Choisir ses identificateurs

Stéphane Bortzmeyer
AFNIC
bortzmeyer@nic.fr

Choisir ses identificateurs

Stéphane Bortzmeyer
AFNIC
bortzmeyer@nic.fr

Plan du tutoriel

1 Introduction

2 Propriétés

3 Futur des identités

4 Nommer par le contenu

5 Conclusion

Introduction

Introduction

1. Dans tout réseau, il faut identifier les **entités**

Introduction

1. Dans tout réseau, il faut identifier les **entités**
2. Machines, humains, programmes en cours d'exécution, fichiers, ...

Introduction

1. Dans tout réseau, il faut identifier les **entités**
2. Machines, humains, programmes en cours d'exécution, fichiers, . . .
3. D'innombrables électrons ont été agités pour les débats sur les « bons » identificateurs

Vecteur d’identité

Vecteur d'identité

1. Les identificateurs ne sont pas qu'un élément technique (cf. les discussions autour des IDN)

Vecteur d'identité

1. Les identificateurs ne sont pas qu'un élément technique (cf. les discussions autour des IDN)
2. Ce sont des vecteurs d'identité, un élément de la présence en ligne

Vecteur d'identité

1. Les identificateurs ne sont pas qu'un élément technique (cf. les discussions autour des IDN)
2. Ce sont des vecteurs d'identité, un élément de la présence en ligne

« Je ne suis pas un login Facebook, je suis un homme libre ! »

Discussions savantes

Discussions savantes

1. URI, URN, URL. . . Quizz : qui connait les différences ?

Discussions savantes

1. URI, URN, URL. . . Quizz : qui connait les différences ?
2. *A name indicates what we seek. An address indicates where it is. A route indicates how to get there* (RFC 791)

Discussions savantes

1. URI, URN, URL. . . Quizz : qui connait les différences ?
2. *A name indicates what we seek. An address indicates where it is. A route indicates how to get there* (RFC 791)
   ⟵ Non, ce n'est pas vrai !

Discussions savantes

1. URI, URN, URL. . . Quizz : qui connait les différences ?
2. *A name indicates what we seek. An address indicates where it is. A route indicates how to get there* (RFC 791)
3. *A name is a unique string, N, in some alphabet, A, that unambiguously denotes some object. . .* (John Day)

Faiblesse des définitions savantes

Faiblesse des définitions savantes

1. Difficiles à comprendre (pensez au modèle en couches. . .)

Faiblesse des définitions savantes

1. Difficiles à comprendre (pensez au modèle en couches. . .)
2. Collent très peu à la réalité de l'Internet (elles peuvent toujours être utiles pour un cours à la fac)

Faiblesse des définitions savantes

1. Difficiles à comprendre (pensez au modèle en couches. . .)
2. Collent très peu à la réalité de l'Internet (elles peuvent toujours être utiles pour un cours à la fac)
3. Une adresse IP n'identifie **pas** un lieu, ni physique (évidemment), ni virtuel (*anycast*)

Faiblesse des définitions savantes

1. Difficiles à comprendre (pensez au modèle en couches. . .)
2. Collent très peu à la réalité de l'Internet (elles peuvent toujours être utiles pour un cours à la fac)
3. Une adresse IP n'identifie **pas** un lieu, ni physique (évidemment), ni virtuel (*anycast*)
4. Un URL n'est pas un localisateur (où est `http://www.google.com/` ?), cf. leur utilisation dans XML

Plan du tutoriel

1 Introduction

2 Propriétés

3 Futur des identités

4 Nommer par le contenu

5 Conclusion

Autre approche pour classer

Autre approche pour classer

1. S'appuyer sur les **propriétés** des identificateurs

Autre approche pour classer

1. S'appuyer sur les **propriétés** des identificateurs
2. Puis faire son **shopping** en fonction des propriétés importantes

Autre approche pour classer

1. S'appuyer sur les **propriétés** des identificateurs
2. Puis faire son **shopping** en fonction des propriétés importantes
3. Se fier aux propriétés plutôt qu'aux étiquettes : par exemple, un URL n'indique **pas** la localisation

Autre approche pour classer

1. S'appuyer sur les **propriétés** des identificateurs
2. Puis faire son **shopping** en fonction des propriétés importantes
3. Se fier aux propriétés plutôt qu'aux étiquettes : par exemple, un URL n'indique **pas** la localisation
4. Être réaliste : on n'aura pas à la fois le beurre et l'argent du beurre `http://www.bortzmeyer.org/nommage-beurre.html`

Les propriétés souhaitables

1. **Unicité**. Nécessaire pour être contacté sans ambiguïté. Les propositions de « racines alternatives » cassent toutes cette propriété. Le problème des identificateurs status.net. La quasi-unicité (clé PGP ou SSH).

