RFC 6280 : An Architecture for Location and Location Privacy in Internet Applications

Stéphane Bortzmeyer
<stephane+blog@bortzmeyer.org>

Première rédaction de cet article le 24 juillet 2011

Date de publication du RFC : Juillet 2011

https://www.bortzmeyer.org/6280.html

Aujourd'hui, de nombreuses applications des réseaux informatiques dépendent de la localisation physique de l'appareil mobile. La récolte de ces données de localisation (« le 31 mai 2011, à 08 :04, l'appareil était en 40,756° Nord et 73,982° Ouest, [Caractère Unicode non montré [1]]150m »), leur transmission à travers le réseau, et leur traitement par le service qui héberge l'application posent de graves questions de protection de la vie privée, surtout quand on lit les conditions d'utilisation de ces services (si on pouvait comprendre le jargon juridique, elles se résumeraient à « nous faisons ce que nous voulons, nous nous occupons de tout, ne vous inquiétez pas »). D'où ce RFC qui essaie de définir une **architecture** de gestion de la vie privée pour ces informations de localisation.

Un exemple d'une telle application est Pages Jaunes <https://market.android.com/details?id=com.pagesjaunes> sur Android : elle permet d'obtenir des réponses qui dépendent du lieu, tel que l'a indiqué le téléphone. Ainsi, si on tape « pizzeria », on obtient en haut du classement les pizzerias les plus proches physiquement. Résultat, les Pages Jaunes savent où vous êtes et on imagine facilement les conséquences pour la vie privée de ce genre de suivi (« L'utilisateur est allé quatre fois dans une clinique spécialisée cette semaine. »). Outre les Pages Jaunes, plusieurs intermédiaires connaissent également votre position (l'opérateur du réseau, qui sait par quelle borne vous vous connectez, les tiers qui écoutent, si le trafic n'est pas chiffré, et le téléphone lui-même qui peut garder cette information très longtemps, comme le fait l'iPhone <http://www.lemonde.fr/technologies/article/2011/04/21/les-iphone-collectent-l-historique-des-deplacements-des-usagers_1510753_651865.html>). À noter que cette information est collectée en permanence, sans action explicite de l'utilisateur (section 1.2). Dans les romans policiers, finies les longues filatures sous la pluie pour savoir où passe le suspect. Le cyber-policier moderne ne bouge plus de sa chaise et obtient via le réseau toutes les informations nécessaires, chaque citoyen emportant avec lui volontairement l'outil qui permet de le suivre à la trace.

1. Car trop difficile à faire afficher par LaTeX

Le RFC (section 1) ne prétend même pas qu'il va régler ce problème, qui est consubstantiel de la récolte et diffusion d'informations de localisation : il affirme juste qu'il tente de limiter les dégâts.

Il y a deux autres points importants à comprendre sur cette « architecture de protection de la vie privée » : d'abord, ce n'est qu'une architecture, pas un protocole, encore moins un logiciel. Elle définit un cadre et des principes, elle ne fournit pas de solutions clé en main. Ensuite, s'agissant d'un travail de l'IETF, elle ne traite que l'aspect technique. Si on suit ce RFC, et les standards futurs qui le développeront, on peut faire un système relativement protecteur de la vie privée. Mais la technique ne suffit pas : le respect du RFC ne garantit pas la sécurité, c'est une norme technique, pas une politique. Des lois de protection de la vie privée (et qui soient d'application générale, pas annulables par des conditions d'utilisation d'un service commercial), bien appliquées et une vigilance citoyenne permanente sont nécessaires si on veut empêcher la vie privée de devenir une marchandise comme les autres.

Ce RFC 6280[2], œuvre du groupe de travail Geopriv <http://tools.ietf.org/wg/geopriv>, qui travaille sur tous les problèmes situés à l'intersection de la localisation et de la vie privée, tourne autour de la notion d'« objet de localisation » (une bonne introduction à Geopriv est le RFC 3693). Cet objet stocke la localisation **et** les règles d'utilisation de cette information. Lorsque l'objet est transmis, les conditions d'utilisation voyagent avec lui (section 1.1). Ainsi, celui qui reçoit une information de localisation ne peut pas prétendre ignorer les règles qui s'y attachent. Ces règles sont, par exemple, « ne transmets ces données à personne » ou, plus subtil, « ne transmets les informations à mon entreprise que pendant les heures de travail ».

