ECRIT Working Group
Internet-Draft
Expires: Sept 6th, 2006

James Polk
Cisco Systems
Andrew Newton
VeriSign
March 6th, 2006

Emergency Context Routing of Internet Technologies Architecture Considerations
draft-polk-newton-ecrit-arch-considerations-02

Status of this Memo

Copyright Notice

Abstract

This document discusses architectural considerations for emergency context routing of Internet technologies. The purpose of this document is to provide a systemic view of emergency context routing, discuss unresolved issues, and explain the relationship of some of the proposals to these issues, while discussing potential directions that might be still be necessary for the working group to investigate.

Table of Contents

1. Introduction

The solution to proper emergency call identification, management and routing over the Internet involves many components and coordination between them. This document describes the necessary interaction between these components. The information given in this document may not be complete, and some of the issues presented in this document may not be resolved by the community. The intent of this document is to describe a "big picture" view of the process, describe prevailing thoughts on this subject and describe unresolved issues in hopes of bringing about consensus within the community on these topics.

The current architecture of Emergency Context Routing of Internet Technologies is composed of the following:

Bootstrapping: delivery of configuration and location information to the client at or near power-up time.

Conversion: conversion of location information into forms usable in mapping and conveyance, if necessary.

LCMS Mapping: conversion of endsystem location information into addresses usable to initiate or progress communication towards an emergency call center (a PSAP).

Conveyance: delivery of endsystem location information to an emergency call center during the emergency call (used for first responder dispatch).

There are many unresolved issues regarding these steps and related matters. The following list is not exhaustive, but includes most of the issues brought grouping discussions to date (and a few new ones).

Universal emergency identifier: there needs to be a universal emergency identifier to prevent highly localized usage and confusion by users and systems of what applies to a certain region, and what does not.

Security: the security properties necessary for the proper protection of LCMS data are not well understood.

Data distribution: the distribution of LCMS data closer to the points of queries within the Internet.

Extensibility: the methods for extensibility in all components of the system must be well understood.

Conflation: many of the components proposed for use in the routing of emergency calls have other uses and most have not been primarily designed for the emergency call routing case.

1.1 Division of Labor

As stated above, not all of the components used in the process of routing an emergency call to the correct emergency call center over the Internet have been defined for the exclusive use of this case, and therefore not all of the specification work is being conducted within the scope of the charter of the ECRIT working group.

Bootstrapping of location information, both geospatial and civic, via DHCP is work in progress in the GEOPRIV working group. Bootstrapping of URI references via DHCP is work in progress in the DHC working group.

The definition of location objects and the use of schemas to describe location is work in progress in the GEOPRIV working group. The specification of conveyance of location objects via SIP is work in progress in the SIP working group.

The mapping of location information to URIs is the primary function of the ECRIT working group. The set of mechanisms and services

working in conjunction with each other to conduct this mapping is referred to as the Location Context Mapping System, or LCMS by this document for the purposes of clarity.

1.2 Terminology, Acronyms and Definitions

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in [RFC 2119].

The following terms and definitions will be used throughout this document:

Application (Layer) Provider (ALP) - provider of application level services such as a voice over IP service that is completely divorced from the Link provider, and may be divorced from the Internet Attachment Provider of an endsystem that is merely providing a layer 3 connectivity service.

ALP - Application (Layer) Provider

Emergency Services Gateway (ESGW) - The special IP to circuit-switched gateway that front-ends an emergency services Selective Router (SR) (which directs all TDM based 911/112 type calls to the appropriate PSAP within a given physical region)

Emergency Services Routing Proxy (ESRP) - a special instance of a SIP Proxy that understands emergency routing of messages identified as emergency messages to a PSAP based on the location of the caller that is included in the message

ESGW - Emergency Services Gateway

ESRP - Emergency Services Routing Proxy

IAP - Internet Attachment Provider

Internet Attachment Provider (IAP) - typically the organization that provides Internet or IP access to the client (i.e. assigns the client an IP address and/or provides the client with IP transit.

Location Context Mapping System A set of coordinated mechanisms and services that map a physical location (geospatial or civic) to a network address.

LCMS Location Context Mapping System

LIS - Location Information Server

PSAP - Public Safety Answering Point

Location Information Server (LIS) Provides a mapping function to relate unique identifiers for IP devices at physical network access points and the geographic descriptions of their location (e.g., civic location/street addresses or geographic coordinates). Responds to queries for location information.

Public Safety Answering Point (PSAP) - the emergency response call center talking the local emergency calls from people in distress. This facility can be logical, and can transfer (reroute) any request sent to it to another facility deemed more appropriate to receive the request.

