[I-D.ietf-sipping-config-framework]
Petrie, D. and S. Channabasappa, "A Framework for Session Initiation Protocol User Agent Profile Delivery", draft-ietf-sipping-config-framework-10 (work in progress), January 2007.

[I-D.rosen-iptel-dialstring]
Rosen, B., "Dialstring parameter for the Session Initiation Protocol Uniform Resource Identifier", draft-rosen-iptel-dialstring-05 (work in progress), March 2007.

[LLDP] "IEEE802.1ab Station and Media Access Control", Dec 2004.

[LLDP-MED]
TIA, "ANSI/TIA-1057 Link Layer Discovery Protocol - Media Endpoint Discovery".

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.

[RFC2396] Berners-Lee, T., Fielding, R., and L. Masinter, "Uniform

Resource Identifiers (URI): Generic Syntax", RFC 2396, August 1998.

[RFC3118] Droms, R. and W. Arbaugh, "Authentication for DHCP Messages", RFC 3118, June 2001.

[RFC3261] Rosenberg, J., Schulzrinne, H., Camarillo, G., Johnston,

A., Peterson, J., Sparks, R., Handley, M., and E. Schooler, "SIP: Session Initiation Protocol", RFC 3261, June 2002.

[RFC3264] Rosenberg, J. and H. Schulzrinne, "An Offer/Answer Model with Session Description Protocol (SDP)", RFC 3264, June 2002.

[RFC3265] Roach, A., "Session Initiation Protocol (SIP)-Specific Event Notification", RFC 3265, June 2002.

[RFC3311] Rosenberg, J., "The Session Initiation Protocol (SIP) UPDATE Method", RFC 3311, October 2002.

[RFC3325] "", 2005.

[RFC3428] Campbell, B., Rosenberg, J., Schulzrinne, H., Huitema, C., and D. Gurle, "Session Initiation Protocol (SIP) Extension for Instant Messaging", RFC 3428, December 2002.

[RFC3515] Sparks, R., "The Session Initiation Protocol (SIP) Refer Method", RFC 3515, April 2003.

[RFC3550] Schulzrinne, H., Casner, S., Frederick, R., and V. Jacobson, "RTP: A Transport Protocol for Real-Time Applications", STD 64, RFC 3550, July 2003.

[RFC3551] Schulzrinne, H. and S. Casner, "RTP Profile for Audio and Video Conferences with Minimal Control", STD 65, RFC 3551, July 2003.

[RFC3693] Cuellar, J., Morris, J., Mulligan, D., Peterson, J., and J. Polk, "Geopriv Requirements", RFC 3693, February 2004.

[RFC3711] Baugher, M., McGrew, D., Naslund, M., Carrara, E., and K. Norrman, "The Secure Real-time Transport Protocol (SRTP)", RFC 3711, March 2004.

[RFC3825] Polk, J., Schnizlein, J., and M. Linsner, "Dynamic Host Configuration Protocol Option for Coordinate-based Location Configuration Information", RFC 3825, July 2004.

[RFC3840] Rosenberg, J., Schulzrinne, H., and P. Kyzivat, "Indicating User Agent Capabilities in the Session Initiation Protocol (SIP)", RFC 3840, August 2004.

[RFC3841] Rosenberg, J., Schulzrinne, H., and P. Kyzivat, "Caller Preferences for the Session Initiation Protocol (SIP)", RFC 3841, August 2004.

[RFC3856] Rosenberg, J., "A Presence Event Package for the Session Initiation Protocol (SIP)", RFC 3856, August 2004.

[RFC3920] Saint-Andre, P., Ed., "Extensible Messaging and Presence Protocol (XMPP): Core", RFC 3920, October 2004.

[RFC4033] Arends, R., Austein, R., Larson, M., Massey, D., and S. Rose, "DNS Security Introduction and Requirements", RFC 4033, March 2005.

[RFC4103] "", 2005.

[RFC4119] Peterson, J., "A Presence-based GEOPRIV Location Object Format", RFC 4119, December 2005.

[RFC4474] "", 2005.

[RFC4676] Schulzrinne, H., "Dynamic Host Configuration Protocol (DHCPv4 and DHCPv6) Option for Civic Addresses Configuration Information", RFC 4676, October 2006.

18.2. Informative References

[I-D.ietf-sipping-service-examples] Johnston, A., "Session Initiation Protocol Service Examples", draft-ietf-sipping-service-examples-12 (work in progress), January 2007.

[RFC3966] Schulzrinne, H., "The tel URI for Telephone Numbers", RFC 3966, December 2004.

Authors' Addresses

Brian Rosen
NeuStar, Inc.
470 Conrad Dr
Mars, PA 16046
US

Email: br@brianrosen.net

Henning Schulzrinne
Columbia University
Department of Computer Science
450 Computer Science Building
New York, NY 10027
US

Phone: +1 212 939 7042
Email: hgs@cs.columbia.edu
URI: http://www.cs.columbia.edu

James Polk
Cisco Systems
3913 Treemont Circle
Colleyville, Texas 76034
US

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

Andrew Newton
SunRocket
8045 Leesburg Pike, Suite 300
Vienna, VA 22182
US

Phone: +1 703 636 8052
Email: andy@hxr.us

Full Copyright Statement

Intellectual Property

copyrights, patents or patent applications, or other proprietary rights that may cover technology that may be required to implement this standard. Please address the information to the IETF at ietf-ipr@ietf.org.

Acknowledgment

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