Providing Emergency Services in Internet Telephony*

Henning Schulzrinne and Knarig Arabshian

Department of Computer Science
Columbia University, New York, NY

ABSTRACT

Assisting during emergencies is one of the important functions of the telephone system. Emergency communications has three components: summoning help during emergencies, coordinating emergency response and notifying citizens and public officials of local emergencies. As we transition to an Internet-based telecommunications system, these functions need to be provided, but there is also an opportunity to add new functionality and improve scalability and robustness. We discuss three aspects of Internet-based communications related to emergencies: First, we describe how Internet telephony can be used to provide emergency call ("911" or "112") services. Secondly, Internet telephony needs to be enhanced to allow prioritized access to communications resources during emergency-induced network congestion. Finally, Internet event notification can be a valuable new mechanism to alert communities to pending or on-going emergencies such as hurricanes or chemical spills.

Keywords: Internet telephony; emergency calls; emergency alerting; emergency communications

1. INTRODUCTION

During emergencies, telecommunications is one of the most important tools to speed response and minimize loss of life and property. Communications systems can help in three different roles: emergency calling, emergency communications, and emergency alerting. Currently, the public switched telephone system (PSTN) and government-run telephone systems are the most important foundation for all three. As we transition from a telephony-focused, circuit-switched network to an integrated-services, packet-switched infrastructure, we need to reconsider how these services are provided. Internet-based communications offers new challenges, as old assumptions related to trust, operation and terminal location no longer hold, but also new opportunities for improved services, more resilient networks and faster response.

Below, we analyze each component in turn and propose an integrated architecture that tries to address Internet-induced problems and leverage the new facilities. Section 2 briefly reviews Internet telephony. Section 3 discusses emergency calling, Section 4 communications during an emergency, and Section 4 a new architecture for alerting citizens and public officials.

We generally assume an architecture outlined in,[1] with the Session Initiation Protocol (SIP)[2] as the signaling framework. This simplifies the discussion and appears reasonable since next-generation Internet telephony networks such as those proposed by PacketCable for cable modems and the Third-Generation Partnership Project (3GPP, 3GPP2) for next-generation wireless use this protocol (UMTS Release 5).

*The work was supported by grants from Cisco Systems, Pingtel, and SIPcomm.
{hgs,knarig}@cs.columbia.edu

2. A BRIEF REVIEW OF INTERNET TELEPHONY

Internet telephony carries audio and video as IP packets across the Internet and intranets. Audio packets are wrapped in RTP,[3] with sessions set up via a signaling protocol. Among signaling protocols, H.323 and SIP are most commonly used.[1,4] Here, for concreteness, we focus mostly on SIP, although the basic mechanisms are likely to apply to other signaling mechanisms as well. In SIP, subscribers are identified either by E.164 telephone numbers[5] such as tel:1-212-555-1234 or by SIP URLs, such as sip:alice@example.com. Each subscriber has at least one *address-of-record* (AOR) that uniquely identifies her. A caller, represented by a user agent, initiates the call by sending a SIP INVITE message to a local *outbound proxy* or a SIP server (proxy) in the destination domain. The SIP URL is independent of the current IP address of the communications devices owned by the user. A user or device creates a binding from its "generic" address-of-record such as alice@example.com to its current network location, such as alice17@pc42.accounting.nyc.example.com. The binding is created by periodically sending a SIP REGISTER request to the home SIP server, with the Contact header conveying the current network address.

SIP messages are identified by randomly-generated tags in their source and destination (From and To headers) as well as a call and request sequence number.

SIP messages can be carried in UDP, TCP or Stream Control Transmission Protocol (SCTP). They are formatted similar to HTTP messages, that is, plain text headers followed by an opaque message body. The message body carries a session description that enumerates the media streams to be used for the call.

SIP has also been extended to generate event notifications[6] and instant messages.[7] Users subscribe to an event with the SUBSCRIBE method and receive notifications via NOTIFY. This event notification facility is used for events that occur during telephone calls, as well as presence notification. We will use this mechanism for emergency alerting (Section 4).

3. EMERGENCY CALLING

Most countries have a means for citizens to use the telephone to summon emergency help, such as an ambulance or the police or fire department. All such systems have four components:

Universal number: A single number for a large geographic region, e.g., 911 in the United States and Canada and 112 in many parts of Europe.

Call routing: Calls from many central offices can reach a single emergency response center; conversely, different callers in the same central office may need to reach different emergency response centers (ERC[†]) since the geographic boundaries of central offices and civil authorities may not be aligned. Individual ERCs or their communications facilities may fail during natural catastrophes, so there is typically a provision to route calls to alternate facilities if necessary.

Caller identification: The emergency response center needs to be able to identify the caller to limit prank calls, to allow call back in case the caller gets disconnected and to log calls for evidence. In almost all cases, the number is delivered even if the caller has suppressed calling number (caller ID) delivery.

Caller location: To speed response and to assist callers unable to identify their current location, the emergency call center should be provided with the street address or geographic location of the caller. For landline phones, this is based on the subscriber billing address, while cellular phones either provide this information derived from a built-in GPS receiver or a network-assisted solution, e.g., based on time-of-arrival differences.[8]

[†]In the U.S., emergency response centers are known as Public Safety Answering Points (PSAPs). For generality, this paper uses the non-standard term ERC.

3.1. United States E911 Service

To set the stage, we briefly describe how the emergency call system works in the United States. Other countries apparently use similar approaches, but details and terminology differ.

In 1968, "911" was established as the universal emergency number. When a 911 call reaches a central office, the switch consults the Selective Routing Database (SRDB) that maps the caller's telephone number to an Emergency Service Number (ESN), a three-to-five digit number that describes a combination of fire, police and EMS agencies. Each ESN is associated with a primary and secondary Public Safety Answering Point (PSAP), where the call is answered and possibly transferred to the appropriate public safety agency. The PSAP also obtains, via a data connection to the telephone company, the caller's street address.