Les propriétés souhaitables

1. **Unicité**. Nécessaire pour être contacté sans ambiguïté. Les propositions de « racines alternatives » cassent toutes cette propriété. Le problème des identificateurs status.net. La quasi-unicité (clé PGP ou SSH).
2. **Stabilité**. Nécessaire pour une publication dans une référence stable (article scientifique). Dépend de procédures humaines, pas juste de la technique.
   `http://www.w3.org/Provider/Style/URI.html`

Les propriétés souhaitables

1. **Unicité**. Nécessaire pour être contacté sans ambiguïté. Les propositions de « racines alternatives » cassent toutes cette propriété. Le problème des identificateurs status.net. La quasi-unicité (clé PGP ou SSH).
2. **Stabilité**. Nécessaire pour une publication dans une référence stable (article scientifique). Dépend de procédures humaines, pas juste de la technique. `http://www.w3.org/Provider/Style/URI.html`
3. **Expressivité**. `www.afnic.fr` contre `2001:67c:2218:2::4:20`. Pas la seule raison d'utiliser les noms de domaine (la stabilité est plus importante).

Les propriétés souhaitables, saison 2

1. **Résolvabilité.** Un identificateur peut servir à. . . identifier mais aussi à accéder à quelque chose. Notez que tout identificateur est résolvable en le mettant dans Google ou dans une DHT. . .

Les propriétés souhaitables, saison 2

1. **Résolvabilité**. Un identificateur peut servir à. . . identifier mais aussi à accéder à quelque chose. Notez que tout identificateur est résolvable en le mettant dans Google ou dans une DHT. . .
2. **Démocratie**. On veut qu'obtenir un identifiant soit rapide, pas cher, sans formalités excessives. . .

Les propriétés souhaitables, saison 2

1. **Résolvabilité**. Un identificateur peut servir à. . . identifier mais aussi à accéder à quelque chose. Notez que tout identificateur est résolvable en le mettant dans Google ou dans une DHT. . .
2. **Démocratie**. On veut qu'obtenir un identifiant soit rapide, pas cher, sans formalités excessives. . .
3. **Sécurité**. Si je suis `www.afnic.fr`, je ne veux pas que quelqu'un d'autre puisse l'utilise par une astuce technique.

Exemples d’identficateurs et de leurs propriétés

Les identificateurs actuels ont-ils toutes ces propriétés ?

Pas toutes à la fois, ne rêvons pas

Noms de domaines

`www.afnic.fr`

1. Uniques grâce à l'allocation décentralisée
2. Assez stables (mais risques juridico-financiers)
3. Très expressifs (c'est bien pour cela qu'ils suscitent des convoitises)
4. Très facilement résolvables, technologie fiable et éprouvée (DNS)
5. Payants, nécessitent de passer par des organismes qu'on n'aime pas forcément (cf. Laurent Chemla, « Confessions d'un voleur ») et qui peuvent censurer (saisies de noms de domaine sur des `.org` la semaine dernière)
6. Pas très sûrs mais en progrès (sécurité de l'enregistrement, DNSSEC)

Host Identifiers

1. Créés par le protocole HIP (*Host Identity Protocol*, RFC 5201). Le HI (*Host Identifier*) est une clé cryptographique publique, le HIT (*Host Identity Tag*) son condensat.
2. Quasi-uniques, comme souvent en cryptographie (clés SSH ou PGP),
3. Très stables (tant qu'on ne perd pas la clé privée)
4. Pas du tout expressifs (binaire pur). Le HIT est un peu plus gérable que le HI.
5. Pas de mécanisme de résolution universel (mais des essais)
6. Gratuits et générés entièrement localement, pas de registre à payer
7. Très sûrs (la machine signe ses paquets avec la clé privée de son nom...)

