Network Working Group
Request for Comments: 5606
Category: Informational

J. Peterson
NeuStar, Inc.
T. Hardie
Qualcomm
J. Morris
CDT

August 2009

Implications of 'retransmission-allowed' for SIP Location Conveyance

Abstract

This document explores an ambiguity in the interpretation of the <retransmission-allowed> element of the Presence Information Data Format for Location Objects (PIDF-LO) in cases where PIDF-LO is conveyed by the Session Initiation Protocol (SIP). It provides recommendations for how the SIP location conveyance mechanism should adapt to this ambiguity.

Documents standardizing the SIP location conveyance mechanisms will be Standards-Track documents processed according to the usual SIP process. This document is intended primarily to provide the SIP working group with a statement of the consensus of the GEOPRIV working group on this topic. It secondarily provides tutorial information on the problem space for the general reader.

Status of This Memo

This memo provides information for the Internet community. It does not specify an Internet standard of any kind. Distribution of this memo is unlimited.

Table of Contents

1. Introduction

The Presence Information Data Format for Location Objects (PIDF-LO [RFC4119]) carries both location information (LI) and policy information set by the Rule Maker, as is stipulated in [RFC3693]. The policy carried along with LI allows the Rule Maker to restrict, among other things, the duration for which LI will be retained by recipients and the redistribution of LI by recipients.

The Session Initiation Protocol [RFC3261] is one proposed Using Protocol for PIDF-LO. The conveyance of PIDF-LO within SIP is specified in [LOC-CONVEY]. The common motivation for providing LI in SIP is to allow location to be considered in routing the SIP message. One example use case would be emergency services, in which the location will be used by dispatchers to direct the response. Another use case might be providing location to be used by services associated with the SIP session; a location associated with a call to a taxi service, for example, might be used to route to a local franchisee of a national service and also to route the taxi to pick up the caller.

Some ambiguities have arisen in the interpretation of Rule Maker policy when PIDF-LO is conveyed by SIP. The following sections explore the problem and provide a recommendation.

2. Problem Statement

The <retransmission-allowed> element of RFC 4119 was designed for use in an environment like that of Section 4 of RFC 3693, in which Location Information (LI) propagates from a Location Generator through a Location Server (LS) to a Location Recipient (LR). In this architecture, it is the responsibility of the Location Server to act on the rules (policy) governing access control to LI, which are in turn set by the Rule Maker. The most important of these responsibilities is delivering LI to authorized Location Recipients and denying it to others. Internal to [RFC4119]-compliant location objects (LOs) are additional privacy rules which are intended to constrain Location Recipients. These include the <retransmission-allowed> element. This element is intended to prevent a compromise of privacy when an authorized recipient of LI shares that LI with third-party entities, principally those who are not authorized by the Rule Maker to receive LI. For example, a user might be willing to share their LI with a pizza shop, but they might not want that pizza shop to sell their LI to a targeted advertising company that will contact the user with coupons for a nearby hair salon.

Bear in mind, however, that <retransmission-allowed> is not intended to provide any protocol-level mechanism to prevent unauthorized parties from learning location through means like eavesdropping. It is merely a way to express the preferences of the Rule Maker to the LR. If the LR were, for example, legally bound to follow the privacy preferences expressed by Rule Makers, then they might incur liability if they ignored the <retransmission-allowed> parameter. No further privacy protection is assumed to be provided by <retransmission-allowed>.

There is a use case for LI that involves embedding it in a SIP request that will potentially traverse multiple SIP intermediaries before arriving at a user agent server (UAS). In this use case, one or more intermediaries might inspect the LI in order to make a SIP routing decision; we will hereafter refer to this as location-based routing. Common examples could include emergency services and other more mundane cases where the originator of a SIP request wants to reach a service in proximity to a particular geographic location, such as contacting a nearby pizza shop. In both such cases, the UAC may intend for selected intermediaries and the UAS to have access to the LI. In the pizza case, for instance, the user agent client (UAC)

shares an address both for location-based routing and additionally so that the pizza shop reached by that routing has the address to which a pizza should be sent.