3.2. New Problems for IP-based Communications

Once we transition to IP telephony, many of the assumptions underlying the current system no longer hold. The problems depend on whether we assume that the PSAPs are aware of IP telephony or are seeing IP telephony only through gateways. We refer to these as cases as an IP-enabled PSAP and a legacy PSAP, respectively.

The current system assumes a central mapping from telephone number to street address, maintained by a single telephone company for each household. However, an IP telephony subscriber can obtain her IP service from one company, and an address (e.g., a SIP URL such as sip:alice@example.com) from another one, just like users currently often have many different email addresses not necessarily assigned by their ISP. It is indeed plausible that the same **user@domain** identifier will be both the subscriber's email address and his IP telephony identifier.

3.3. Locating IP Devices

As with email addresses, SIP URLs are not associated with a fixed location or even IP address. Since SIP signaling typically traverses multiple routers, multiple proxy servers (or possibly network address translation devices), an IP-enabled PSAP receiving a SIP call will not necessarily even have access to the caller's MAC or IP address. Thus, none of the traditional identifiers that are roughly equivalent to phone numbers can reliably identify a terminal or terminal location. The addition of virtual private networks (VPNs) aggravates this problem, as it may cause terminals to appear to be local to a LAN, even though they are physically located across a dial-up connection or a completely unrelated LAN.

Independent of the overall architecture, we need to be able to ascertain the location of the indoor, wired IP device. It appears likely that several methods will be used, as all the methods described below have different trade-offs in cost, reliability, and compatibility with existing systems. The FCC requires that the location of cell phones can be ascertained to within 50 m 67% of the time and to within 150 m 95% of the time. There is no mandate to convey elevation, although this is often necessary within high-rise apartment and office buildings.

We can roughly divide the location techniques into three categories: target-based, querier-based and hybrids. In target-based mechanism, the device whose location needs to be known determines its own location and then forwards the location to the querier. Queries-based location determination proceeds in the opposite direction, with the querier trying to map terminal identification to location. Hybrids combine both approaches.

We have implemented a queried-based location mechanism that combines network management, the Cisco Discovery Protocol (CDP),[9] ICMP traceroute and ARP into a single mapping function. The mapping function maps a host name or IP address to an interface and board number. First, the location tracer uses the well-known traceroute mechanism to find the last-hop router closest to the end system. Depending on the configuration, that address may also be usable for SNMP queries. If the device is manufactured by Cisco, it will support the Cisco Discovery Protocol (CDP) that allows to find other attached devices and their management interfaces.

Next, the tracer uses SNMP to query the router ARP table to obtain the MAC address of the device. (It appears to be difficult to do this without downloading the whole ARP table, unfortunately.) Given the MAC address, the MAC

forwarding tables for each VLAN are queried in turn, until the MAC address is found. The device may either be connected directly to that switch, or it may in turn be connected via another switch. The tracer tries to determine, via CDP, whether there is another switch connected to the port, and then queries that switch recursively. Without CDP, a single port with multiple MAC addresses also indicates that another switch is found on the path to the device. The main limitation of this approach is that it is likely limited to queriers within the same administrative domain, as others are unlikely to have access to the SNMP community string or other SNMP authentication information even for read access.

Thus, one can determine the physical Ethernet jack and thus, via an asset database, location (e.g., room and building or longitude, latitude and elevation) for each device. Unfortunately, this only works with managed hubs and switches, and still requires an accurate wiring database. If only the switch location is known, CAT5 or fiber wiring can easily induce uncertainties of several floors or even miles.

IP telephones can utilize other mechanisms for determining their physical location, such as a combination of the following:

Manual entry: If the phone were to force the user to enter a physical location each time the phone is moved, manual entry may be viable. IP phones typically have one or more "owners" which the phone or a server can contact by email or other mechanisms. This is hardly ideal, but as long as phones do not move frequently, a practical approach. If the location is periodically conveyed to the local management system, it can check whether the location is likely to be valid.

Ethernet enhancement: Ethernet switches could be enhanced by sending a periodic broadcast packet on each port identifying the location. In a typical multi-stage switched Ethernet, each device would receive multiple location packets, but these would provide incremental information, e.g., "building 4" and "jack room 4F523". Such functionality is also useful for asset management.

Even without modifying switches, this approach can be easily implemented for crude location, to within a building, as most broadcast domains in modern networks are likely to span no more than a building. (However, many VLANs span a whole campus or service provider.)

Smart jacks: There are commercial products (e.g., made by Panduit) where jacks themselves are active components, allowing them to be queried as to the MAC addresses attached. Recently, a vendor has introduced Ethernet jacks that contain an Ethernet switch.

Wireless-like approaches: While standard GPS does not work indoors, assisted GPS may.[10] There have also been suggestions to use the signal of digital television stations for location.[11, 12] Typically, however, cellular location is accurate to around 100 m, which is not sufficient in an office building or high-rise apartment complex.

Wireless LANs: Some IP phones will use 802.11a/b to communicate with the wired infrastructure. Given the limited range of these technologies, the base station location provides a reasonable approximation to the device location. This estimate can be improved by measuring the signal strength of one or more base stations and correlating it with a map of measured signal strengths.[13] This measurement can be done either by the base station or the mobile.

Loud noises: If there are loud noises, such as police sirens, that can be heard throughout a building, it may be possible to triangulate the location by time-of-arrival differences, assuming tightly synchronized clocks. Since IP phones already have microphones, the additional cost is minimal.

In a variation on this theme, small audio transmitters operating above or below human hearing range could transmit occasional identifying tone sequences.[14, 15]

IR/RF location: There are a number of asset-tracking products that use IR and RF transmitters and sensors installed in ceilings. Such approaches may be appropriate for commercial environments, although they add about $50-$100 to the cost of each device.

X10-based location: X10 is a very simple means of overlaying low bit-rate information by modulating the zero-crossings of household AC circuits. The transmitter and receiver electronics only cost a few dollars. X10 information generally does not travel beyond the a single fused circuit, so that small beacons can be installed in each fuse panel, with receivers in the AC adapter for each IP device. Unfortunately, this does not work well for Ethernet-powered devices which draw their 48 V power from spare pairs in the Ethernet cable[‡].