2. Bootstrapping

Bootstrapping delivers configuration information to the client. The most obvious use of bootstrapping is for an endsystem to ask for and receive its IP address, Subnet Mask and Default Gateway addresses. This could also be used to deliver the geospatial or civic address of the client to it for future use. This is typically done at device or individual application boot times. Unlike other parts of the architecture, bootstrapping is the only phase of the emergency call routing process that does not have a single solution. This is due to the many network configuration techniques used by access networks.

There are three well-known methods of bootstrapping:

1. Using the DHCP protocol
2. Using the PPP protocol
3. Using IEEE 802.1 LLDP-MED

For location information, there are two types: location information regarding a civic address (e.g. 123 Main St) and geospatial information (e.g. x, y, z coordinates).

The set of configuration data types has not been discussed or resolved. But the following types of configuration information have been noted: 1) references to LCMS servers, 2) references to PSAPs, and 3) references to location information servers.

References to LCMS servers obviate the need for global hierarchies of LCMS data directories (which have proven politically difficult in other voice over IP matters) and reduce the coordination to only the necessary jurisdictional boundaries.

References to PSAPs obviate the need for the mapping steps in cases where location is not likely to be a determining factor in emergency call routing (e.g. location is fixed and the emergency call center is known).

References to location information servers enable better separation

for knowledge of location from the access network.

3. Conversion

Conversion of location information from one format to the other format is conducted for the benefit of the LCMS servers and conveyance of the location information. There are two aspects to this conversion: syntactic conversion of the location information from the binary formats used in bootstrapping to the XML format used in PIDF-LO, and conversion of the civic location information to geospatial location information or vice versa.

Syntactic conversion is a necessary function, and it can take two different forms: conversion from the binary DHCP format to the PIDF-LO conveyance format and conversion to a LCMS format if the LCMS interface does not use PIDF-LO.

Conversion of location information from a civic address to geospatial coordinates or from geospatial coordinates to a civic address is much more controversial. There is no doubt that civic addresses are much easier to consume by humans than geospatial coordinates, however conversion from geospatial coordinates to a civic address can be error prone and in cases involving large areas (e.g. a farm, an outdoor park, etc) the resulting civic address can be of limited utility. Further, civic coordinates only address a fraction of the land of this planet, as civic addresses are to a great degree tied to the existence of a nearby street, which are prevalent in cities, but are scarce to non-existent in rural areas. Therefore, a combination of the two formats will be required regardless of mankind's consumption preference.

4. LCMS Mapping

The creation of an LCMS, which will convert location information into references (i.e. URIs) to emergency call centers, is the primary area of work for the ECRIT working group. There are two times in which this LCMS function can occur successfully:

1. before the device attempts contact with a PSAP, and

2. after the device initiates contact with the PSAP, but before the correct PSAP is determined by the ESRP making the LCMS query.

Accomplishing mapping of a device's location with the proper PSAP can be done statically or dynamically, or in a layered - combination - approach.

Statically - If a site is small enough, for example residential or

Small or single building business scale, all devices in either of these types of locations can be configured to download, or have

downloaded to them, their location information and the SIP or SIPS URI for contacting the appropriate PSAP for that location.

Dynamically - for the above types of locations, or for larger locations, the client's location can be used to "look up" the appropriate PSAP SIP or SIPS URI, and have this configuration information downloaded to each client requesting it. This can also be pushed to all clients regardless of whether they asked for the information or not. This pushing of the PSAP URI(s) can be at some interval to maintain freshness of the URI(s), as stale URI(s) are a concern to some.

For the static configuration, for each given endpoint or section of a network, a known PSAP SIP or SIPS URI is downloaded to the client(s) without the clients providing their individual location to perform the LCMS function. In the case of dynamic configuration of a URI, the client provides their location to an LCMS server to do this look-up, with the answer sent to the client for future use by any application on that client, including for emergency services.

In a layered approach, there does not need to be a one-size-fits-all solution, but a combination of ways in which the mapping resolution is accomplished towards the goal of having the emergency call set-up reach the appropriate PSAP. For example, a solution with the most risk can be used last, but in a way it does not rely on any other steps to have occurred to function properly. In this scenario, the simplest means of mapping with the least risk can be performed initially, before the device ever knows it will generate an emergency call set-up message. In this way, this first mechanism is done at boot-time, and the mapping during the actual emergency call can still happen whether or not the bootstrapping took place or not. This layered approach would be with a goal of solving the function of mapping one of the independent steps towards entering the appropriate PSAP SIP or SIPS URI into the INVITE message. When this URI is learned should not matter, as long as it is the appropriate URI.

