Conveying Location Data Using P3P
W3C Workshop on the Future of P3P

Helena Lindskog, Johan Hjelm, Mikael Nilsson
Ericsson AB
{Helena.Lindskog,Mikael.Nilsson}@ericsson.com
Johan.Hjelm@era.ericsson.se

Introduction

The location of an individual or a mobile terminal can be sensitive with regard to the privacy of an individual. Awarding the individual self-determination of the way position information is used will be crucial to the acceptance of the technology. This paper suggests a solution to how location data can be conveyed in privacy-friendly way using P3P combined with the PiMI method for data conveyance, first described in [Nilsson et al. 2001], and further developed in [Lindskog et al. 2002].

P3P and Data Conveyance

With the PiMI method, first described in 2001, P3P is used for data conveyance. The method was first developed to meet the need of transmitting device capabilities and user preferences in mobile Internet applications to origin servers, but suites the legacy Internet just as well.

The basic concept is that a specific data schema [P3P & UAProf] is used to describe which device capabilities an application needs. When the device has matched the policy file with the user preferences, the device capabilities are sent with the request for the resource using UAProf, which is based on CC/PP. Thus, the policy file is used to request data.

The data does not need to be device capabilities, but could also be for example the user's contact information, besides user-agent information, like browser-name or Bluetooth version. It might also be preferences like "I don't want images, just give me the text". All this information can be baked into the HTTP header and transmitted with the request, using CC/PP.

Look at this scenario:

- Greta has her name and address stored in a profile.
- Greta goes to werespectyou.com to order a violin.
- Greta's user-agent fetches their privacy policy.
- It is OK to surf around—no alert.
- Greta goes to the specific site where they ask for her name, address and credit card number.
- Greta gets a warning from her user-agent, stating exactly what they want, for what purpose, and their retention policy and all the other stuff that a policy can tell you (chapter 9 again).
- Greta knows why they are asking—she just ordered—and now she learned about their policy, and she then presses OK.
- Within the request header comes all the information needed:

```
<?xml version="1.0"?>
<RDF xmlns="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:prf="http://www.mysite.org/Buddies/ccppschema-20020727#">
        <rdf:Description ID="Profile">
                <prf:component>
                        <rdf:Description ID="UserData">
                                <prf:Gender>Female</prf:Gender>
                                <prf:Age>27</prf:Age>
                                <prf:Languages>
                                        <rdf:Bag>
                                                <rdf:li>Swedish</rdf:li>
                                                <rdf:li>English</rdf:li>
                                                <rdf:li>German</rdf:li>
                                                <rdf:li>Russian</rdf:li>
                                                <rdf:li>Spanish</rdf:li>
                                        </rdf:Bag>
                                </prf:Languages>
                                <prf:Name>Greta</prf:Name>
                                <prf:Location>Stockholm</prf:Location>
                                <prf:JobTitle>Warrants-
Consultant</prf:JobTitle>
                        </rdf:Description>
                </prf:component>
        </rdf:Description>
</RDF>
```

- The server-side application gets her request at the other end notices that “hey, this is a data-conveyance-compliant user-agent at the other end”, and accepts the data without asking her to fill in a form—alternatively pre-fills in the form, and sends it back for acceptance.

Data conveyance means convenience for the user. Such convenience is always a benefit, but it is especially important in mobile environments.

Location

Location, the position of a user (or rather, the terminal of a user), can be determined in several ways (for an introduction, see [Hjelm 2002]. In brief, the location determination is dependent on sensing the position of a piece of equipment, which can either be dedicated (as in GPS) or a general communications terminal (such as a mobile phone). Having sensed the location of the user, and related it either to the network or a global reference system (i.e. the methods used in mobile-phone based network positioning and GPS), it can be expressed in a number of coordinate systems (which are immaterial to this discussion).