3.3.1. Legacy PSAPs

A legacy PSAP will see an incoming call coming from a telephone number belonging to the gateway that terminated the SIP call. However, that gateway may not be anywhere near the original caller. Dispatching the fire department to the gateway location is not likely to be helpful. With ISDN it is possible to adjust the caller ID to reflect user location.

Even if the caller ID were to reflect the real location, if the gateway connects to the emergency number, it will reach a local SRDB that will not have access to address and ESN information of far-away locations.

In cases where the gateway is located in the same location as all phones, and as long as IP phones do not wander off the premises (e.g., via virtual private networks), we can address this problem in some circumstances. Cisco has proposed[16] to assign a unique telephone number to each Ethernet jack. The location is maintained in the standard ALI mapping databases.

One intermediate approach that allows gateways to be physically distant from the IP telephones is to publish a directory of the Emergency Service Routing (ESR) numbers for each PSAP, along with the PSAPs geographic serving area. (The ESR is the ten-digit routing phone number that gets a tandem switch translates 911 into.) Each gateway would then consult this database, based on location information that it knows about the IP telephone. It has also been proposed to assign the exchange 911 in each area code to PSAPs.[17]

3.3.2. IP-enabled PSAPs

Transitioning to Internet-enabled PSAPs adds a number of new capabilities, but also poses new problems. Internet-enabled PSAPs allow a much richer communications environment. For example, emergency operators could use video to communicate with people speaking sign language or to gain a better understanding of the emergency situation or monitor first-aid efforts. Video from the communicator to the emergency caller can allow the emergency operator to instruct the caller in first aid. Text-based messaging is currently only available via specialized Telecommunications Device for the Deaf (TDD) equipment, but such equipment is not widely available in offices or public places. Internet-based communications can also easily accommodate biometric data, e.g., from patients that are monitored while living at home.

Also, current PSAPs require a highly specialized infrastructure. An IP-enabled PSAP only requires network connectivity and a commodity PC, making it easy to move operations if, for example, the primary location is affected by natural disasters. Indeed, emergency operators do not even have to be in one location; they can be distributed across multiple locations, including their homes, if equipped with a DSL or cable modem connection.

SIP has built-in capabilities for call distribution ("forking") where a single incoming call is routed to multiple destinations either in parallel or sequentially, until somebody answers the call. Parallel forking allows load-balancing among PSAPs, while sequential forking supports overflow routing, where calls that exceed a certain waiting time are automatically routed to a backup PSAP.

[‡]Ethernet power is becoming popular since it simplifies providing battery-backed power to IP telephones along with the LAN infrastructure, to enable lifeline service.

The problems for IP-enabled PSAPs are similar to those for a legacy PSAP, with the additional issue that an IP phone cannot even rely on a gateway to reach *some* PSAP, even if an inappropriate one. Below, we present an architecture that tries to address these issues.

3.4. An Architecture for Internet Telephony

It is likely that the existing analog PSTN will be around for several decades, but it makes sense to run IP-enabled PSAPs even while many callers still use circuit-switched systems. Third-generation wireless systems, as mentioned, will use IP for voice communications, thus encouraging direct IP-connectivity, rather than gatewaying back into the circuit-switched environment.

An outline of the architecture is shown in Fig. 1. The architecture supports a mixture of IP-enabled and legacy PSAPs. As PSAPs become IP-enabled, database entries in only a few places need to be updated. During an emergency call, the IP phone contacts the local outbound proxy, as it does for every call. (The outbound proxy is located using DHCP.[18]) A special identifier, "sos@aor-domain", has been proposed as the universal destination for emergency calls.[19] The "aor-domain" is the domain corresponding to the address-of-record of the caller. This approach requires no configuration when the terminal roams into networks away from home. The outbound proxy intercepts the session setup request (INVITE) and tries to determine the caller's location. If the end system can determine its location, it includes it in the request; otherwise, the outbound proxy tries to use the MAC-backtracking mechanism described. If all else fails, the outbound proxy has to assume that the device is located close to the outbound proxy, but indicates the uncertainty, and then relies on human interaction to determine the precise location.

Since the outbound proxy should not need to keep track of a database of PSAPs, we propose that a national or regional SIP-based call router registers with the proxy as user "sos", subject to appropriate authentication. Following terminology in,[20] we label this an emergency provider access directory (EPAD). Thus, any call will automatically be redirected to the EPAD. The router then maps the location information provided by the proxy to an emergency provider. The EPAD can route the call, acting as a SIP proxy, or simply provide the SIP URI or telephone number to the proxy receiving the call, acting as a SIP redirect server.

For redundancy, multiple EPADs can register. Normal SIP forking rules ensure that servers will be contacted in some order, trying all until a working one has been found. It is not yet clear how the EPADs find the proxy. While one can imagine various service location mechanisms, the EPADs are not likely to change frequently, so that simple manual configuration may be sufficient. The owner of a SIP proxy obtains the addresses and provides the EPADs with a secret that they can use to register. This prevents rogue EPADs from registering and avoids that anybody but the owner of a domain can invoke EPAD registration.

We are implementing an initial version of this architecture that makes use of the sip-cgi[21] functionality in SIP proxies such as our sipd. Sip-cgi provides language-neutral programming interface similar to the common gateway interface (cgi) for web servers. In our implementation, the sip-cgi script invokes a script returning the user location. The location is then passed to an external database engine run by a national emergency location database provider, using a proprietary XML-over-HTTP protocol. The database provider returns the 10-digit routing number; the script then causes the call to be proxied there, through the local gateway. (The gateway map needs to be set up so that normal call restrictions limiting long-distance calls are bypassed.) Results are cached, as the same query is likely to be repeated. This also ensures continuity of operation should the connection to the database provider be temporarily disrupted.