Another combined approach can be attained in the following scenario: if the endsystem knows of an authoritative LCMS server regardless of which network or domain the client is connected to, the endsystem can contact this server to get its PSAP SIP/SIPS URI based on its location provided by the local access network. In this scenario, an endsystem can have a trust association established with a particular server (or server service) that it contacts as soon as it either learns its location from a local network/domain or somehow determines it has moved while remaining "connected" to that network/domain.

For the device configuration of a PSAP SIP or SIPS URI, currently

only DHCP is being proposed as a solution [ID-DHCP-URI]. This proposal is not an LCMS function because it does not send location to a server and receive the mapping answer containing a URI. DHCP

is used here to only deliver the URI to be used as the Request-URI of the emergency SIP INVITE.

To date there are three protocol proposals for LCMS: LUMP, ECON, and DNS SOS. LUMP is an XML-based, SOAP solution with emphasis on data distribution. ECON is an XML-based, IRIS solution with emphasis on lightweight data exchange, and DNS SOS is a DNS-based solution with emphasis on re-using DNS semantics.

Finally, some jurisdictions may find it necessary to withdraw the LCMS protocol from public view and place its function within an ESRP. At the option of the jurisdiction, more than one ESRP function may be implemented in series, to provide increasingly precise routing to the appropriate PSAP.

5. Conveyance

5.1 Location Conveyance

Once the address of the PSAP is known, either through bootstrapping or through LCMS mapping, a call can be initiated with the PSAP. Location information is sent to the PSAP as meta-data of the call using PIDF-LO. This facet is not part of the ECRIT WG, but cannot be overlooked. Even if the caller contacts the appropriate PSAP, that PSAP will still require knowledge of where the caller is in order to dispatch emergency responders (i.e. help). Issues regarding the acquisition of this knowledge are discussed in Section 7.2.

Passing location information within the voice application protocol is commonly referred to as "location-by-value". There exists another concept where a reference to a location server is passed within the voice application protocol instead of the actual location information. This is known a "location-by-reference". Location-by-reference is not without controversy, and its plusses and minuses will be discussed in a future version of this document.

5.2 Identity Conveyance

There is a general desire on behalf of PSAP operators to have the identity of a caller conveyed within a call. This identity has two parts: an identity asserted to be authentic and a call-back reference for re-establishment of communications.

Of the two parts of this identity conveyance, the authentication of the identity is the most contentious and burdensome to solve. For example, if a traveler with a phone purchased in London were to make an emergency phone call in New York, what trust relationship exists between the authorities of New York and a phone retailer in London?

Making matters more complicated, conveyance of identity for

emergency calling is not a work item for any IETF working group.

6. Universal Emergency Identifiers

Throughout the world, there are many different numbers in use to signal emergency phone calls. Some counts have this number as high as 60 unique number sequences worldwide. This lack of uniformity also leads to collision. For example, in some areas of the world, dialing the number 0 is used for calling for help, whereas in other parts of the world, this would not accomplish its intended emergency meaning, resulting in the caller being told to hang up and dial another number.

Therefore, one of the ECRIT requirements is for a universal emergency identifier to signal an emergency. The need for it to be universal (or well-known) is threefold: so that all the components in the emergency call routing process may properly operate based on its presence in a message, to avoid collisions with other purposes (as stated above), and so that clients may localize its meaning to end users. The issue is that there is not just one single identifier related to emergency calls, but that there are many identifiers related to emergency calls for various specific types of emergencies.

Multiple identifiers lead to confusion and many have overlapping meaning. For example, the separate identifiers "mountain" and "rescue" could mean the same thing to a user needing to be rescued from a mountainous area. Additionally, some jurisdictions have custom identifiers that are either unused globally or have a limited applicability.

Each LCMS proposal takes a different approach to solving this problem. ECON takes the simplest approach, specifying a simple list of 3 identifiers. DNS SOS specifies a list of six identifiers. LUMP specifies a hierarchy of identifiers.

What is not clear, or has not been well defined, is the need for even the simplest of these approaches. It is not even well understood if end users, in an emergency situation, will be able to rationalize the difference between "emergency" and a simple list of "police", "fire", and "medical". While some have suggested this is in practice in some parts of the world today, that does not necessarily mean this will become universal in usage. It appears that if there is a single master identifier with more than one sub-identifier, that this arrangement should be used where it is understood, and perhaps adopted elsewhere as jurisdictions decide to segment this capability based on education within that area.