ISBN

1. Uniques grâce à l'allocation décentralisée
2. Très stables
3. Pas du tout expressifs (mais très courants et bien acceptés dans la communauté : l'expressivité est une notion relative)
4. Pas du tout résolvables (aucun service standard aujourd'hui, c'est encore Google qui est le plus efficace bien qu'on pourrait faire une DHT des ISBN).
   `978-2-87772-409-8` est bien trouvé.
5. Pas de sécurité particulière (pas forcément utile)

BitMessage

1. Un système de messagerie pair-à-pair fonctionnant par inondation. Mon adresse est `BM-2D8rwZkR1KvUMCnBhH7MGzTwVRXnDvhMC9`, le condensat d'une clé privée (comme un HIT)
2. Quasi-uniques, comme souvent en cryptographie
3. Très stables (tant qu'on ne perd pas la clé privée - possibilité de dérivation stable à partir d'un mot de passe)
4. Pas du tout expressifs (exemple ci-dessus)
5. Résolution par inondation, ce qui est coûteux
6. Gratuits et générés entièrement localement, pas de registre à payer
7. Très sûrs (messages signés avec la clé privée correspondante)

afnic

ARK

1. Uniques grâce à l'allocation décentralisée
2. Identifiant stable choisi par la BNF
3. Pas expressif, exprès (un identificateur parlant empêche de changer d'avis, exemple `http://blog.example.org/machin-est-un-con.html` ne peut plus être modifié sans faire des 404)
4. Résolvable par une astuce : syntaxe standard pour faire un URL à partir d'un ARK. L'ARK `12148/bpt6k101412s` devient l'URL `http://catalogue.bnf.fr/ark:/12148/bpt6k101412s`. Si `catalogue.bnf.fr` disparait, on peut remplacer par un autre.

Namecoin

1. Uniques grâce à l'allocation pair-à-pair **si et seulement si** tout le monde reconnait la même chaîne
2. Stables tant qu'on ne perd pas ses fichiers
3. Parlants (ce sont des chaînes de caractères normales)
4. Pas résolvables mais on peut le combiner avec le DNS (pseudo-TLD `.bit`)
5. Sûr (autant que Bitcoin)

[Troll] Mots-clés Google

On peut constater tous les jours que beaucoup de gens utilisent les mots-clés dans une recherche Google comme identificateur.
« Pour voir le site d'Alcatel, tapez "Alcatel" dans Google »

1. Pas uniques, loin de là.
2. Aucune stabilité, change tous les jours (changements des « concurrents », changement de l'algorithme)
3. Très expressifs
4. Résolvables très rapidement aujourd'hui, grâce à l'excellente infrastructure de Google
5. Très sécurisés, la NSA gère les sauvegardes

Contradiction entre propriétés

Contradiction entre propriétés

1. L'unicité implique un système centralisé ou arborescent. Elle s'oppose donc au désir d'avoir des noms obtenables localement.

Contradiction entre propriétés

1. L'unicité implique un système centralisé ou arborescent. Elle s'oppose donc au désir d'avoir des noms obtenables localement.
2. Le caractère parlant d'un nom s'oppose à sa stabilité (conflit de possession sur les noms de domaine : tout le monde veut `sex.com` alors que personne ne me réclame `2001:660:3003:2::4:20`)

Un peu de théorie

L'impossibilité d'avoir toutes les propriétés à la fois est-elle prouvée ?

Pas au sens mathématique. Disons que c'est une conjecture.

Plan du tutoriel

1 Introduction

2 Propriétés

3 Futur des identités

4 Nommer par le contenu

5 Conclusion

Comment sera t-on identifié dans le futur ?

1. Jean Dupont, né le 3 octobre 1978 à Bois-le-Roi,
2. `jdupont43` sur Twitter,
3. `2001:db8:1:76a:bd2::1`,
4. `jean.dupont.fr`,
5. « Cherche Jean Dupont sur Google ».