Since the audio packets from the call itself do not have to traverse the proxies, it does not much matter which proxy intercepts the SOS call. The call will eventually reach the caller's home domain, so that the caller can ensure (by selecting an appropriate service provider, say) that emergency calls are handled appropriately. The same functionality is needed for telematics applications, where cars are equipped with automatic dialers that contact an emergency call center operated by a private service provider.

Calls can also be made directly to a PSAP by designating a universal URL such as “sip:sos.cc” that reaches the main PSAP within the country with top-level domain “cc”. Once the head PSAP gets the 911 call and determines the caller’s location, it will forward the call to the local PSAP within the caller’s area. The number of PSAPs is modest (about 6,800)[22] and the total United States 911 call volume of 250,000 calls/day could easily be managed by one SIP server. XXXXX Such an approach has two disadvantages: configuration and robustness. is that it requires the end system to be configured with the terminal’s current country, unless an international organization is willing to perform country-level routing of calls directed to “sos.int”. If database lookup scaling is a concern, the hierarchy can be extended to states or provinces, or even counties or metropolitan areas. The robustness problem remains, as all emergency calls require that communication to a destination that is potentially far away works reliably. This can be addressed to some extent by using regional mirror servers, with appropriate DNS-based redirection.

While the local registration mechanism using EPADs appears to be the better solution, both can coexist easily.

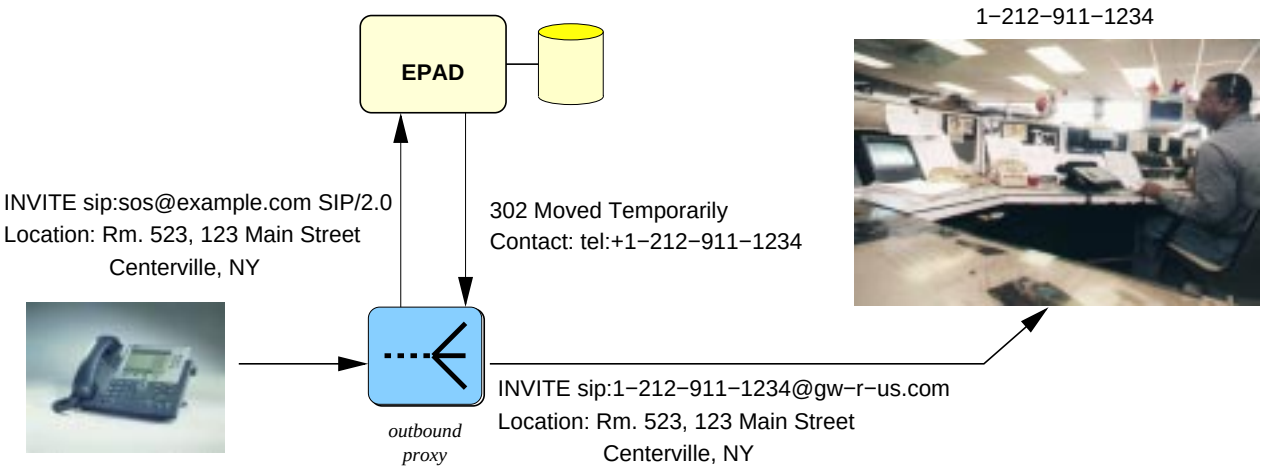

Figure 1. SIP call routing for emergency calls

3.5. Location and Identity Privacy

It appears that most jurisdictions require that callers reveal their identity and location to the emergency call center, to discourage prank calls and speed emergency response. However, there may be concern that including the caller’s name or location in the request may compromise the privacy of the caller. Location and caller identity information[23] can be protected in two ways, by encrypting the signaling information using IPsec or TLS or by just encrypting the identity and location information with the destination public key. (More precisely, a random symmetric key is protected via public-key cryptography, using S/MIME mechanisms[24]). The latter approach has the disadvantage that the client needs a mechanism for obtaining the destination’s public key. There currently is no standardized mechanism in the Internet that supports this, although efforts are underway within the SACRED working group.[25]

Conversely, the network may want to authenticate the caller, as the SIP end system itself is unlikely to have a personal certificate. The subscriber and his network provider already share a key for registration updates, so that the provider can use this information to sign the request.[26]

3.6. Emergency Calls for Other VoIP Protocols

While this paper focuses on how a SIP-based Internet telephony architecture can make emergency calls, similar approaches apply to the two other major protocol architectures, Megaco and H.323, as well.

In the Megaco architecture, a Media Gateway Controller (MGC) drives one or more Media Gateways (MGs), such as PSTN gateways or desk phones. MGCs are then connected via SIP or, less commonly, H.323, so that the same considerations apply.

If H.323 is used, each zone has a gatekeeper that routes calls from local terminals. This gatekeeper would intercept emergency calls and forward them to the appropriate location. Instead of REGISTER, the EPAD could use the H.225.0 RAS (registration, admission, status) protocol to register with the gatekeeper.

4. EMERGENCY COMMUNICATIONS

During emergencies, telecommunications facilities are often strained by both official and private communications. Rescue workers and law enforcement need to coordinate activities, while ordinary citizens want to find out about the whereabouts and health of friends and relatives.

There are at least two areas in the existing Internet architecture that need to be modified for such use, namely the IP layer and the signaling layer.[27] At the IP layer, differentiated services already offers a mechanism to give better service to certain users. The main problem is authenticating the users that should have access to such service. Since one does not want to add authentication information to each packet header, some form of boundary filtering and admission control is needed.

For "I'm alive" notifications, it may make sense to give each device a set of tokens that they can expend on elevated-priority packets, thus encouraging frugal notification options such as email, instant messaging or short calls. For signaling priority, there are two cases, namely access to the existing PSTN and prioritizing resources in SIP proxy servers. For the former, existing military and civilian emergency networks offer multi-layer preemption priority. For example, the U.S. defense network defines levels ranging from "routine" to "critic-ecp". This functionality needs to be made available to IP-based systems. We have proposed[28] that a simple SIP header field indicates the desired resource access priority, addressing priority handling in proxies and gateways. The header field supports multiple different namespaces, as different organizations have chosen different sets of labels. If desired, the same mechanism could also be used for email (if authenticated) and HTTP, although its usefulness is in doubt there.