What appears obvious to avoid is to have different identifiers for help in different parts of the world moving forward. If only 'sos.police@...' reaches the police in a country where Alice does

not live when she sends a SIP INVITE to 'sos@...', ECRIT as a WG has not likely accomplished its goal.

Locales that choose to have sub-identifiers for granularity must have an architected solution for the higher level identifier as well.

7. Security Considerations

7.1 Security of the LCMS

It is the goal of the working group to develop a dynamic LCMS protocol that is both secure and responsive, two features that tend to conflict with each other. Security for this mapping solution has fallen into two broad categories: object signing and channel security.

Object signing has three benefits: integrity of data during distribution, the potential for utilization in UDP packets, and pre-calculation of cryptographic data. However, in cases where partial matching of the query are to be allowed (i.e. parts of a civic address are to be ignored in the query) or the query cannot be known ahead of time (i.e. the whole set of geospatial coordinates is known but not in practical terms), object signing will require "on-line" signing which negates advantages in data distribution and cryptographic pre-calculations.

In addition, the use case regarding the invalidity of a signed object may be no different from that of a validly signed object. Users confronted with an emergency may not be able to appreciate the difference in validity, and even if they did, may not alter their course of action (i.e. they continue with the emergency call anyway).

Channel security requires expensive cryptographic calculations that cannot be computed ahead of time and requires multiple packet exchanges (i.e. roundtrips) to establish. However, this approach has the benefit of securing all parts of the transaction, and unlike object signing, is well used and well understood on the Internet.

The security properties of each of the three LCMS proposals is as follows:

LUMP uses both channel security (TLS connections to the query server) and object signing (signed entries in the database).

DNS SOS uses both channel security (TLS in connections to fetching polygons and other information) and object signing (in DNSSEC for the protection of NAPTR records).

ECON uses channel security (TLS connections to the query server) as

an option setup by the operator of the service and a lightweight UDP transfer protocol for scenarios where security is not needed.

7.2 Security of Location Conveyance

There is a general desire to protect PSAPs from malicious calls. Yet, how this is to be accomplished is not clear or well defined.

Complicating this issue is the simple fact that many PSAPs will accept a call without location information related to the caller. Additionally, many PSAPs give priority or parity to location information collected by a human operator from a human caller. Due to this fact, it has been observed that any security mechanism put into place by ECRIT can simply be routed around by directly contacting a PSAP.

In cases where a PSAP would wish to disregard calls of unknown provenance, no guidelines have clearly been stated as to how such trust relationships would be erected.

7.3 Security of Bootstrapping

Where critical decisions might be based on the value(s) of the bootstrapping process, such as a URI option from [ID-DHC-URI], DHCP authentication in [RFC3118] SHOULD be used to protect the integrity of the DHCP options.

Since there is no privacy protection for DHCP messages, an eavesdropper who can monitor the link between the client and destination DHCP server to capture any URIs in transit.

When implementing a DHC server that will serve clients across an uncontrolled network, one should consider the potential security risks.

All that said, if DNS is not secure, and bootstrapping is difficult to secure based on what it takes to accomplish [RFC3118], is securing the mapping service worth the effort and pain to achieve?

7.4 Security of Conversion

Location is a vital part of emergency messaging. As discussed earlier, an endsystem will not likely be in control of which format of location it receives from a roamed to network. For more fixed endsystems, this should not be the case. If an endsystem does receive location in a format it knows an application on that endsystem does not prefer, the endsystem can contact a server or

service, if one is known, to convert this format to the other format.

As a non-emergency example, most humans understand street addresses much better than GPS coordinates. If a roaming device, say using 802.11 at a hotspot, acquires its location via DHCP Option 123 [RFC 3825], it can determine if an application used by that device prefers the civic format when using an instant messaging application on that device. Before the IM application is launched, or as the app is launched, the device can seek a conversion server to perform this format conversion operation. How does a client learn of a server that can do this? [ID-DHC-URI] provides one means for a device to learn the URL of a server that can do this function, or this can be preconfigured in the device as a trusted source for this conversion, wherever it is - as long as there is connectivity to that trusted source.

In the emergency case, perhaps the device knows it needs to convert to the civic format to have an ESRP perform an LCMS query, but the local network gave it a geospatial location only. If the conversion server is preconfigured, this provides the ability to have the two devices, say the phone and the conversion server, establish a trust relationship, perhaps with pre-shared keys. This reduces the round trip times, making it more efficient. This also provides a more secure means of communication since both entities 'know' each other.