Lier les règles d'usage aux informations de localisation est contraignant mais permet de d'assurer que le destinataire a accès à ces règles. Le RFC cite l'exemple des licences libres comme Creative Commons, qui imposent une telle liaison. Il y a eu de nombreuses polémiques (par exemple au sujet de la GFDL), sur l'opportunité d'imposer l'inclusion du (long) texte de la licence dans chaque document couvert par cette licence mais le but est le même : couper court à toute tentative de jouer l'ignorance. Un autre exemple donné par le RFC est celui des classifications utilisées par les militaires (« Secret », « Confidentiel », etc). Tous les documents sont marqués avec leur niveau de sécurité et tout militaire qui les manipule sait donc exactement ce qu'il peut faire ou ne pas faire avec ce document. (Un exemple détaillé figure dans les règles de classification du ministère de la défense états-unien.)

La solution radicale serait évidemment d'arrêter la collecte et la transmission de ces informations. Mais, outre qu'elles peuvent permettre d'améliorer certains services (comme dans l'exemple des pizzérias plus haut), des systèmes comme le GSM ne peuvent pas fonctionner sans connaître la localisation de l'utilisateur (pas moyen de lui faire suivre un appel entrant si on ne sait pas quelle borne le dessert en ce moment). Comment limiter les dégâts ? Historiquement, les décisions en matière de vie privée étaient prises uniquement par le récepteur des données, qui en faisait à peu près ce qu'il voulait. Le seul choix de la personne concernée était de partager ses données ou pas, et ce choix binaire ne permettait pas d'indiquer ce qu'il acceptait qu'on fasse des données (section 1.3). Comme, évidemment, les intérêts des personnes ne coïncident pas avec ceux des entreprises privées et organisations étatiques qui reçoivent les données, les abus étaient inévitables. Par exemple, un utilisateur envoie volontairement son adresse électronique sur un site Web pour recevoir ensuite des informations, mais il ne pouvait jamais s'opposer à ce que cette adresse soit ensuite revendue à des spammeurs. (Aujourd'hui, il est courant de voir une case à cocher « J'accepte de recevoir du spam, pardon, que mon adresse soit revendue à des tiers » mais il faut noter que c'est le site Web destinataire qui définit les choix possibles, pas le titulaire de l'adresse.) Geopriv pose donc comme principe que l'utilisateur, pas le destinataire, définit les usages acceptables, et les indique avec les données. À noter que ce n'est qu'une indication. Il n'existe pas de moyen technique de s'assurer que ces choix de l'utilisateur seront respectés, une fois les données transmises. Sur

2. Pour voir le RFC de numéro NNN, https://www.ietf.org/rfc/rfcNNN.txt, par exemple https://www.ietf.org/rfc/rfc6280.txt

un réseau distribué comme l'Internet, il n'y a pas de possibilité de concevoir un mécanisme **technique** de respect de ces règles. Celui-ci doit être assuré par des moyens autres, comme la loi (loi Informatique & Libertés, par exemple), des régulateurs efficaces et/ou le marché (note personnelle : pour des raisons idéologiques, le marché est mentionné par le RFC comme mécanisme possible de rétroaction sur les entreprises qui violeraient les usages spécifiés par l'émetteur ; l'expérience déjà ancienne de l'Internet en matière de protection de la vie privée montre bien la vanité de cette idée). Le mécanisme technique de Geopriv sert donc juste à s'assurer que le destinataire des données était conscient des règles et donc, s'il les a violées, qu'il était de mauvaise foi (ce qui est important pour d'éventuelles poursuites judiciaires).

Avec la section 3 commence la partie la plus technique du RFC : elle décrit l'architecture utilisée. Les deux objectifs de cette architecture sont :

— S'assurer que les données de localisation ne soient distribuées qu'aux entités autorisées,
— S'assurer que ces entités soient au courant des conditions d'usage.

Comme on l'a vu, s'assurer que les entités réceptrices ne peuvent pas techniquement violer les règles est un non-but : il n'est pas réalisable techniquement. Pour atteindre les deux objectifs, il y a deux classes de règles :

— Les contrôles d'accès qui indiquent les entités autorisées,
— Les règles d'usage, qui indiquent ce que celles-ci ont le droit de faire.

Un exemple figure en section 3.1, décrivant le « cycle de vie » des données. Alice se déplace. Son mobile apprend sa position (par le réseau de communication, ou bien par GPS) et elle veut partager cette information avec ses amis via un service de présence (un tiers, qu'on suppose de confiance). Avec cette information viennent les règles d'usage qu'Alice a définies, par exemple que les amis ne peuvent pas garder cette information plus de N jours. Si un ami, mettons Bob, veut transmettre l'information sur la localisation d'Alice à des amis à lui, il devra également vérifier si les règles d'usage lui permettent (rappelez-vous que ce RFC définit une architecture de sécurité, pas des moyens techniques de protection ; comme le savent ceux qui ont essayé de mettre en œuvre des DRM, il n'existe aucun moyen **technique** d'empêcher Bob de tricher et de faire suivre ces données à des gens non autorisés).

