Requirements for the Support of Continuously Varying Values in Presence

draft-thomson-simple-cont-presence-val-req-02

Status of this Memo

This Internet-Draft is submitted to IETF in full conformance with the provisions of BCP 78 and BCP 79.
Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.
Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as “work in progress.”
The list of current Internet-Drafts can be accessed at http://www.ietf.org/ietf/1id-abstracts.txt.
The list of Internet-Draft Shadow Directories can be accessed at http://www.ietf.org/shadow.html.
This Internet-Draft will expire on January 4, 2010.

Copyright Notice

Copyright (c) 2009 IETF Trust and the persons identified as the document authors. All rights reserved.
This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents in effect on the date of publication of this document (http://trustee.ietf.org/license-info). Please review these documents carefully, as they describe your rights and restrictions with respect to this document.

Abstract

The attributes of continuous-valued data are examined in respect to presence systems. The limitations of the existing presence system with respect to continuous-valued data is examined. Requirements are formulated that would enable the use of the presence system for this data, with an emphasis on providing the watcher with a means of control over the measurement process.
1. Introduction

The provision of continuously varying parameters using presence requires specific support by the presence infrastructure (Day, M., Rosenberg, J., and H. Sugano, “A Model for Presence and Instant Messaging,” February 2000.) [RFC2778]. Existing methods assume that perfect information is always available to the Presence Agent (PA). These methods rely on filtering, which limit the data provided to a watcher, but do not provide the watcher any control over the quality of the
actual data. The current presence system is ill-equipped to handle imperfect data.

Perfect information is available with no delay; perfect information is absolutely accurate. The absence of perfectly accurate information, can be inversely characterized as the existence of uncertainty. The measurement of any physical property is subject to some degree of uncertainty. The impact of this depends on the purpose that the information is intended for. If a low standard of quality is demanded of the data, the existence of uncertainty might be ignored without negative consequences; near-perfect information is often perfectly adequate. However, in some cases uncertainty can be significant to the intended use of the information.

Timely availability of information also affects the service provided by the PA. For a continuously varying value, the amount of time required to acquire a value increases as the required accuracy increases. The relationship between time and accuracy is likely to be non-linear; gains in accuracy progressive require increasing amounts of time to acquire. Furthermore, as a value changes over time, the value provided to a watcher could be invalidated by the time the watcher receives the information.

Dealing with imperfect information using presence requires additional protocol support. To properly support continuous-valued data the presence system needs to provide a way for a watcher to indicate their preferences for data quality and timeliness. This document outlines requirements for presence that enable the use of continuously varying parameters.

Location information is used throughout this document as an exemplary example of a continuous-valued datum. The requirements in this document are intended to provide a basis for the definition of presence features that support location while giving due consideration to support of other continuous variables.

This document specifies requirements that might already be covered by existing publications. To ensure the integrity of the overall concept - and guarantee is completeness - those requirements are retained. References are provided to relevant documents.

1.1. Conventions used in this document

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 [RFC2119] (Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels," March 1997).
2. Continuous-Valued Data

Continuous-valued data are defined by this document as having a continuous value space. Therefore, the set of possible values for any continuous-valued datum is infinite. In contrast, many existing presence elements are discrete-valued. Discrete-valued elements have a limited number of possible values. Continuous-valued data is most regularly numerical, or is composed of numerical data.

Physical properties are the prototypical example of the use of continuous-valued data. Any value expressed in the SI units (metres, kilograms, seconds, ampere, kelvin, mole or candela) and units derived from these are the focus of this document.

Examples of continuous-valued presence elements could include descriptions of the physical characteristics of a presentity or its immediate environment. This could include position in space; weather related values such as ambient temperature and wind speed; light or noise intensity; remaining battery life of a device; weight and size.

2.1. Continuous and Discrete Values in Presence

The many elements defined in RFC 4480 (Schulzrinne, H., Gurbani, V., Kyzivat, P., and J. Rosenberg, “RPID: Rich Presence Extensions to the Presence Information Data Format (PIDF),” July 2006.) [RFC4480] are all discrete-valued. Some of these discrete-valued presence elements are a product of an underlying continuous value, others are created subjectively. For instance, the sphere element could be derived from the location of the presentity.