Inconvénients de certaines identités

1. Jean Dupont, né le 3 octobre 1978 à Bois-le-Roi : pas unique et pas résolvable
2. `jdupont43` sur Twitter : vous met à la merci d'une boîte privée, à but lucratif, et PRISM-compatible
3. `2001:db8:1:76a:bd2::1` : cool, ça fait e1eeT mais peut-être pas très maniable. Convient pour les routeurs, moins bien pour les humains.
4. `jean.dupont.fr` : OK, mais de moins en moins utilisé au fur et à mesure que tout le monde publie son identité Facebook. Problèmes d'utilisabilité ? De confiance ?
5. « Cherche Jean Dupont sur Google » : Michu-compatible.

Nouveaux protocoles qui comblent les vides

Un nom de domaine seul ne donne pas accès à assez d’informations structurées.

Nouveaux protocoles qui comblent les vides

Un nom de domaine seul ne donne pas accès à assez d'informations structurées.

1. `.tel`, un TLD spécial pour avoir ces informations ?

Nouveaux protocoles qui comblent les vides

Un nom de domaine seul ne donne pas accès à assez d'informations structurées.

1. `.tel`, un TLD spécial pour avoir ces informations ?
2. Qui se souvient à quoi servait le port 79 ?

Nouveaux protocoles qui comblent les vides

Un nom de domaine seul ne donne pas accès à assez d'informations structurées.

1. `.tel`, un TLD spécial pour avoir ces informations ?
2. Qui se souvient à quoi servait le port 79 ?
3. Plus prometteur (car indépendant du TLD), WebFinger `draft-ietf-appsawg-webfinger-15` à l'IETF. *Given a person, how do I find out what services that person uses ?* (John Panzer) Webfinger permet de récupérer des données à partir d'une adresse de courrier (première version) ou d'un URI (version IETF).

Plan du tutoriel

1 Introduction

2 Propriétés

3 Futur des identités

4 Nommer par le contenu

5 Conclusion

Le principe

Le principe

1. Désigner une entité par son contenu et pas sa localisation

Le principe

1. Désigner une entité par son contenu et pas sa localisation
2. Exemple le CCN (*Content-Centric Networking*) du PARC (Van Jacobson et al.)

Le principe

1. Désigner une entité par son contenu et pas sa localisation
2. Exemple le CCN (*Content-Centric Networking*) du PARC (Van Jacobson et al.)
3. Attention, pub ! Contrairement à ce que disent les promoteurs du CCN, les autres identificateurs ne s'appuient pas forcément sur la localisation

Le principe

1. Désigner une entité par son contenu et pas sa localisation
2. Exemple le CCN (*Content-Centric Networking*) du PARC (Van Jacobson et al.)
3. Attention, pub ! Contrairement à ce que disent les promoteurs du CCN, les autres identificateurs ne s'appuient pas forcément sur la localisation
4. Exemple d'un identificateur CCN : `/parc.com/media/art/carla-bruni.mp3` Vous voyez la différence avec un URL (moi, pas) ?

Identificateurs réellement fondés sur le contenu

Identificateurs réellement fondés sur le contenu

1. Ils sont formés d'un condensat du contenu,

Identificateurs réellement fondés sur le contenu

1. Ils sont formés d'un condensat du contenu,
2. Et de zéro, un ou plusieurs localisateurs.

Identificateurs réellement fondés sur le contenu

1. Ils sont formés d'un condensat du contenu,
2. Et de zéro, un ou plusieurs localisateurs.

ARK avait déjà cette idée du localisateur facultatif, complétant l'identificateur
`http://gallica.bnf.fr/ark:/12148/bpt6k107371t`

Auto-validation

L'intérêt du condensat est de pouvoir valider le contenu une fois récupéré.

Mais cela rend ce mécanisme très sensible à des changements de détail.

Magnet

Magnet

1. Le magnet a été popularisé par BitTorrent (mais peut être utilisé ailleurs)

Magnet

1. Le magnet a été popularisé par BitTorrent (mais peut être utilisé ailleurs)
2. Identificateur par défaut dans ThePirateBay

Magnet

1. Le magnet a été popularisé par BitTorrent (mais peut être utilisé ailleurs)
2. Identificateur par défaut dans ThePirateBay
3.

```
magnet:?xt=urn:btih:aee6206d41c92bc6ff5ad007597b
dn=Doctor+Who+S01+1963+The+Daleks+Serial+%7E%2AS
tr=udp%3A%2F%2Ftracker.openbittorrent.com%3A80&
tr=udp%3A%2F%2Ftracker.publicbt.com%3A80&
tr=udp%3A%2F%2Ftracker.istole.it%3A6969&tr=udp%3
```