Décrire ensuite tout ce qui se passe nécessite un vocabulaire élaboré (section 3.2). On y trouve de nombreux rôle (un même acteur pouvant tenir plusieurs rôles) :

— Le **sujet** ("*target*") est la personne dont on veut protéger la vie privée (Alice, dans l'exemple précédent),
— L'**engin** ("*device*") est la machine que porte le sujet (par exemple son "*smartphone*") ; techniquement, c'est lui qui est localisé, pas le sujet (pensez au nombre de films où un personnage jette son téléphone pour ne pas être repéré),
— Le **décideur** ("*rule maker*") est celui qui définit les règles d'usage ; c'est souvent le sujet, mais cela peut être le responsable sécurité de son entreprise,
— Le **localisateur** ("*location generator*") est le matériel ou logiciel qui détermine la localisation ; cela peut être un récepteur GPS ou bien la partie GSM de l'appareil, qui obtient sa localisation à partir du réseau GSM,
— Les **serveurs de localisation** ("*location server*"), terme impropre : un serveur de localisation est une entité susceptible de transmettre l'information de localisation, en appliquant les règles d'usage. Dans l'exemple plus haut, Alice, le serveur de présence, et Bob connaissaient tous les trois la localisation d'Alice et étaient susceptibles de la transmettre (et donc étaient des serveurs de localisation potentiels),
— Les **destinataires de la localisation** ("*location recipient*") sont toutes les entités qui connaissent la localisation du sujet et s'en servent mais ne la transmettent pas forcément. Bob, le serveur de présence, les autres amis qui ne relayaient pas l'information, sont tous des destinataires de localisation.

J'ai simplifié en mettant un seul décideur et un seul localisateur mais il peut y avoir des cas plus complexes.

Pour mettre en œuvre l'architecture de sécurité Geopriv, on a besoin de deux formats de données :

— Les **objets de localisation** (LO pour "*Location Object*", cf. RFC 5491), qui stockent une position, exprimée en géodésique ("*geodetic*", c'est-à-dire longitude, latitude, altitude) ou bien en civil ("*civic*", comme 120, rue de la Gare, 75013, Paris..., voir par exemple le RFC 5139),
— Les **règles d'usage** (PR pour "*Privacy Rules*", et RFC 4745 pour un exemple de langage pour exprimer ces règles).

En utilisant cette terminologie, la section 4 du RFC décrit le cycle de vie de l'information de localisation. Elle passe par trois étapes :

— **Détermination** de la localisation (un "*fix*", dans la terminologie GPS),
— **Distribution** de la localisation, par les serveurs de localisation,
— **Utilisation** de la localisation, par les destinataires.

Les trois étapes sont ensuite étudiées une à une, chacune posant des problèmes spécifiques de protection de la vie privée.

La section 4.1 parle de la détermination. Elle peut être faite par l'engin lui-même (GPS, observation des étoiles et bien d'autres, y compris pour le cas trivial où l'utilisateur rentre manuellement l'information), par le réseau auquel il se connecte, ou bien via une coopération entre tous les deux. Dans le premier cas, guère de problème de sécurité dans cette étape, l'engin étant tout seul (les satellites GPS ne dialoguent pas avec l'engin, ils ne savent pas qui les utilise). Bien plus délicat est le cas où c'est le réseau qui détermine la position. C'est ce qui se produit en GSM (le réseau sait avec quelle base parle l'engin, et l'intensité du signal, ainsi que sa direction, peuvent donner une idée plus précise de la position de l'engin) mais évidemment aussi en connexion filaire, où le réseau sait exactement où on se trouve. Tout mécanisme du genre GeoIP est également un cas de détermination de la position par le réseau. Il n'y a pas forcément besoin d'un protocole réseau. Mais le point important est qu'un tiers (pas seulement le sujet) connait dès le début la position de l'engin. Enfin, le troisième cas est celui de la détermination assistée par le réseau. Par exemple, un engin va mesurer un certain nombre de choses comme la force du signal radio, l'adresse MAC obtenue, etc, et les envoyer à un localisateur qui en déduira une position précise. Cette fois, il y a un protocole réseau (entre l'engin et le localisateur), qu'il faut protéger, par exemple contre l'écoute.