Discrete-valued data is easy to manipulate and use. However, the process by which discrete-valued data is determined varies. When discrete-valued data is derived from continuous variables, the process might require heuristics that could be incompatible with the needs of all watchers. Alternatively, where the process is subjective, active user input is required.

Derivation of discrete-valued data from continuous primitives places assumptions and restrictions on the resulting value. Access to uninterpreted continuous values avoids any limitations that these might impose and can enable a wider range of applications.

2.2. Measurement

The act of determining a continuous value is an act of measurement. The value that is represented in a presence document is the product of a measurement process. All measurement processes are subject to

This document applies to the inclusion of continuous-valued data in presence, with any of the following constraints:

1. The process of measurement has non-zero - or at least non-negligible - cost. Cost can be defined in any number of ways, but could include time, processing resources, network utilisation, labour or financial.

2. The measurement process has limited accuracy, resulting in uncertainty that is - or could be - significant.

Cost and accuracy are often related. Improvements in one result in degradation of the other. For continuous-valued information that is not subject to these constraints, some requirements from this document could still apply.

From the above constraints, one aim of any system that handles such continuous-valued data is to minimise both cost and uncertainty using different trade-offs. A common factor amongst systems that handle continuous-valued data is the desire to be able to ignore the effects of these constraints. Where ignorance is not viable, systems exhibit a range of traits: iterative measurement collection, caching of results, fuzzy logic, maximum likelihood estimation and other forms of statistical analysis.

A significant example of a continuously varying parameter is the position of an entity in space, or its location. Location information as an element of presence was established in [RFC4079] (Peterson, J., “A Presence Architecture for the Distribution of GEOPRIV Location Objects,” July 2005.) and a format specified in [RFC4119] (Peterson, J., “A Presence-based GEOPRIV Location Object Format,” December 2005.).

Uncertainty in location information, described in more detail in [I-D.thomson-geopriv-uncertainty] (Thomson, M. and J. Winterbottom, “Representation of Uncertainty and Confidence in PIDF-LO,” November 2009.), is a product of the method of location generation used and can vary greatly; likewise, the amount of time required to ascertain location within specific quality constraints is highly variable.

3. Logical Model: Presence Sources

The entity that measures continuous-valued data - and the process of measurement - are crucial in determining the resulting information. RFC 3856 (Rosenberg, J., “A Presence Event Package for the Session Initiation Protocol (SIP),” August 2004.) [RFC3856] notes that presence data can be sourced by a variety of means, but does not attach any
significance to the source of presence data. This document adopts a model that distinguishes between the PA and the presence source.

Figure 1: Logical Model

Specifics of the interaction between presence source and PA are out of scope for this document. There might be formalised protocols, or the PA might use unspecified or informal methods to acquire data. Formalised exchanges include SIP PUBLISH (Niemi, A., “Session Initiation Protocol (SIP) Extension for Event State Publication,” October 2004., [RFC3903]); informal interactions include those that are briefly described in Section 7 of [RFC3856] (Rosenberg, J., “A Presence Event Package for the Session Initiation Protocol (SIP),” August 2004.). For location information, the GEOPRIV architecture (Cuellar, J., Morris, J., Mulligan, D., Peterson, J., and J. Polk, “Geopriv Requirements,” February 2004., [RFC3693] defines the Location Generator. The Location Generator is the presence source of this model; the entity that generates, or measures, the continuous value. In the GEOPRIV architecture the Location Server is the analogue of the PA and the Location Recipient corresponds to the watcher. Other forms of continuous-valued data might have similarly formal architectures and nomenclature.

In an end-to-end system, the model can be applied iteratively. An entity that acts as a presence source to a PA could equally acquire its information from a separate source; from a different perspective, the PA becomes a watcher and the presence source becomes a PA.

Figure 2: Iterative Application of the Logical Model
3.1. The Presence Agent is a Cache

Caching is a tool that is widely used in protecting resources where multiple watchers require access to the same information. After a presence datum is acquired from a source, requests within a set period are served from a cache to reduce the time required to serve the request and to avoid the costs in gathering the data. For discrete-valued data, it is often feasible to provide data that is correct at the time that a watcher receives a notification. The caching behavior of the PA is effectively hidden. Continuous-valued data differs in that the information provided can never be absolutely correct. As the number of watchers increases, the extent to which a PA relies on caching necessarily increases. The PA provides a value from some time in the past, with a certain accuracy.