This location-based routing use case for LI has a number of important disconnects from the RFC 3693 model. Unlike the RFC 3693 model, there is no LS designating to which specific entities LI will be sent. There may be multiple intermediaries between the UAC and UAS, some of which will need or want to inspect LI (which would seem to qualify them as LRs) and some of them will not. While SIP proxy servers generally are not [RFC4119]-aware and do not need to inspect SIP request bodies in order to perform their function, nothing precludes proxy servers inspecting or logging any SIP message bodies, including LI. Furthermore, it is very difficult for the UAC to anticipate which intermediaries and which eventual UAS a SIP request might reach.

This architecture is further complicated by the possibility of sending location information by-reference, that is, placing a URL where LI can be retrieved in SIP requests instead of using a PIDF-LO body (commonly called including the PIDF-LO by value). Depending on the qualities of a reference, further authorization checks may be performed before LI is retrieved, LI may be customized depending on who is asking, and so forth. As will be discussed in greater detail below, the conveyance of a reference may have very different privacy properties than conveying a PIDF-LO body by-value in a SIP request.

In this architecture, the question of who is an "authorized recipient" from the point of view of the Rule Maker has been muddy.

The SIP elements along the path are authorized to receive and forward the SIP message; does that make them automatically authorized recipients of the LI it contains? The final target of the SIP message will receive the LI along with other information, but it may be different than the initial target in a variety of scenarios; is it authorized to read the LI?

These questions and concerns are particularly problematic when <retransmission-allowed> is set to "no" (the default case). This core concern might be put as "to whom does <retransmission-allowed> apply in location-based routing?" More specifically:

Is any entity that reads LI bound by <retransmission-allowed>? If so, does that mean a proxy that performs location-based routing is unable to forward a request and complete a SIP call if <retransmission-allowed> is "no"? Alternatively, must they strip the location body from the message in order to complete the call?

If the proxy does not understand RFC 4119, it may forward a SIP message containing a policy statement <retransmission-allowed> set to "no". Is any proxy that does understand RFC 4119 required to parse the LI for this statement, even if it would not do so in order to route the message?

Is there a need for SIP-level indications regarding retransmission for the benefit of entities that do not understand RFC 4119?

Since the UAC cannot anticipate who may receive a SIP request, how do we understand who the intended LR is in the location-based routing case? Can a UAC have intended for there to be multiple serial LRs in a transmission? If so, if one LR is authorized to retransmit to another LR, how will it know it is not also authorized to transmit LI to other third parties (i.e., how will the serial LRs know to whom they are authorized to retransmit)? How could all of this be designated?

3. Recommendation

The following sections provide a recommendation for how the <retransmission-allowed> flag should be understood in a SIP environment. The core semantics of this recommendation represent the consensus of the GEOPRIV working group. While Section 3.5 proposes a syntax that might be adopted by the SIP WG to implement these semantics in its protocol, the actual syntax of SIP is the responsibility of the SIP WG.

3.1. Goals

After extensive discussion in both GEOPRIV and SIP contexts, there seems to be consensus that a solution for this problem must enable location-based routing to work even when the <retransmission-allowed> flag is set to "no". A solution should also give the Rule Maker the ability to allow or forbid the use of LI for location-based routing and the ability to allow or forbid the use of LI for the consumption of the endpoint.

3.2. Core Semantics

Consensus has emerged that any SIP entity that receives a SIP message containing LI through the operation of SIP’s normal routing procedures or as a result of location-based routing should be considered an authorized recipient of that LI. Because of this presumption, one SIP element may pass the LI to another even if the LO it contains has <retransmission-allowed> set to "no"; this sees the passing of the SIP message as part of the delivery to authorized recipients, rather than as retransmission. SIP entities are still

enjoined from passing these messages outside the normal routing to external entities if <retransmission-allowed> is set to "no", as it is the passing to third parties that <retransmission-allowed> is meant to control.

This architecture is considerably different from the presumptions of RFC 3963, in that authorized recipients pass the LO on to other authorized recipients, but it seems to be the most sensible mechanism given SIP's operation.

To maintain the Rule Maker's ability to affect the consumption of this information, two different mechanisms may be used to limit the distribution of LI and one may used to limit the sphere in which it may be used; these are discussed below.

3.3. Limiting Access

3.3.1. Limiting Access Using Public Key Encryption

One way of limiting access to LI is to encrypt the PIDF-LO object in a SIP request. If the originator knows which specific entity on the path needs to inspect the LI, and knows a public key for that entity, this is a straightforward matter. It is even possible to encrypt multiple instance of PIDF-LO, containing different policies or levels of location granularity, in the same SIP request if multiple entities along the path need to inspect the location.