To be usable in an information service context, the position has to be communicated to the point where information is filtered or assembled (often from a number of different sources). This can be done in the mobile terminal (GPS terminals often have a map database built in), but this either requires that the data set be static (as in GPS terminals), or that the data set communicated is small (due to the limited bandwidth available to transmit information to most mobile equipment). In practice, the filtering and assembly of information is done in a server in the network, which also can be the same server that handles all or some of the databases to be used in the creation of the information service.

Location-dependent information services constitute a special case of context-dependent services, where a set of parameters relating to the users situation are used in creating a specially adapted piece of information that relates to the specific user in that specific situation. An example is services which create a map which relates the users current situation to the desired situation of the user, according to some set of rules; such as "WHEN 'I am on vacation' AND 'the sun is shining' ARE 'true', THEN 'show me the way to the beach'". For an introduction and further about location dependent services creation and maintenance, see [Hjelm 2002].

The creation of this type of services requires that the position of the terminal be communicated to the server, for it to be usable as a parameter in the filtering or construction of information. This is where the privacy risk occurs. If all possible information is communicated to the terminal, and all filtering and assembly done there, what happens to the information is opaque to the entity maintaining the service. However, this is impractical in reality, as mentioned. If registered, the position of the user (especially if identifiable and registered over time) in combination with the service (implicitly) requested will enable profiling of the user and thus increase the risk for privacy violations.

The information set describing the users position has been standardized by the Location Interoperability Forum (for an introduction, see [Hjelm 2002]). The information is defined as an XML document, and the assumption is that the actual information set will be handled at the application layer of the OSI stack, like the information constituting the service. An intermediary, the Mobile Positioning Center, is required to take the data describing the users position in the network or GPS receiver, and put it into a format which is usable by applications, e.g. database management systems.

The GEOPRIV Working Group

If information is assembled from many different sources, and the user can be identified as the same unique user, a very detailed demographic profile can be created over time, which implies that there is a risk of significant privacy violations. One of the favourite use cases of the location information industry is positioned push advertising, where advertisements would be pushed to your mobile phone when you walk past a storefront.

The risk of privacy violation has been of concern to the location information industry and regulatory bodies for some time. It has also been the focus for work in the Location Interoperability Forum, as well as the GEOPRIV working group in the IETF.

Originally founded to create a generic positioning mechanism for the network layer, the IETF GEOPRIV working group has changed its charter to focus on the privacy issues. Still assuming that a (as yet undefined) object containing the position information will be communicated (through some undefined means) to the server creating the service, the working group have created a requirements document, which takes some initial steps towards describing the mechanisms. However, the problem defined as the work item is easier to handle in the application layer, using the LIF and P3P protocols, instead of at the network layer.

In addition, the group has discussed some alternate methods to mask the users position, e.g. by generating a number of spurious positions from which a request could have occurred, only one of them being the users actual position. This "privacy by obscurity" in position dependent systems is a fallacy, and based on the premise that the user is only ever going to make a request to the information provider, and only that user ever will make that request.

In a system where the user hides his position by generating a number of spurious addresses to hide his real position, providing a set of addresses to the service, saying "I am at one of these". Since, however, the probability that the user is in one location (e.g. in a street corner) is higher than another (e.g. in the middle of a lake), this probability can be used to deduce where the user is, especially since it is likely that all users making requests for information pertaining to a location (e.g. directions) will be making requests from a limited set of positions; the actual positions of the users making the request will then form a pattern, whereas the spurious positions will be

distributed at random, and this means that if there is a desire to violate the users privacy by knowing his position, that can easily be determined.

Setting up Policies around Location

An alternative method to assuming that the position information is communicated at the network layer using some obscure and undefined method is to assume that it is communicated at the applications layer, e.g. as part of the users profile using the PIMI method when the user requests an information service. This implies that the service can accept a statement about the policy of the user concerning his privacy, e.g. if expressed as P3P, and handle that in the same way as any policy concerning any request.