In presence, caching is also used by the watcher user agent, which maintains a cached presence document for each presentity. This cache is updated when the PA notifies the watcher of changes.

How useful cached information is depends on how the information is applied. This caching behavior can become apparent to watchers when the information that they receive is inadequate for their needs. Many caching management mechanisms, such as that defined in HTTP (Fielding, R., Gettys, J., Mogul, J., Nielsen, H., Masinter, L., Leach, P., Berners-Lee, T., and J. Reschke, “HTTP/1.1, part 6: Caching,” March 2010.) [I-D.ietf-httpbis-p6-cache], rely solely on age. As a metric with a single axis, time can be precisely controlled. There is no need to trade off multiple conflicting constraints, the data is either current enough, or not. Acquisition of continuous-valued data often involves quality and timeliness constraints in addition to age constraints.

3.2. The Role of the Presentity

As the subject of presence information, there are two main aspects to the presentities involvement in this model. The presentity can, in some circumstances, act as a presence source, even though the model does not require any direct involvement. More significantly, in all cases the presentity has the most significant stake in ensuring that it retains some measure of control over the process. The presentity exercises control over the dissemination of presence information through use of policy. Enforcement of policy is performed by
the PA. Policy is used to restrict what presence information is made available to watchers. This document describes requirements for policy that also limit what controls the watcher is given over the measurement of continuous-valued presence information.

Note: The GEOPRIV model assigns the label of "Rule Maker" to the role assumed by the presentity when setting policy. The Rule Maker role is further extended to include other stakeholders. The requirements in this document do not preclude this abstraction.

In some cases, a presentity also acts as a presence source. A presentity could exploit this confluence by restricting the information that it provides as a presence source, thereby obviating the need for certain aspects of policy. This can be an advantageous circumstance, particularly where the PA limits the complexity and size of the policy it accepts.

4. Problem Statement

Continuous-valued data cannot always be treated in the same fashion as discrete-valued data. This section outlines a number of considerations that distinguish continuous-valued data. These considerations affect how the presence system uses continuous-valued data.

4.1. Unguided Measurement

A presence source is responsible for the measurement of presence data. Unless the watcher is able to communicate preferences to the presence source, the process of measurement is unguided. In the event that a PA and watcher operate independently of the presence data source, there is a risk that continuous-valued data is measured inappropriately. Values might be measured too frequently or with unneeded degree of accuracy; that is, accuracy is favored too highly over cost. Conversely, if cost is optimised, measurements might be insufficient in frequency or accuracy. Communicating watcher preferences to the presence source addresses this problem by making watcher preferences known to the presence source. With this information the measurement process is no longer unguided. Communicating watcher preferences to the presence source is especially critical if the watcher also incurs costs relating to the measuring of the continuous value. The most direct cost is in time: the watcher might be forced to wait for measuring of information to an accuracy more stringent than needed. However, it need not be limited to time.
4.2. Presence Filters

It is possible to resolve the problem of unguided measurement without resorting to explicit protocol mechanisms. Presence filters ([RFC4660] (Khartabil, H., Leppanen, E., Lonnfors, M., and J. Costa-Requena, “Functional Description of Event Notification Filtering,” September 2006.), [RFC4661] (Khartabil, H., Leppanen, E., Lonnfors, M., and J. Costa-Requena, “An Extensible Markup Language (XML)-Based Format for Event Notification Filtering,” September 2006.)) are a vessel for conveying a watcher's preferences to the PA. If the PA has a means of communicating with a presence source, it is able to request measurement of continuous-valued data according to the preferences it is aware of. By communicating the watcher's preferences in this manner, the presence source is able to measure accordingly. Additional changes to presence subscriptions ([I-D.ietf-sipcore-event-rate-control] (Niemi, A., Kiss, K., and S. Loreto, “Session Initiation Protocol (SIP) Event Notification Extension for Notification Rate Control,” February 2010.)) and filter extensions ([I-D.ietf-geopriv-loc-filters] (Mahy, R., Rosen, B., and H. Tschofenig, “Filtering Location Notifications in the Session Initiation Protocol (SIP),” March 2010.)) make more information available to the PA; this information could be passed on.