Magnet

1. Le magnet a été popularisé par BitTorrent (mais peut être utilisé ailleurs)
2. Identificateur par défaut dans ThePirateBay
3.

```
magnet:?xt=urn:btih:aee6206d41c92bc6ff5ad007597b
dn=Doctor+Who+S01+1963+The+Daleks+Serial+%7E%2AS
tr=udp%3A%2F%2Ftracker.openbittorrent.com%3A80&
tr=udp%3A%2F%2Ftracker.publicbt.com%3A80&
tr=udp%3A%2F%2Ftracker.istole.it%3A6969&tr=udp%3
```

On y voit le condensat cryptographique, et des localisations possibles

Les (chevaliers qui disent) NI

Les (chevaliers qui disent) NI

1. RFC 6920 `http://www.bortzmeyer.org/6920.html`

Les (chevaliers qui disent) NI

1. RFC 6920 `http://www.bortzmeyer.org/6920.html`
2. Autorité (le localisateur) facultatif
   `ni://www.bortzmeyer.org/sha256;6OuucQ1RgugCDVinI`
   (notez que Google résoud parfaitement ce condensat incompréhensible, sans avoir besoin de l'autorité)

NI contre Magnet

NI contre Magnet

1. NI est normalisé. Magnet n’est même pas réellement documenté (lis le code source, Luke)

NI contre Magnet

1. NI est normalisé. Magnet n'est même pas réellement documenté (lis le code source, Luke)
2. Magnet permet plusieurs localisateurs

NI contre Magnet

1. NI est normalisé. Magnet n'est même pas réellement documenté (lis le code source, Luke)
2. Magnet permet plusieurs localisateurs
3. Magnet permet d'autres protocoles que HTTP pour les localisateurs

NI contre Magnet

1. NI est normalisé. Magnet n'est même pas réellement documenté (lis le code source, Luke)
2. Magnet permet plusieurs localisateurs
3. Magnet permet d'autres protocoles que HTTP pour les localisateurs
4. Magnet est plus déployé

NI contre Magnet

1. NI est normalisé. Magnet n'est même pas réellement documenté (lis le code source, Luke)
2. Magnet permet plusieurs localisateurs
3. Magnet permet d'autres protocoles que HTTP pour les localisateurs
4. Magnet est plus déployé
5. Aucun des deux ne semble reconnu nativement par les navigateurs Web

Plan du tutoriel

1 Introduction

2 Propriétés

3 Futur des identités

4 Nommer par le contenu

5 Conclusion

Que faire ?

1. Éducation : aujourd'hui, comprendre les identificateurs doit faire partie de la culture numérique de base.

Que faire ?

1. Éducation : aujourd'hui, comprendre les identificateurs doit faire partie de la culture numérique de base.
2. Politique/juridique : travailler à sécuriser les identificateurs, par exemple contre le *reverse hijacking* (quand une grosse compagnie essaie de piquer un nom de domaine à une entité plus petite) ou contre les saisies par un État, sans procédure contradictoire.

Que faire ?

1. Éducation : aujourd'hui, comprendre les identificateurs doit faire partie de la culture numérique de base.
2. Politique/juridique : travailler à sécuriser les identificateurs, par exemple contre le *reverse hijacking* (quand une grosse compagnie essaie de piquer un nom de domaine à une entité plus petite) ou contre les saisies par un État, sans procédure contradictoire.
3. Sécurité : DNSSEC et ses copains.

Que faire ?

1. Éducation : aujourd'hui, comprendre les identificateurs doit faire partie de la culture numérique de base.
2. Politique/juridique : travailler à sécuriser les identificateurs, par exemple contre le *reverse hijacking* (quand une grosse compagnie essaie de piquer un nom de domaine à une entité plus petite) ou contre les saisies par un État, sans procédure contradictoire.
3. Sécurité : DNSSEC et ses copains.
4. Nouveaux services comblant les manques afin de concurrencer les gros silos du Web 2.0 (par exemple Webfinger).

Merci !

www.afnic.fr
contact@afnic.fr