5. EMERGENCY NOTIFICATION SYSTEMS

Emergency notification systems, the third component of an emergency communication system, allow government officials to notify a community of an emergency and the subsequent precautionary measures that should be taken. Such notifications can be between government agencies, e.g., from a national law enforcement agency to local police departments, or, more interestingly, from government agencies to citizens. Systems similar to the one we are describing below can also be useful for private enterprises, e.g., for alerting personnel working for a chemical plant.

While PSTN-based emergency notification systems exist, they are limited in scale, relatively slow and provide only fairly basic information. Thus, we explore below how event notification protocols, such as SIP,[6] can be used for this application.

5.1. Overview of Current Emergency Notification Systems

The Emergency Broadcast System (EBS) was developed in 1963 to notify the United States public of emergency situations. In December 1995, the Federal Communications Commission (FCC) started replacing the EBS with the Emergency Alert System (EAS).[29] While the EBS was limited to use by the President, the EAS can also be called upon by state and local authorities.

The EAS distributes information across AM, FM and television stations. Each station listens to at least two other stations for EAS alerts and automatically rebroadcasts them for its local area. An emergency announcement consists of an alert tone, an FSK-encoded digital data stream of about eight seconds, an audio message and an end-of-message

(EOM) indicator. The digital data stream contains information about the type of warning (e.g., hurricane or civil disturbance), the county or part of a county it applies to, the date and time issued and the authority issuing the alert. The format is similar to the weather alerts issued by the National Weather Service.

In addition to the EAS tailored to radio and TV, there are emergency notification networks and products that offer emergency alerting to a local area.[30–33] Old systems used sirens, but they provide only minimal information content, beyond getting people to turn on their TV or radio. Some community alert systems use loudspeakers, while others rely on telephone circuits. For example, Reverse 911[31] dials phone and fax numbers from a list or within a specific geographic area.

5.2. SIP-based Emergency Notification System

We propose to enhance the EAS and community alert systems with a SIP-based event notification system. In principle, any network-based event notification system could be used, but since end users will likely already have SIP-based event notification capabilities[6] on their 3G handsets and PC desktops[§], for example, it makes sense to use this "commodity" technology rather than invent a new one specific to emergency alerts.

The basic architecture is straightforward. We envision a hierarchical subscription system, where information is disseminated from national governments to state and local governments, and then citizens, with information generated at any of these levels. Users subscribe to servers at the next higher level and may in turn become servers for the levels below (Fig. 2).

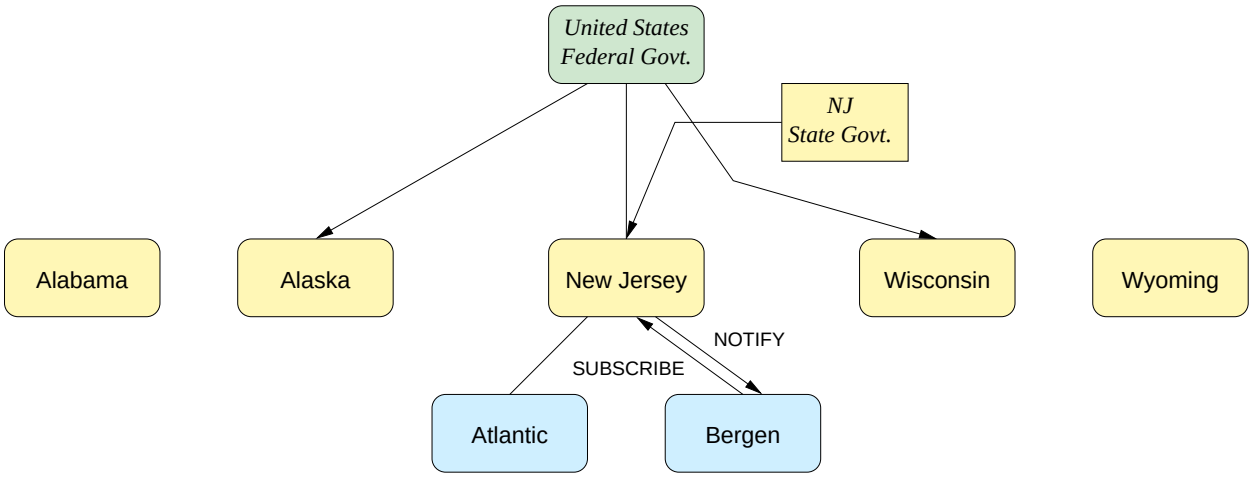

Figure 2. Alerting hierarchy

As shown in Fig. 3, a client sends a SUBSCRIBE request to the appropriate server. The request contains the event description (Event header) and the network destination for notifications (Contact header), as well as any authentication information. If the server approves the subscription, it adds the subscriber to the appropriate event list and generates NOTIFY requests to the subscriber when an alert occurs. Subscriptions time out to prevent devices that are no longer interested or capable of receiving alerts from consuming network resources. The Expires header indicates the duration of the subscription. Subscribers simply update and refresh their subscription periodically, e.g., once a day. Subscribers can explicitly remove themselves from notifications by setting the expiration time to zero.

[§]Microsoft XP includes a SIP-capable instant messaging client.