Figure 3 (Presence Agent Model) shows a simplified model of the existing system of filtering for presence data. The system assumes that presence information is perfect. If this is not the case, the only option for a PA is to project the illusion that this is the case. The only information available to the PA is the filter and throttle data made available to it in the subscription. This information is then used to advise presence sources and to shape the notifications the PA provides.
4.3. Quality

Ensuring that a watcher is able to access information that meets their needs is an important goal of presence filters. When it comes to information quality, there are two separate goals that need to be distinguished:

*A filter is used to determine when (and if) the PA notifies the watcher of changed presence information. This filter is usually
strictly boolean, or "hard": unless the condition is met, no notification is generated.

*A target quality is used to inform the presence source (the measurement process) of the desired quality of the result. This target is "soft": if the stated quality cannot be provided, a notification might still be generated. In this way, this information acts as a goal for the presence source, rather than a condition on the notification process.

The difference between the soft target and hard filter could alternatively be characterised as "want" and "need". The concept of a soft target is not necessary for discrete-valued presence data.

Suppose a watcher, Bob, is watching the location of a presentity, Alice. Bob specifies a hard filter that specifies a maximum uncertainty of 50m. If this is used to advise the presence source, the source attempts to determine a location within that uncertainty. If the source is unable to meet the requirement, the filter ensures that no notification is sent to Bob.

However, Bob could handle uncertainty of up to 1000m; it's not ideal, but some information is better than none. If he specified a larger constraint using the hard filter, the source might use cheaper methods that never produce a result with the accuracy Bob really wants. Therefore, Bob provides two values: a soft target of 50m, and a hard cutoff at 1000m uncertainty.

From this example, it can be seen that soft quality preferences are useful where continuous-valued data is the subject. In comparison, the hard filter is not tolerant of variations in quality. The soft preferences are used to guide the measurement process, but aren't restrictive and don't unnecessarily filter out poor quality results.

Location-specific standards such as the Mobile Location Protocol (MLP) ([OMA.MLS]) provide a means to specify either hard or soft quality constraints. A service class of "assured" implies that constraints are hard, whereas "best effort" service class indicates the use of a soft constraint.

Quality constraints can be used to bypass caching of data. For instance, a constraint might be placed on the timestamp that forces the measurement of new data. This is discussed in more detail in Section 4.6.1 (Immediate Triggering of Measurement).

4.4. Watcher Feedback

A watcher requires adequate feedback from the PA about how its preferences are being acted upon. To a watcher, a PA that fully communicates watcher preferences to a presence source, remains indistinguishable from a PA that is either less able to acquire data (because it relies on PUBLISH (Niemi, A., “Session Initiation Protocol (SIP) Extension for Event State Publication,” October 2004.) [RFC3903]), or simply chooses not to perform this communication. In some cases, a watcher is able to compensate for any limitations imposed by the PA. As shown in Section 4.6.3 (Value-based Triggering of Measurement) adaptive polling can be used to seek a particular value. However, if a watcher uses polling for value-seeking, this potentially duplicates similar efforts from the PA or presence source. In other circumstances, shortcomings cannot be addressed by the watcher. For instance, perhaps the presence source is capable of providing information to a certain high degree of accuracy, but unless requested it does not do so (the cost is too high). If the PA never requests highly accurate information, that information is never available to the watcher.

Providing the watcher feedback on what efforts the PA and presence source make to acquire information ensures that the watcher is able to avoid replicating behaviour. It also ensures that the watcher is able to make requests with some assurance that the PA is attempting to address them.

4.5. Active and Passive Sources

For a particular presence datum, the means by which the PA is able to acquire updated information varies. For some forms of discrete-valued data - such as basic status - the PA might be considered absolutely authoritative, in that the information is absolutely correct by definition. Data can be made available to the PA either passively or actively:

**Passive Acquisition:** The PA is a passive recipient of updated information with no means to trigger measurement. SIP PUBLISH (Niemi, A., “Session Initiation Protocol (SIP) Extension for Event State Publication,” October 2004.) [RFC3903] is an example of how the PA could passively acquire data.

**Active Acquisition:** The PA is able to seek updated information, and does so based on one or more triggers.

Providing the watcher information on the nature of a particular presence source is an important part of the feedback provided by the PA. If the PA only passively acquires data, much of the considerations presented in
this document become moot. More significantly, if the watcher is not able to influence how the PA interacts with the presence source, the watcher is not able to specify constraints on the data it receives. Thus, only when the PA conveys watcher preferences, does active acquisition becomes truly effective in improving the usefulness of presence data.