8. Data distribution

There is a desire to locate LCMS data in the LCMS close to the points of query in the Internet for performance reasons. In addition, some jurisdictions distribute authority for this data upon hierarchical lines. However, the needs for data distribution beyond these high-level requirements are not well known. For instance, the known life expectancy of data distributed to caches is not well known, nor are update procedures in the distribution of this data.

Each of the three LCMS proposals addresses this problem in a different way:

DNS SOS relies upon the cache machinery of DNS. The population of DNS caches with location information is accomplished through validation of caller locations (a process during provisioning of a callers location and before any emergency call). This proposal does not address early cache expiration due to limited cache memory, by accepted practices of DNS operations, or by routine maintenance of DNS servers.

LUMP defines a caching mechanism that identifies objects by hash value in order to accomplish a mesh of caches between nodes. The population of the caches with location information is accomplished through validation of caller locations (a process

conducted during provisioning of a callers location and before any emergency call).

ECON defines no preference for data distribution due to the limited requirements available. However, it does describe two methods that could be employed: the serialization of data to files for distribution via standard transfer protocols and an on-line, incremental approach capable of distributing entries before their activation date.

9. Extensibility

Within the ECRIT working group, there appears to be rough consensus on the need for extensible mechanisms, and hence an extensible LCMS protocol capable of extensions in its query interface and resulting output. This desire for extensibility is born from a general sense that not all the problems of emergency call routing over the Internet will be fully exposed until deployment of a first generation solution and from a more specific sense of the incompleteness of the civic address schema in PIDF-LO.

As an example of the more general case, the document [ID-ECRIT-JAPAN] describes a numeric address code equivalent to the civic elements <A1> through <A5> in PIDF-LO used in conjunction with geospatial coordinates to conduct emergency call handling in Japan. As an example of the more specific case, PIDF-LO does not contain an element to describe street section numbers as used in Taiwan.

10. Conflation

As mentioned in the introduction, many of the components used in the process of routing emergency calls were not designed primarily for this task and are being developed in working groups that do not have emergency call routing explicitly defined in their chartered scope.

As a general example, conveyance of location information within a call also has applicability to delivery businesses, such as pizza restaurants that need to know the location of the caller for delivery purposes. In a more technical sense, much of what is known about civic addresses worldwide is a result of the study of postal delivery, and therefore schemas used as location input for emergency call routing may not be tuned properly.

Within the ECRIT working group, there are no requirements regarding the resilience of the emergency call routing process as it relates to inputs that have not been designed for this specific purpose.

11. Rerouting/Transfer

Another facet of the ECRIT WG that has not been addressed is what to

do if a PSAP receives an emergency call and the call should not have been routed to that PSAP. What does the PSAP do next, and can this

action be automated? Does the PSAP have an additional screening capability in some ESRP at the PSAP interior edge to do a final check that the call set-up is to the appropriate PSAP, taking steps not yet defined if this PSAP is not the appropriate one for this particular call set-up?

This is more a rerouting function of the call set-up, or of a call transfer function if the call is answered before determining this is an inappropriate PSAP. For example, VPNs will likely cause some emergency calls to go erroneously to the city that the caller's corporate offices are located in rather than to where the caller is. This has not been considered to date, yet likely should be - as message reroute should be possible anytime an ESRP misdirects a message towards a PSAP, just not he appropriate PSAP.

12. Acknowledgements

Nadine Abbott provided text regarding ESRP usage and comments regarding the inclusion of discussion of location-by-value vs. location-by-reference. Richard Statsny suggested this document would be a more complete introduction to the problem space if it included information regarding identity conveyance.

13. References

13.1 Normative References

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

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

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

13.2 Informative References

[ID-DHC-URI] J. Polk, "A Dynamic Host Configuration Protocol Option for Requesting and Receiving Uniform Resource Identifiers", draft-polk-dhc-uri-03.txt, "work in progress", March 2006

[ID-ECRIT-JAPAN] H. Arai, M. Kawanishi, draft-arai-ecrit-japan-req-01.txt, "work-in-progress", May 2005

Author's Address

James M. Polk

3913 Treemont Circle
Colleyville, Texas 76034
USA

Phone: +1-817-271-3552
Fax: none
Email: jmpolk@cisco.com

Andrew Newton
21345 Ridgetop Circle
Dulles, VA 20166

Phone: +17039483382
Email: andy@hxr.us

Appendix A. Additional stuff

Intellectual Property Statement

Disclaimer of Validity

an "AS IS" basis and THE CONTRIBUTOR, THE ORGANIZATION HE/SHE
REPRESENTS OR IS SPONSORED BY (IF ANY), THE INTERNET SOCIETY AND THE

Copyright Statement

Acknowledgment

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