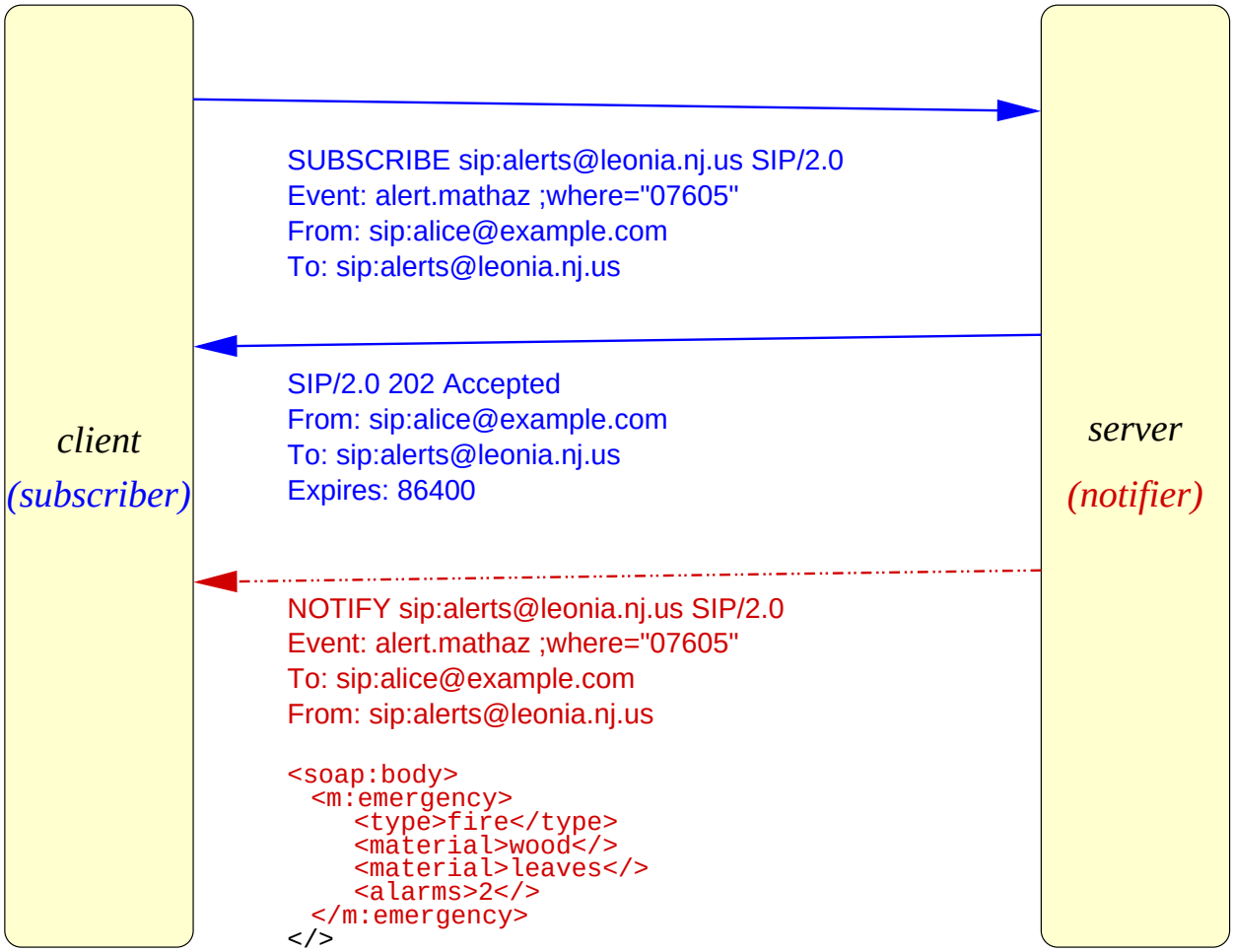

Figure 3. Protocol exchanges for event alerting

User agents can submit multiple SUBSCRIBE requests if they are interested in a range of events. We also envision extending the subscription mechanism to include a geographic range, limiting the type of notifications sent to the subscriber. The subscriber can also indicate what type of media it can accept, e.g., whether audio or text notifications are appropriate. The SUBSCRIBE request may contain a message body in a standardized format that further describes the subscriber capabilities.

The emergency notification is sent to the address that subscribed earlier. This can either be a specific host, identified by a host name or IP address, or a more generic "user@domain" address. SIP supports request routing, where intermediaries, so-called SIP proxies, can rewrite the destination address and forward the request. This has the advantage that changes in address do not have to be propagated to the source, affording privacy to the subscriber and improving system scalability.

The notification may also contain a message body that further describes the nature of the emergency in a machine-parseable way. We propose the use of the XML-RPC schema, as used for the Simple Object Access Protocol (SOAP), as it already offers the necessary data abstraction functionality and there are existing implementations that can be re-used for emergency alert. For example, a forest fire notification might contain details about projected movements of the fire, evacuation instructions and similar information. This can then be rendered appropriately and maybe even integrated into, say, a geographic information system. Notifications to emergency response personnel are likely to be far more detailed than those to citizens.

5.3. Finding Servers

We have glossed over the issue as to how a subscriber finds the server that is appropriate for her. This is an example of the more general wide-area service location problem,[34, 35] and it is likely that a solution that supports more than finding alert servers is most appropriate. In the absence of such a general solution, we can envision a number of ad-hoc solutions. For citizen subscriptions, the simplest is to advertise the subscription address via out-of-band means, such as web pages and newspaper advertisements. Also, one could provide a well-known address, similar to the current 911 number. The SIP redirect server at that address would not provide notifications directly, but rather redirect the subscription request to the appropriate server, based on the geographic location (postal code, say) and event type specified. Local agencies then register with the central server, using the normal SIP REGISTER binding mechanism (Fig. 4).

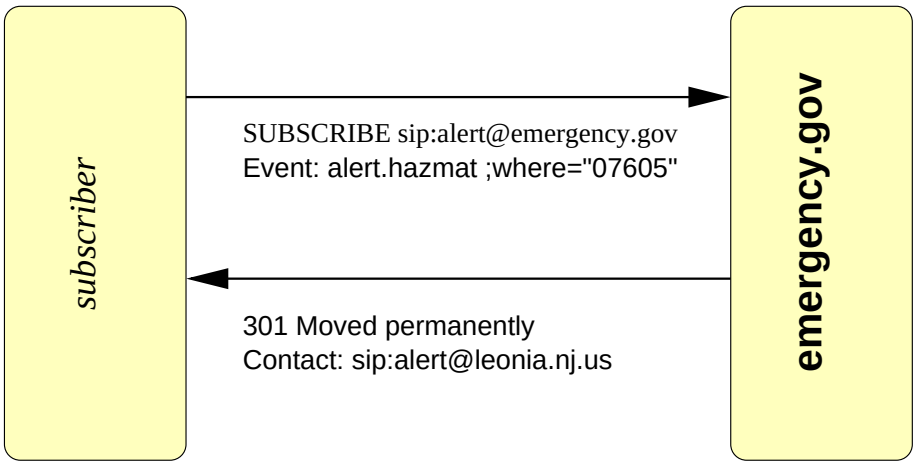

Figure 4. Creating an alert binding

Government agencies presumably have information distribution arrangements that can provide authentication credentials and logical server addresses (e.g., sip:tornado@nws.noaa.gov). Since the subscription address can remain fixed and is not subject to interruptions like area code changes or physical moves of agencies, it is likely to remain constant for many years.