Back to Bob and Alice: This time, Bob wants to know when Alice passes by the library where Bob works. Bob sets a filter, requesting that he only be notified when Alice passes within 100m of the library. If the PA passively receives updates about Alice's location from a presence source and does not inform the source of Bob's desired result, it could be that it never receives information when Alice passes by. The presence source might not publish the necessary information at the moment that Alice passes by and Bob is not informed even though the event he was interested in actually occurs.

4.6. Triggering Measurement

Stimulating the generation of a value for the continuous variable can be done through several methods:

* immediate or on-demand triggers

* time-based or periodic triggers (polling)

* value-based triggers

4.6.1. Immediate Triggering of Measurement

An immediate request for measurement provides the simplest means of directly providing the presence source with necessary information. A watcher makes a request to the PA, including its preferences. The PA then acquires information from the presence source, passing watcher preferences directly to the source.

An immediate request for updated information is particularly useful for suppressing caching behavior at the PA or other entities. PA caching can interfere with certain use cases, particularly those where currency of information is valued over efficiency. For instance, if a watcher wants to know what the ocean temperature off the coast of Madagascar is now; if cached data is too old to useful, a caching
mechanism could interfere with that goal. Measures to suppress and control caching are necessary.

The simplest mechanism for avoiding cached data is to specify a quality constraint that limits the age of information. Setting this age value to zero indicates to any caching entity that old information is no good. As long as the caching policy allows it, the cache is avoided. Values larger than zero indicate how old the cached information can be to be used. This ensures that the benefits of caching are achieved without negative affecting the watcher.

The HTTP Cache-Control header [I-D.ietf-httpbis-p6-cache] (Fielding, R., Gettys, J., Mogul, J., Nielsen, H., Masinter, L., Leach, P., Berners-Lee, T., and J. Reschke, “HTTP/1.1, part 6: Caching,” March 2010.) provides this function. A requester is given explicit control over whether a cache is used. Cache-Control directives relate to the age of the information; a header set with no-cache prevents the use of data from cache.

4.6.1.1. Immediate SIP SUBSCRIBE

For SIP presence, the means of triggering immediate measurement is the SUBSCRIBE request. This presents a challenge for the SIP presence framework if the presence source is not able to provide the requested information immediately. RFC 3265 (Roach, A., “Session Initiation Protocol (SIP)-Specific Event Notification,” June 2002.) [RFC3265] states:

When a SUBSCRIBE request is answered with a 200-class response, the notifier MUST immediately construct and send a NOTIFY request to the subscriber.

If the PA is unable to provide an immediate notification due to lack of information, it must indicate this to the watcher. A presence document might be a composit of continuous- and discrete-valued data. Any solution for continuous-valued data cannot affect the conveyance of discrete-valued data. Compatibility with existing the existing presence framework is also desirable. Therefore, an immediate notification is still necessary.

Some time after the initial request, when the continuous-valued data becomes available, a second notification can be sent. However, another conflict arises with event rate throttling (Niemi, A., Kiss, K., and S. Loreto, “Session Initiation Protocol (SIP) Event Notification Extension for Notification Rate Control,” February 2010.) [I-D.ietf-sipcore-event-rate-control]. When information becomes available, the notification might be suppressed due to a short elapsed time since the initial notification.
Any means of triggering immediate measurement needs to consider these problems.

4.6.2. Periodic Triggering of Measurement

Periodic notifications of the value of presence data are not assumed to be necessary by the existing work. Values are assumed to change infrequently, or if frequent changes are necessary, solutions have concentrated on means of throttling notifications. RFC 3265 (Roach, A., “Session Initiation Protocol (SIP)-Specific Event Notification,” June 2002.) recommends that event packages specify a minimum interval between notifications; and [I-D.ietf-sipcore-event-rate-control] (Niemi, A., Kiss, K., and S. Loreto, “Session Initiation Protocol (SIP) Event Notification Extension for Notification Rate Control,” February 2010.) provides a direct means of controlling this interval.

If the watcher has an interest in the value of a particular datum over time, it is useful to be able to have the PA provide notifications at regular intervals.

A similar time-based problem exists for periodic notification as did for the immediate request. The PA cannot request presence information at the time that a notification is required - the presence source is unlikely to be able to produce a result so quickly. The PA needs to allow the presence source sufficient time to make a measurement. This time can be allocated before the notification is required so that all the necessary data is available for the notification.

