

RFC 7459 : Representation of Uncertainty and Confidence in PIDF-LO

Stéphane Bortzmeyer

<stephane+blog@bortzmeyer.org>

Première rédaction de cet article le 27 février 2015

Date de publication du RFC : Février 2015

<https://www.bortzmeyer.org/7459.html>

Le format PIDF-LO (*"Presence Information Data Format - Location Object"*, RFC 4119¹) permet de représenter la position présente d'un objet mobile. Ce nouveau RFC est consacré à la représentation de l'**incertitude** sur la position de cet objet, et ajoute un moyen d'indiquer la **confiance** qu'on a sur la position de cet objet. Ce RFC fournit également des indications sur le comportement à avoir lorsqu'on reçoit de l'information imprécise ou incertaine.

Le cadre général sur le service de localisation est présenté dans le RFC 6280. Idéalement, la localisation de la cible (oui, le RFC utilise le terme de *"target"*, ce qui est un peu effrayant) est connue parfaitement. Dans la réalité, on n'a pas une précision infinie, et on a même des doutes sur l'exactitude de la localisation. Il faut donc pouvoir représenter ce manque de précision et ces doutes, à la fois par un vocabulaire commun et par un format standard (une extension du PIDF-LO, *"Presence Information Data Format - Location Object"* du RFC 4119). Notez que l'IETF, dans ses RFC, ne normalise pas **comment** on arrive à déterminer une localisation. Il existe de nombreuses méthodes (triangulation entre les relais de téléphonie mobile, déclaration par la cible elle-même, lorsqu'elle est munie d'un GPS ou équivalent, etc) mais l'IETF ne se préoccupe que du format et de la distribution de cette information, pas de comment elle a été obtenue.

D'abord, revenons sur la notion d'**incertitude**, si importante en métrologie. La section 2 de notre RFC donne la définition suivante : « l'incertitude exprime de manière quantitative ce que l'on sait à propos d'une grandeur mesurée ». Par exemple, on sait que les instruments de mesure ont une précision limitée et on intègre cette imprécision dans l'incertitude. Ou bien on sait que la grandeur varie très vite et cette variation devient une partie de l'incertitude. Techniquement, l'incertitude est en général représentée par une distribution de probabilité. Celle-ci peut être trop complexe pour être manipulée simplement donc

1. Pour voir le RFC de numéro NNN, <https://www.ietf.org/rfc/rfcNNN.txt>, par exemple <https://www.ietf.org/rfc/rfc4119.txt>

on adopte en général un modèle simple : toute grandeur est accompagnée d'un intervalle et d'un degré de confiance. Par exemple, si je mesure une longueur, je vais l'indiquer « $L = 1,00742 \pm 0,0043$ m à 95 % de confiance » ce qui veut dire qu'il y a 95 chances sur 100 que L soit entre 1,00312 et 1,01172 mètres. (On note que, si la « vraie » mesure a une dimension - c'est un point, la mesure avec l'incertitude a deux ou trois dimensions, dessinant une région d'incertitude. Les formes de cette région peuvent être variées, cf. RFC 5491.) Si vous voulez en savoir plus sur l'incertitude, vous pouvez lire le « *Guide to the expression of uncertainty in measurement (GUM) 98 :1995* » de l'ISO ou la *Technical Note 1297*, « *Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Result* » du NIST.

On a dit plus haut que l'incertitude pouvait être exprimée par une distribution de probabilité. Celle-ci est représentée par une fonction, la PDF (*"probability density function"*) qui indique la probabilité, en chaque point, que la « vraie » valeur figure en ce point. Deux PDF sont très répandues, la gaussienne, aussi appelée distribution normale, qui provient en général de la combinaison de beaucoup de petites erreurs aléatoires, et la PDF rectangulaire, qui représente en général une mesure avec une source d'erreurs principale. La PDF est centrée sur la moyenne des mesures, la PDF gaussienne est caractérisée par l'écart-type, la rectangulaire par sa largeur (le RFC utilise la demi-largeur).

L'ancien RFC 3693 utilisait également les termes de précision et de résolution, qui sont désormais abandonnés, l'expérience ayant montré que leur définition ne suffisait pas en pratique. Par contre, comme alternative aux termes **quantitatifs** comme les caractéristiques de la PDF, notre nouveau RFC 7459 utilise le concept **qualitatif d'exactitude** (*"accuracy"*). L'exactitude d'une mesure indique qualitativement le degré de proximité avec la réalité. On dira ainsi qu'une mesure est « plus exacte » qu'une autre.