This is most likely to be effective in cases where the originator does not wish the LI to be inspected by intermediate entities and has the public key for the target of the SIP message, as it is very difficult for the originator to anticipate the intermediaries through which a SIP message will pass. It may also be useful in limited environments where the originator has a trust relationship with a specific SIP element (e.g., a "home" or first-hop proxy) and it wants to reveal that LI only to that element.

Note that even in the case where the originator intends to encrypt LI for the benefit only of the target of the message, it may be quite difficult to anticipate the eventual endpoint of the message. These encrypted LIs will not be useful in any case where the anticipation of the originators is not met.

An additional problem posed by this approach is that it requires some sort of public key discovery system, which compounds the operational complexity significantly. While this method is included for completeness, it is the consensus of the working group that the deployment scenarios in which this is appropriate will be relatively few; we do not believe it is an appropriate baseline approach.

3.3.2. Limiting Access Using Location-by-Reference

Another, more feasible approach is leveraging location by reference. When a SIP request conveys a reference, it cannot be properly said to be conveying location; location is conveyed upon dereferencing the URI in the question, and the meaning of <retransmission-allowed> must be understood in the context of that conveyance, not the forwarding of the SIP request.

The properties of references, especially the security properties, vary significantly depending on the nature and disposition of the resource indicated. Clearly, if the referenced PIDF-LO is available, in the same form, to any entity along the SIP signaling path that requests it, then inserting a reference has no advantages over inserting LI by value (and introduces wasteful complexity). However, if the Rule Maker influences the results of the dereferencing process, including determining who can receive LI at what degree of granularity and what policies are bound with the LI, the security properties are different.

It might superficially appear that this suffers from the same problems as the encryption approach, since the Rule Maker must anticipate a set of entities who are authorized to receive location information. The difference is that this set does not need to be communicated in the SIP request in order for authorization decisions to be made. There is a world of difference between managing a whitelist of a thousand parties that might ask for LI and sending a SIP request containing a thousand differently encrypted adumbrations on LI -- the former is commonplace and the latter is impossible. Additionally, some Rule Maker policies might not even require the establishment of an exhaustive whitelist. For example, it may be that there exists a finite set of commercial requestors that the Rule Maker would like to block, in a manner similar to the way ad-blockers operate in modern web browsers.

In any system where one makes an authorization decision, a certain cost in authentication must be paid -- the greater the assurance the greater the cost. The precise cost will of course depend on the URI scheme of the reference. For SIP, Digest has a low computational cost but requires pre-established keys, which limits applicability. RFC 4474 Identity does not require any pre-association, but it does make signaling more heavyweight and requires the deployment of additional features in the network, including a web-like public key infrastructure (PKI).

But even if no authentication takes place, in the Location-by-Reference (LbyR) case the meaning of <retransmission-allowed> is unambiguous -- each entity to which LI is conveyed in the dereference

process is bound by the retransmission policy. The cost of the reference itself is of course the server that maintains the resource. While not every SIP client has access to an appropriate server for this purpose, the fact that PIDF-LO builds on the typical SIP presence service makes this less implausible than it might be. Moreover, the LbyR approach casts the conveyance architecture in a manner familiar from RFC 3693, with a Location Server receiving requests from Location Recipients, which may be accepted or denied. This allows the preservation of the original semantics of <retransmission-allowed>.

3.3.3. Refraining from Including Location Information

The most fundamental mechanism for limiting access to location information is simply not including it. While location-based routing might conceivably occur in almost any SIP message in the future, there is no requirement that location be included in the general case to support it. If it is not included and is required, an appropriate error indicating the lack may be returned and the choice made to continue communication with the information included. This challenge and response will slow the establishment of communication when it is required, but it is the most basic way to ensure that location distribution is limited to the times when it is required for communication to proceed.

3.4. Choosing among the Available Mechanisms

Refraining from including location is the most appropriate choice for systems that do not wish to reveal location to any party in the SIP path.

Location-by-Reference is generally recommended as the most deployable mechanism for limiting access to LI which is passed via a SIP message. It is significantly easier to deploy than public key discovery systems, allows for both whitelists and blacklists, and can scale in ways that the inclusion of multiple encrypted bodies cannot. Encryption may be used in a limited set of circumstance where location-by-value must be used.