Deciding how much time to allow the presence source for measurement presents another problem. Relying upon a value specified by the watcher ensures that the measurement process is guided appropriately.

4.6.3. Value-based Triggering of Measurement

A watcher might only be interested in a value if that value enters a particular range. For instance, for location, the watcher might be interested when the presentity is in the vicinity of a particular landmark, or even when two presentities approach each other. A watcher that monitors the temperature of a presentity might be interested if the temperature exceeds a certain threshold.

Value-seeking can be performed in different ways. Given certain assumptions or knowledge about the rate of change of a value and the desired responsiveness, an adaptive polling method can be used. The rate of polling can increase in frequency in proportion to the proximity of the current value to the target range. In addition, time and quality constraints can be relaxed or made more stringent as appropriate. The
advantage of polling is that it can be performed based on the value only. The additional knowledge available to a presence source is not necessary; a PA or watcher is able to use polling for value-seeking. A presence source can potentially use alternative information to assist in value-seeking. For data that is derived from other measurements, the value of the unprocessed measurements might provide an adequate indication of the value to obviate any need to continue measurement. For instance, the set of wireless transmitters that can be observed by a wireless device can be used as a rough indication of location. Value-seeking can be of great benefit to applications, particularly for simplifying the application logic. Without this feature, an application needs accept and interpret notifications and check values against a target range. As demonstrated in [I-D.thomson-geopriv-uncertainty] (Thomson, M. and J. Winterbottom, “Representation of Uncertainty and Confidence in PIDF-LO,” November 2009.), this can be non-trivial. If a trigger can be set based on a certain value, the notification provided by the PA indicates that the condition has been met.

Lee is a surfer. He uses a presence service to inform him of when the waves are good at his favourite beaches. During the week, when Lee is usually at work, Lee sets a trigger that notifies him when the waves reach a certain size at any one of these beaches. If the waves reach the size that means that surfing becomes a higher priority than working, Lee is sent a text message. Upon receiving the text message, he drops what he is doing to catch the big surf.

This case can also be used to demonstrate one reason for using continuous-valued data instead of discrete-valued data. Harriette might also be an equally keen surfer, but less flexible employment. She has to set a more stringent set of criteria before she is willing to renege on her responsibilities. Giving both surfers access to the raw data lets them set their own definition for what is worth skipping out on work. Any discrete value that is assigned reflects a certain set of values that cannot be universally accurate.

5. Requirements

The requirements documented in this section relate to subscriptions for continuous-valued data. These requirements do not make any assumptions about the nature of the relationship between PA and presence source. A presence source is assumed to be present only to the extent that the measurement of a particular presence element affects the results observed by the watcher.
A solution that addresses these requirements MAY address them based on a specific type of data. A generic solution might not be feasible in all cases.

5.1. Quality Requirements

**Q1.** The system MUST provide a watcher the ability to express non-binding requirements on information quality for continuous-valued presence data.
Motivation: Having a means to express preferences in non-binding fashion can have the desired effect in influencing the measurement process at the presence source without other side-effects.

**Q2.** The system MUST provide a means to indicate how information quality preferences are used or propagated.
Motivation: Providing adequate feedback to a watcher assists the watcher in making decisions about its behaviour. From the perspective of a PA, this requirement can only be fulfilled based on the knowledge it has available. A PA can know whether or not it is the presence source, but it cannot assume that the entity it requests information from is the ultimate source of the information; a PA needs to - in turn - rely on the information provided by its source. This ensures that a watcher is able to learn how its preferences are getting to the ultimate source.

For instance, the PA or presence source might limit the extent to which quality parameters are able to bypass caches. This information can be integrated with any inherent limitations indicated by the source, or the mechanism used to acquire the presence data.

5.2. Immediate Triggering Requirements

**IT1.** The system MUST provide a means for a watcher to explicitly and immediately request measurement of a specified continuous-valued datum, bypassing any cached data.
Motivation: A direct request for data ensures that the presence source is properly advised of the constraints of a request. It also ensures that a watcher has access to updated information.

**IT2.** The system MUST provide a means for the watcher to communicate time constraints on measurement processes.
Motivation: Specifying time constraints ensures that the information is available to the watcher when required. It also ensures that the presence source applies appropriate methods in measuring.