Après ces considérations générales sur l'incertitude, les sections 3 et 4 du RFC décrivent précisément comment celle-ci est représentée dans le format PIDF-LO (format fondé sur XML). La section 3 précise le modèle utilisé : la cible qui nous intéresse est considérée comme un point. (Ce n'est pas vrai si la cible est de grande taille, par exemple un aéroport, ou si elle est quantique, mais le modèle adopté par l'IETF est délibérément simplifié.) Le format PIDF-LO du RFC 4119 permet de décrire les incertitudes, et d'indiquer la forme de la région d'incertitude, en utilisant les formes géométriques standard de la norme « *GML 3.1.1 PIDF-LO Shape Application Schema for use by the Internet Engineering Task Force (IETF) - Candidate OpenGIS Implementation Specification 06-142r1* » <http://portal.opengeospatial.org/files/?artifact_id=21630> ». Par contre, le format PIDF-LO original ne permet pas d'indiquer la confiance, le RFC 5491 se contentant d'une valeur fixe de 95 %. À noter que l'absence d'incertitude dans le fichier PIDF-LO ne signifie pas que la mesure était parfaite : il se peut qu'on n'ait pas cherché à évaluer l'incertitude, ou bien qu'on la connaisse mais qu'on la cache délibérément, par exemple pour des raisons de protection de la vie privée. Notez enfin que le problème de l'incertitude existe aussi pour les localisations civiles (comme « 8, rue de la Pompe, 12345 Villeneuve ») du RFC 5139.

La section 4 de notre RFC étend l'élément `<location-info>` de PIDF-LO (format normalisé dans le RFC 4119) en ajoutant un élément `<confidence>`. La confiance est exprimée sous forme d'un pourcentage (comme le « 95 % » indiqué plus haut) ou de la valeur spéciale `unknown`. Les anciens logiciels ignoreront cet élément (qui n'existait pas au début de PIDF-LO) mais les nouveaux pourront s'en servir pour améliorer les traitements et les messages aux utilisateurs. Un attribut `pdf` indique l'apparence de la fonction PDF : inconnue, normale (gaussienne) ou rectangulaire. Voici un exemple :

```
<pdf:presence
  xmlns:pdf="urn:ietf:params:xml:ns:pdf"
  xmlns:con="urn:ietf:params:xml:ns:pdf:geopriv:conf"
  ... >
<dm:device id="sg89ab">
  <pdf:status>
    <gp:geopriv>
```

```
<gp:location-info>
  <gs:Circle srsName="urn:ogc:def:crs:EPSG::4326">
    <gml:pos>42.5463 -73.2512</gml:pos>
    <gs:radius uom="urn:ogc:def:uom:EPSG::9001">
      850.24
    </gs:radius>
  </gs:Circle>
<!-- NOUVEAU --> <con:confidence pdf="normal">67</con:confidence>
</gp:location-info>
</gp:geopriv>
</pidf:status>
</dm:device>
</pidf:presence>
```

Le schéma XML complet de ces extensions à PIDF-LO figure dans la section 7. L'espace de noms nécessaire, `urn:ietf:params:xml:ns:pidf:geopriv:conf` est désormais enregistré à l'IANA <<https://www.iana.org/assignments/xml-registry/xml-registry.xhtml#ns>>.

Avoir l'information ne suffit pas. Il faudra l'utiliser et on peut se demander, par exemple (section 4.3 du RFC) comment présenter une confiance à l'utilisateur, de manière à la fois techniquement correcte et compréhensible pour le non-spécialiste. Par exemple, si la confiance est indiquée de manière trop subtile, avec une légère variation de couleur, l'utilisateur pourrait croire que deux positions sont équivalentes alors que l'une d'elles est plus exacte.

La section 5 du RFC se penche ensuite sur les calculs effectués à partir de localisations comportant une incertitude. Elle est bien plus riche en mathématiques et sa lecture est plutôt pour celles et ceux qui sont à l'aise avec les centroïdes, la forme de la Terre, le calcul matriciel, et la géométrie en général. Ces calculs sont d'autant plus complexes que les opérations sur des données comportant une incertitude ne sont pas intuitives. Par exemple, l'incertitude sur la différence entre deux positions est la somme des incertitudes, pas l'incertitude maximale, encore moins la différence des incertitudes comme un calcul naïf risquerait de le faire. Les fanas de maths peuvent aussi lire les annexes A et B, qui contiennent des formules utiles.

Un petit mot sur la sécurité, pour finir (section 9 du RFC). Indiquer l'incertitude peut donner des indications sur le mécanisme de localisation utilisé, ce qui peut être vu comme excessivement bavard. Ceci dit, la section 13.5 du RFC 6772 rappelle que les services de localisation distribuent de l'information et qu'il est très difficile de diminuer les fuites (sans rendre le service inutilisable).