A statement about privacy practices for location data can be made in two ways. One is by very generally stating that location data is used and needed by the web site, by using the <location/> attribute, as in the example below:

```
<STATEMENT>
 <PURPOSE><current/></PURPOSE>
 <RECIPIENT><ours/></RECIPIENT>
 <RETENTION><stated-purpose/></RETENTION>
 <DATA-GROUP>
  <DATA ref="#dynamic.miscdata" optional="no">
   <CATEGORIES>
    <location/>
   </CATEGORIES>
  </DATA>
 </DATA-GROUP>
</STATEMENT>
```

The other option is to use a specific schema. We suggest that the LIF structure be used as a basis for such a schema. This is not necessary if the policy is not to be used for data retrieval, or if the location in X and Y coordinates is the only thing needed.

However, in cases where other attributes are needed, a privacy statement can be used. Such attributes can for instance be the uncertainty area of the users position, which can be defined in the LIF document as an area within which the user is with some degree of certainty; e.g. “my position is in the circle surrounding this point with a certainity of 60 %”. In those cases, a statement like the following can be used.

```
<STATEMENT>
 <PURPOSE><current/></PURPOSE>
 <RECIPIENT><ours/></RECIPIENT>
 <RETENTION><stated-purpose/></RETENTION>
 <DATA-GROUP>
  <DATA ref="#lif.shape"><DATA>
  <DATA ref="#lif.time"></DATA>
 </DATA-GROUP>
</STATEMENT>
```

User-Agent Response

If the privacy statements of the origin server corresponds to the preferences of the user, the user-agent will respond with a structure containing either a simple X and Y coordinate structure, or else the elements requested for through the suggested data schema.

The simple response expressing only the coordinates of the user would look like this (the "msid" is the identifier of the terminal, in this case the telephone number; this can be replaced by any identifier):

```
<slia ver="3.0.0" >
   <pos>
   <msid="461011334411">
          <pd>
               <coord>
                    <X>301628.312</X>
                    <Y>451533.431</Y>
               </coord>
          <pd>
     </pos>
</slia>
```

This is hardly meaningful, however, since it cannot be assumed that the user is at this specific point, but somewhere near it. Also, there is no time parameter, thus making it impossible to know if this was a recent position, or if it relates to the users position as of a week ago. If more accuracy is required, an area has to be declared for the position, and a time to enable freshness. Assuming the user has made the declaration that he will give out the time and shape, as in the example above, and the server has declared that it will not misuse it, the users response will include the shape and time, as follows (this is actually the normal declaration):

```
<slia ver="3.0.0" >
   <pos>
   <msid="461011334411">
          <pd>
               <time utc_off="+0200">20020623134453</time>
               <shape>
                    <CircularArea srsName="www.epsg.org#4004">
                         <coord>
                              <X>301628.312</X>
                              <Y>451533.431</Y>
                         </coord>
                    <radius>240</radius>
               </shape>
          <pd>
     </pos>
</slia>
```

About the Method

It is obvious that the method does not offer a complete solution to the problem with user trust in location-based services. There need to be more mechanisms at place – such as a method for guaranteeing that the service provider is trustworthy.

However, when used in combination with other PETs, this method is a very good way of achieving data conveyance while adhering to the user's privacy preferences.

References

[Hjelm 2002] Creating Location Services for the Wireless Web. John Wiley & Sons, New York, 2002.

[Lindskog et al. Taking Charge of Profile Information Conveyance. Proceedings of e2002.

2002]

[Nilsson et al. 2001] M. Nilsson, H. Lindskog, S. Fischer-Hübner. "Privacy Enhancement in the Mobile Internet". *In Proceedings of Security and Control of IT in Society-II, IFIP SCITS-II*. Bratislava, Slovakia, June 15-16, 2001. http://privacy.lindskog.ws/pimi.pdf

[P3P & UAProf] H. Lindskog. "A Sample P3P Policy for UAProf". May 2001. http://privacy.lindskog.ws/p3p_policy4uaprof.html