**IT3.** The system MUST provide a means for a PA to limit support for immediate requests.

Motivation: Making a direct request to the presence source increases the load on the presence source and PA. A PA implementation might want to constrain access to immediate requests to prevent denial of service.

**IT4.** The system MUST provide a means for a PA to indicate support for immediate requests, including any limits or constraints on that support.

Motivation: Providing adequate feedback to the watcher ensures that the watcher is able to modify its behaviour accordingly.

### 5.3. Periodic Triggering Requirements


**PT1.** The system MUST provide a means to request periodic measurement of a continuous-valued datum at a specified interval.

Motivation: Periodic measurement is the most basic means of enabling tracking of a value over time. More sophisticated methods might be used, but this sets a minimum level of capability that can be exploited for any type of continuous-valued data.

**PT2.** The system MUST provide a means for a PA to limit support for periodic requests.

Motivation: Excessive polling rates might result in denial of service through excessive requests.

**PT3.** The system MUST provide a means for a PA to indicate support for periodic measurement, including any limits or constraints.

**PT4.** The system MUST provide a means for a watcher to explicitly limit the rate of notifications.

Motivation: Even without perfect information, the defining characteristic of continuous variables is the potential for continuous change. This manifests in presence as a continuous
stream of data from the PA to the watcher. For a range of reasons - protection of network utilization not-withstanding - this is not desirable.

5.4. Value-Seeking Requirements

VT1. The system MAY provide a means for a watcher to indicate a particular range of values to seek.
Motivation: If a watcher is only interested in a certain range of values, limiting measurement and notification protects resources on both the presence source and watcher.

VT2. The system MUST provide a means for a PA to indicate support for value-seeking requests, including any limits or constraints on that support.
Motivation: Providing adequate feedback to the watcher ensures that the watcher is able to modify its behaviour accordingly.

5.5. Timeliness Requirements

T1. Continuous-valued data MUST always have a timestamp that indicates when the value was measured.
Motivation: A continuously varying datum is, by its nature, inevitably out of date at the time that the watcher receives it. In practice, this is only a problem if the time between measurement and receipt of the data is large. Since this is a matter of degree, providing a timestamp ensures that a watcher is able to make a judgment about validity.

If time information is not included, the recipient of continuous-valued information has no means of judging how current the information is. PIDF (Sugano, H., Fujimoto, S., Klyne, G., Bateman, A., Carr, W., and J. Peterson, “Presence Information Data Format (PIDF),” August 2004.) [RFC3863] specifies an optional timestamp element, which is described as being necessary if available. This requirement makes the value mandatory.

The corollary to this that an item of continuous-valued data is automatically invalid if it is not timestamped.

T2. The system MUST provide a watcher the ability to limit the age of data that it is provided.
Motivation: This ensures that caching of data is only used to the extent that it is acceptable to the watcher.


T3. The system SHOULD provide a means for a watcher to indicate how long measurement of continuous-valued data is allowed to take.

Motivation: This indication assists the PA and presence source in making decisions about how much time to allocate to measurement; for periodic measurement, it provides a hint on when to start measurement. This guidance is often better served by other quality constraints in situations where the delay is not evident to the watcher. However, a time-based value serves in a wide range of scenarios and guides behaviour where a PA might not understand the quality constraints specific to a data type.

5.6. Privacy Requirements

P1. The system MUST provide a means for a presentity (rule maker) to control how watchers are able to request continuous-valued data.

Motivation: A presence agent acts for the presentity in disseminating presence information. Policy is the primary means of privacy control available a presentity.

Many of the requirements in this document describe more flexible means of acquiring presence information. Policy provides a means for a presentity to limit a watcher's access to these features and thereby restrict access to private data.


Each of the individual aspects of watcher-agent-source interaction are appropriate subjects for policy:

Quality: A presentity might want to limit the quality of the information provided. Quality can be purposefully degraded post-
Immediate Triggers: A presentity might want to restrict access to immediate requests, or force the use of cached data.

Periodic Triggers: A presentity might want to prevent periodic requesting, or place a limit on the rate that can be requested by watchers.

Each of the mechanisms that allow for greater control at the watcher MUST also consider policy controls to limit or block use of that mechanism. This guarantees that a presentity (or rule maker) is able to communicate appropriate rules to the PA.