5.4. Benefits

There are a number of benefits of using an Internet-based emergency alert system, as described above:

Device neutrality: There are likely to be range of SIP-based end systems, ranging from IP telephones, 3G wireless handsets, IM/presence software to embedded devices. Thus, the emergency alert system can migrate to new devices, without having to explicitly be extended to handle them.

More information: The information embedded in the current EAS system is rather limited and hard to extend without upgrading end systems. The SIP event notifications can carry much more detailed information, tailored to different needs. For example, automated systems cannot benefit from voice announcements, but need detailed information on duration of events and the ability to cancel emergency alerts. Even for human users, it becomes much easier to provide a multitude of languages.

Automated action: Related to the previous items, SIP event notifications can provide more detailed guidance. For example, the message can contain pointers to web pages providing detailed instructions on appropriate behavior or fugitive information.

Currently, citizens get all EAS emergency alerts, limiting the system to major emergencies. The more fine-grained event subscription mechanism allows events to be distributed only to the small population that may be affected, e.g., by a water main break.

As described, we embed an RPC-like mechanism so that the alert can trigger appropriate automated actions.

Stronger authentication: The existing authentication mechanism relies on manual codebooks and the difficulty of spoofing an over-the-air signal. The mechanism described here can use true cryptographic authentication which is more amenable to automated processing and less likely to be spoofed. (Under the existing system, it would not be hard to drive by an EAS receiver with a small transmitter and make it distribute a false alarm.)

Lower resource consumption: A one-minute alert call consumes about 480 kbytes (one way), while an alert notification is likely to be at most a few hundred bytes long. Thus, the same bandwidth can reach approximately 1,000 times as many people in the same time period. It is also easy to leverage web hosting and similar facilities with large amounts of bandwidth.

Integration with current systems: It would be straightforward to feed EAS and EDIS (emergency digital information system) into the SIP emergency alert system. The combination is then more likely to reach more people, even those not listening to the radio or watching TV. Since the system reach can be tailored very narrowly, the system can also easily integrate alerts that are of a less urgent nature, such as traffic accidents or other police activity.

Out-of-area notification: Current notification systems assume that only those in close physical proximity of the emergency event need to know about the event. However, there are cases where the recipient of the information is currently away, but still needs to be alerted. For example, owners of summer cottages need to be aware of impending storms so that they can summon appropriate assistance or have their property checked on after the fact. Those at work may need to know about conditions affecting their home.

5.5. Authentication and Authorization

Authentication and authorization are vital for an emergency alert system, for both subscriptions and notifications. Subscriptions need to be authenticated for distribution of events to government officials, but authentication is also useful to prevent a single citizen subscribing multiple times or, worse, accidentally or intentionally redirecting the subscription of their neighbor. (This could occur if somebody spoofs the SIP From header and then inserts its own address as the Contact value.)

The authentication of citizens and officials is likely to require different approaches. There are far fewer emergency response agencies, which are also more likely to already have access to a mechanism for distributing information securely. In addition, the emergency alert server needs to prove its identity when issuing notifications, to prevent creating panic with bogus alerts.

SIP currently uses either HTTP Digest or transport and network-layer security to authenticate requests. For digest security, a server challenges the client within an error response, which then causes the client to reissue the original request with it the challenge encrypted with a shared secret. This scheme does not expose the user's password to those listening in.

5.5.1. Authenticating Residents

A relatively simple authentication mechanism suffices for authenticating residents. For example, a resident wishing to receive alerts signs up via a web page and is then e-mailed a secret key that she can use to authenticate the subscription. While email delivery is hardly secure, it should be good enough to prevent random users from creating nuisance subscriptions or "stealing" the subscription of a user.

5.5.2. Authenticating Government Officials

Shared secrets can be distributed by any secure mechanism. Alternatively, CMS-based encryption[36] can provide both authentication and confidentiality using public-key cryptography. This replaces the problem of distributing shared secrets with distributing private/public key pairs. Instead of basing access to notifications on identity, it is probably easier to base it on role and agency, i.e., having signed capabilities.

5.6. Notification Architecture

As described earlier, we envision a hierarchical system, with multiple entry points for alerts. For emergency alert systems, robustness is of prime importance. One way to achieve this is similar to the multiple-source mechanism of the EAS, namely by having lower levels of the alerting hierarchy subscribe to multiple upper levels. Notifications may also need to be sent among peers, for example, between neighboring states or counties. Again, cross-subscriptions are likely to be helpful here.

For general alerts, each level also subscribes to events generated by its children. That way, it becomes less critical as to where somebody subscribes to events and allows alerting across the civil hierarchy. For example, a local police department in Alabama could generate a fugitive alert to New York authorities if it had reason to believe that the person in question may have boarded a plane bound for New York. This avoids having to keep updated contact lists for peer law enforcement and emergency response units; instead, addressing is by type of emergency and geography, with destinations determining their coverage and expertise.

6. CONCLUSION

IP telephony and the associated protocol infrastructure need to offer at least the same level of emergency-related services that citizens have come to expect from the existing telephone system. However, rather than simply replicating the existing system using packets, there is an opportunity to create more functional, robust and flexible systems to enhance existing capabilities.