L'étape de détermination est le premier endroit où on peut appliquer des mesures de protection, comme l'utilisation d'identificateurs qui ne permettent pas un lien simple avec l'utilisateur, et comme le chiffrement pour empêcher l'écoute, dans le cas où la détermination emploie un dialogue sur le réseau.

La deuxième étape est la distribution de l'information de localisation (section 4.2). C'est là que les règles d'usage commencent à jouer un rôle, en limitant la redistribution. Voici quelques exemples de règles :

— Ne retransmettre cette information qu'aux machines dont le nom se termine en `example.com`,
— Ne transmettre de l'adresse que la ville et le pays (pas la rue),
— Dans la même idée de dégradation délibérée de la précision, ne transmettre de la longitude et latitude que des valeurs arrondies à la minute d'angle la plus proche (environ deux kilomètres)...

Les deux derniers exemples montrent qu'un serveur de localisation a deux armes à sa disposition, ne pas transmettre la localisation du tout, ou bien la transformer pour limiter les risques.

Pour cette étape de distribution, les choses les plus importantes sont de toujours transmettre les règles d'usage avec la localisation, et de respecter ensuite ces règles. Ces règles pouvant dépendre de l'**identité** de celui à qui on transmet, utiliser des techniques d'authentification fiables est donc impératif. De même, il faut veiller à ce qu'un tiers non-autorisé ne puisse pas accéder aux données qui circulent sur le réseau, donc le chiffrement est fortement souhaité. On est là dans un domaine plus traditionnel pour l'IETF, la conception de protocoles réseaux sûrs.

Enfin, la troisième étape du cycle de vie de l'information de localisation est l'utilisation de l'information (section 4.3). Le destinataire doit respecter les règles d'usage, qu'il a reçu avec l'objet de localisation. Normalement, ce respect est obligatoire et des mesures administratives ou légales doivent être déployées

pour s'assurer que ce soit le cas. Cette étape ne comprend pas de transmission de l'information et ne nécessite donc pas de protocole réseau.

La section 6 du RFC fournit des exemples de scénarios concrets d'utilisation de cette information de localisation. Le premier scénario est très simple : un engin détermine sa position et la transmet à une application Web dans une requête HTTP de type `POST` ou bien une requête SIP de type `INVITE`. Le serveur fournit le service demandé, sans stocker l'information de localisation. Dans ce cas ultra-simple, pas de place pour les architectures compliquées, les rôles sont bien définis (le RFC oublie toutefois de rappeler que la connexion entre l'engin et le serveur doit être protégée contre l'écoute).

Plus complexe, en section 6.2, le cas où un navigateur Web qui fait appel à des API de détermination de sa position (comme la "*Geolocation API*" <http://dev.w3.org/geo/api/spec-source.html>, dont l'implémentation fait souvent appel à l'aide du réseau d'accès) et exécute du code JavaScript qui va appeler des services d'assistance à la détermination de la position (voyez un bon tutoriel en « "*Geocoding A User's Location Using Javascript's GeoLocation API*" <http://www.bennadel.com/blog/2023-Geocoding-A-User-htm> »). Notez qu'une partie des risques peut provenir d'une action du navigateur **après** avoir obtenu sa position. Par exemple, s'il demande à Google Maps une carte des environs, Google Maps, bien que n'étant pas partie lors de la détermination de la position, la connaîtra désormais. Il y a plus de deux acteurs, cette fois (notez que, même sur une machine unique, il peut y avoir plusieurs composants logiciels ne se faisant pas mutuellement confiance) et donc davantage de complexité.

Enfin, le troisième scénario est plus critique puisqu'il concerne un appel d'urgence, domaine très régulé, puisque beaucoup d'opérateurs ont l'obligation légale de fournir automatiquement la localisation de l'appel. Ici, la distribution de l'information se fera typiquement en utilisant LoST (RFC 5222).

Comme tout RFC, notre RFC 6280 inclut une section sur les problèmes de sécurité (section 5). Ici, elle sert surtout de rappel car tout le RFC est consacré à la sécurité. Parmi les rappels : le fait qu'il existe déjà des protocoles fournissant les fonctions de sécurité importantes (IPsec et TLS, par exemple) et que les solutions de localisation devraient les utiliser. Autre rappel : la sécurité de bout en bout est préférable à la sécurité transitive et il est donc recommandé qu'on sécurise l'ensemble du trajet, pas juste chaque étape (pensez au cas de Bob faisant suivre un objet de localisation : l'a-t-il modifié ?). Une bonne étude générale des problèmes de sécurité de Geopriv est dans le RFC 3694.