3.5. Indicating Permission to Use Location-Based Routing in SIP

The discussion in Section 3.3.2 describes 3 mechanisms for limiting the distribution of LI to specific entities. There remains the problem of limiting the use to which LI included by value or by reference may be put. In order to meet the need to limit that use, this document recommends the creation of a syntactical element in SIP to carry this information. As an exemplary concrete proposal, we recommend a "Location-Routing-Allowed" header as described below.

When "Location-Routing-Allowed" is set to "Yes", the Rule Maker is indicating permission to use the included LI for location-based routing. When "Location-Routing-Allowed" is set to "No", the originator is indicating that this use is not permitted. "Location-Routing-Allowed" being set to "No" has no protocol-level mechanism for enforcement of this behavior; like the PIDF-LO <retransmission-allowed> being set to "no", it is a way for the Rule Maker to express a preference to the SIP elements, which are LI recipients. It may, however, present a significant optimization. Where a location-by-reference is included with "Location-Routing-Allowed" set to "No", the SIP elements along the path know that they do not need to attempt to dereference the location information; this is significantly faster than attempting the dereference and being denied at the authentication stage.

We recommend that "Location-Routing-Allowed" be made mandatory-to-implement for elements complying with [LOC-CONVEY].

We recommend that it appear in any SIP message that contains a location, whether by reference or by value.

We recommend that any SIP message containing a location but no "Location-Routing-Allowed" header should be treated as containing a "Location-Routing-Allowed" header set to "no".

We recommend that a UA be allowed to insert a "Location-Routing-Allowed" header even when it has not included a location, in order to set the policy for any locations inserted by other SIP elements.

This allows the UA to assert that it is a Rule Maker for locations, even when the network architecture in which the UA is present inserts the location into SIP messages after the UA has originated the SIP exchange.

We recommend that any SIP element inserting a location, whether by reference or by value, insert a "Location-Routing-Allowed" header if one is not already present. If one is present, it should not be overridden by the SIP element inserting the location.

We recommend that any SIP element not the originator of a message and not inserting a location be enjoined from inserting a "Location-Routing-Allowed" header.

3.6. Behavior of Back-to-Back User Agents

Back-to-back user agent (B2BUA) behavior is often difficult to proscribe. There are many uses of B2BUAs, and the rules that apply to location would depend on the actual use case. This section suggests what any SIP mechanism arising from this document might wish to consider with regard to B2BUA behavior.

In most uses of B2BUAs, they act as a simple intermediary between the nominal originating and nominal terminating UAs, that is, a proxy that does something proxies aren't allowed to do. In such cases, the B2BUA must conform to any new routing-allowed mechanism if it chooses an outgoing route. As this document advises proxies, <retransmission-allowed> does not apply to the B2BUA in this case, and the B2BUA must copy the LI, the new routing-allowed, and existing <retransmission-allowed> values.

Where the B2BUA in fact does act as an endpoint (terminating the session and originating a different session), <retransmission-allowed> applies to it, and it must not copy location if <retransmission-allowed> is "no". If it chooses a route for the outgoing leg, any new routing-allowed mechanism applies to it.

Encryption lets the originator control who, including B2BUAs, is allowed to see location. On the other hand, using encryption with LI, which is needed for routing, is problematic, in that it is often difficult to know in advance which elements do location-based routing. Similarly, using Location-by-Reference instead of location-by-value provides additional control to the originator over B2BUA behavior by controlling who can dereference. See Section 3.4 for more guidance on this trade off.

4. Security Considerations

The privacy and security implications of distributing location information are the fundamental subject of this document.

5. Acknowledgements

James Polk provided a series of questions regarding the specifics of the Location-Routing-Allowed mechanism, and this resulted in the recommendations in Section 3.4. Thanks to Brian Rosen for the text on B2BUAs.

6. Informative References

[LOC-CONVEY] Polk, J. and B. Rosen, "Location Conveyance for the Session Initiation Protocol", Work in Progress, March 2009.

[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.

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

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

Authors' Addresses

Jon Peterson
NeuStar, Inc.

EMail: jon.peterson@neustar.biz

Ted Hardie
Qualcomm

EMail: hardie@qualcomm.com

John Morris
Center for Democracy & Technology

EMail: jmorris@cdt.org