ACKNOWLEDGMENTS

The device location algorithm was implemented by Jisoo Lee.

REFERENCES

1. J. Rosenberg and H. Schulzrinne, "The IETF Internet telephony architecture and protocols," *IEEE Network* **13**, pp. 18–23, May/June 1999.
2. J. Rosenberg, H. Schulzrinne, G. Camarillo, A. Johnston, J. Peterson, R. Sparks, M. Handley, and E. Schooler, "SIP: Session initiation protocol," RFC 3261, Internet Engineering Task Force, May 2002.
3. H. Schulzrinne, S. Casner, R. Frederick, and V. Jacobson, "RTP: a transport protocol for real-time applications," RFC 1889, Internet Engineering Task Force, Jan. 1996.
4. J. Toga and J. Ott, "ITU-T standardization activities for interactive multimedia communications on packet-based networks: H.323 and related recommendations," *Computer Networks and ISDN Systems* **31**, pp. 205–223, Feb. 1999.
5. A. Vaha-Sipila, "URLs for telephone calls," RFC 2806, Internet Engineering Task Force, Apr. 2000.
6. A. Roach, "SIP-specific event notification," Internet Draft, Internet Engineering Task Force, Mar. 2002. Work in progress.
7. J. Rosenberg *et al.*, "SIP extensions for instant messaging," Internet Draft, Internet Engineering Task Force, July 2001. Work in progress.
8. S. Tekinay, "Wireless geolocation system and services (special issue)," *IEEE Communications Magazine* **36**, Apr. 1998.
9. Cisco, "Catalyst token ring switching implementation guide - frame formats," documentation, Cisco, Mar. 1998.
10. G. M. Djuknic and R. E. Richton, "Geolocation and assisted GPS," *IEEE Computer* **34**, pp. 123–125, Feb. 2001.
11. M. Rabinowitz and J. Spilker, "Positioning using the atsc digital television signal," rosum corporation whitepaper, Rosum, Redwood City, California, 2001.
12. P. Fyfe, C. Kelley, D. Martoccia, and E. Martin, "Augmented navigation signal sources," in *Proc. of ION GPS*, (Salt Lake City, Utah), Sept. 2001.
13. P. V. Bahl and V. Padmanabhan, "Radar: An in-building rf-based user location and tracking system," in *Proceedings of the Conference on Computer Communications (IEEE Infocom)*, (Tel Aviv, Israel), Mar. 2000.
14. N. B. Priyantha, A. Chakraborty, and H. Balakrishnan, "The cricket location-support system," in *ACM/IEEE International Conference on Mobile Computing and Networking (MobiCom)*, pp. 32–43, (Boston, Massachusetts, USA), Aug. 2000.
15. L. Girod and D. Estrin, "Robust range estimation using acoustic and multimodal sensing," in *Proc. of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2001)*, (Maui, Hawaii), Oct. 2001.
16. Cisco, "Emergency responder version 1.1," data sheet, Cisco Systems, San Jose, California, Dec. 2001.
17. National Emergency Number Association, "9-1-1 tutorial." http://www.nena9-1-1.org/desktop/9-1-1
18. H. Schulzrinne, "DHCP option for SIP servers," Internet Draft, Internet Engineering Task Force, Mar. 2002. Work in progress.
19. H. Schulzrinne, "Universal emergency address for SIP-based internet telephony," Internet Draft, Internet Engineering Task Force, Feb. 2002. Work in progress.
20. ComCARE, "Communications for coordinated assistance and response in emergencies." http://www.comcare.org/.
21. J. Lennox, H. Schulzrinne, and J. Rosenberg, "Common gateway interface for SIP," RFC 3050, Internet Engineering Task Force, Jan. 2001.
22. T. E. Wheeler, "Statement of Thomas E. Wheeler president and chief executive officer Cellular Telecommunications and Internet Association before the Senate Committee on Commerce, Science and Transportation regarding implementation of E-911 service," Oct. 2001.

23. J. Peterson, “A privacy mechanism for the session initiation protocol (SIP),” Internet Draft, Internet Engineering Task Force, Apr. 2002. Work in progress.
24. B. Ramsdell and Ed, “S/MIME version 3 message specification,” RFC 2633, Internet Engineering Task Force, June 1999.
25. A. Arsenault and S. Farrell, “Securely available credentials - requirements,” RFC 3157, Internet Engineering Task Force, Aug. 2001.
26. J. Peterson, “Enhancements for authenticated identity management in the session initiation protocol (SIP),” Internet Draft, Internet Engineering Task Force, Apr. 2002. Work in progress.
27. F. Baker, “IEPS requirement statement,” Internet Draft, Internet Engineering Task Force, Nov. 2001. Work in progress.
28. J. Polk and H. Schulzrinne, “SIP communications resource priority header,” Internet Draft, Internet Engineering Task Force, Nov. 2001. Work in progress.
29. Federal Emergency Management Agency (FEMA), “Background on the emergency notification system (EAS).” http://www.fema.gov/pte/rep/easrep.htm.
30. Community Alert Network. http://www.can-intl.com/.
31. Reverse 911, “Interactive community notification system.” http://www.reverse911.com.
32. Intelligent Wireless Solutions, “Web-enabled emergeny notification.” http://www.inwireless.com/.
33. Longboat Key Police Department, “SAM emergency alerting system.” http://www.longboatkey.org/Departments/police/sam.htm.
34. P. Castro, B. Greenstein, R. Muntz, P. Kermani, C. Bisdikian, and M. Papadopouli, “Locating application data across service discovery domains,” in *ACM/IEEE International Conference on Mobile Computing and Networking (MobiCom)*, pp. 28–42, (Rome, Italy), July 2001.
35. J. Rosenberg and H. Schulzrinne, “Internet telephony gateway location,” in *Proceedings of the Conference on Computer Communications (IEEE Infocom)*, pp. 488–496, (San Francisco, California), March/April 1998.
36. R. Housley, “Cryptographic message syntax,” RFC 2630, Internet Engineering Task Force, June 1999.