6. Security Considerations

A number of requirements involve a PA making information about its operation available to a watcher. This information might be consider sensitive by the operators of a PA. Any solution that addresses these requirements MUST provide an option to suppress this information and MAY include guidance on when this might be appropriate.

Many of the requirements in this document potentially result in PA and presence source behaviour being controlled to some extent by a watcher. A malicious watcher could attempt a denial of service attack by exploiting the fact that some of this behaviour requires greater expenditure resources from the PA and presence source than is incurred in making the request. In particular, periodic requests or requests that includes value-seeking require significant processing in response to a relatively simple request. Any solution that addresses these requirements MUST consider the implications of denial of service on the PA and SHOULD also consider the presence source.

Caching of results can limit the load imposed by multiple requests for the same information. To protect against denial of service, a PA MAY choose to suppress any options for bypassing cached data.

Privacy is a major consideration in any architecture that defines more flexible means of access to private data. Section 5.6 (Privacy Requirements) describes requirements that are aimed at providing a presentity (or other rule maker) greater control over the access watchers have to private data. These focus on the application of policies that direct the actions of the PA.
7. Acknowledgements

This document is a clumsy attempt to formalize the output of discussions on the nature of presence and its applicability to location information. Adam Roach has been helpful in destroying misconceptions about presence. Jonathan Rosenberg pointed out a major omission in ignoring the role of the presentity an earlier version of this document. Thanks also to Richard Barnes, Jon Peterson, Robert Sparks, James Winterbottom.

8. References

8.1. Normative References

| RFC2119 | Bradner, S., “Key words for use in RFCs to Indicate Requirement Levels,” BCP 14, RFC 2119, March 1997 (TXT, HTML, XML). |

8.2. Informative References

<table>
<thead>
<tr>
<th>Reference</th>
<th>Title</th>
<th>Status</th>
<th>Date</th>
<th>Link</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Thomson, M. and J. Winterbottom, “Specifying Location Quality Requirements in Location</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix A. Presence Agent and Source Interactions

[[Ed: this section has some informational value, but need to determine how much that contributes to the document.]]

This document specifies requirements that expose the nature of the association between PA and presence source. While there are no direct requirements on this interface and proprietary solutions are entirely appropriate, it is hoped that these requirements will influence the design of any protocol mechanisms used on this interface.

Associations between PA and presence sources could be largely static in nature, as is true of the methods described in [RFC3856] (Rosenberg, J., “A Presence Event Package for the Session Initiation Protocol (SIP),” August 2004.). Establishing dynamic associations between PA and presence source is an option where disparate presence sources are required. This is especially true for location information, where close physical proximity between presentity and source is usually required; consequently the presence source can dynamically change as the presentity moves.


[I-D.ietf-geopriv-lbyr-requirements] provides a means whereby a relationship between any entity and the Location Generator can be established. By contacting the Location Generator, a requester is able to specify preferences for how location information is generated/measured. A location reference forms the basis for establishing an association between Location Generator and a Location Server.

For presence, Indirect publish (Garcia, M., Tschofenig, H., and H. Schulzrinne, “Indirect Presence Publication with the Session Initiation Protocol (SIP),” March 2009.)

[I-D.garcia-simple-indirect-presence-publish] describes how a PA is able to use the indirection provided by a URI. The URI is used to establish a link between the presence source and the PA. The URI is distinguished by two characteristics:

*the host serving the URI is the presence source

*additional information in the URI provides enough information to uniquely identify the presentity to the presence source
## Author's Address

<table>
<thead>
<tr>
<th>Martin Thomson</th>
<th>TOC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Andrew</td>
<td></td>
</tr>
<tr>
<td>PO Box U40</td>
<td></td>
</tr>
<tr>
<td>Wollongong University Campus, NSW 2500</td>
<td></td>
</tr>
<tr>
<td>AU</td>
<td></td>
</tr>
<tr>
<td>Phone: +61 2 4221 2915</td>
<td></td>
</tr>
<tr>
<td>Email: <a href="mailto:martin.thomson@andrew.com">martin.thomson@andrew.com</a></td>
<td></td>
</tr>
<tr>
<td>URI: <a href="http://www.andrew.com/">http://www.andrew.com/</a></td>
<td></td>
</tr>
</tbody>
</table>