SAN JUAN — Indicateurs de santé des technologies des identificateurs
Mardi 13 mars 2018 – 17h à 18h30 AST
ICANN61 – San Juan, Porto Rico

CATHY PETERSON : Bonjour à tous. Nous allons commencer la séance sur les indicateurs de santé des technologies des identificateurs dans quelques minutes. Nous vous accordons simplement quelques minutes de plus. Merci.

ALAIN DURAND : Bonjour. Il s'agit de la séance sur les indicateurs de santé des technologies des identificateurs. C'est un projet qui a été lancé depuis un certain temps, et aujourd'hui, nous allons vous montrer quelques données chiffrées intéressantes [inaudible]. Chiffres qui devraient vous intéresser. Ils m'ont intéressé.

Dans cette séance, nous aurons trois présentateurs. Le premier sera Paul Wilson, l'actuel président de la NRO. Il va nous faire une mise à jour sur ce que la NRO a fait dans ce domaine.

Le deuxième et le troisième se concentreront sur la partie principale du projet, les choses qui sont gérées par l'ICANN. La deuxième présentation sera faite par Christian Huitema sur les

indicateurs actuels et les données que nous constatons sur les indicateurs actuels.

La dernière présentation sera faite par Geoff Huston sur la proposition d'un ensemble de nouveaux indicateurs.

Alors sans plus tarder, je vais laisser la parole à Paul, qui va nous parler des activités dans l'espace des numéros. Paul ?

PAUL WILSON : Merci, Alain et bonjour tout le monde. Nous faisons tourner cinq registres Internet régionaux et chacun de nous fait fonctionner un registre WHOIS en utilisant un service WHOIS assez familier. Il existe donc cinq différents registres qui sont exploités par cinq RIR. Ils sont assez étroitement coordonnés entre eux. Techniquement ils sont capables d'incohérence et, bien sûr, d'erreur et de lacunes et ainsi de suite. Ainsi les RIR travaillent ensemble sous la bannière de la NRO pour s'assurer que ces registres ont un sens les uns par rapport aux autres et qu'ils jouent leur rôle en ce qui concerne les enregistrements qu'ils stockent.

C'est ce que nous faisons depuis très longtemps, mais je pense que les choses sont en train de changer quelque peu au cours des dernières années parce qu'un plus large groupe de parties liées à ces bases de données font preuve d'un plus grand intérêt

concernant l'exactitude et l'efficacité de ces bases de données. En outre, le rythme de mises à jour s'accroît aussi relativement bien. Donc, avant que nous atteignions la pénurie actuelle d'adresses IPv4, les allocations étaient assez statiques. Elles étaient faites aux parties qui gardaient ces attributions et les utilisaient. Mais ces jours-ci, nous voyons passer beaucoup de transferts. À l'intérieur des régions et entre les registres Internet régionaux, il se passe ainsi beaucoup de transferts, qui doivent évidemment être mis à jour dans la base de données. C'est une autre raison pour laquelle nous mettons de plus en plus l'accent sur l'exactitude et l'exhaustivité et ainsi de suite.

Comme je l'ai dit, nous sommes préoccupés depuis la création des RIR par le fait que les registres fassent leur travail. Nous n'avons pas parlé de la santé en tant que telle, mais l'idée d'un identificateur de santé est quelque chose de nouveau qui, je pense, est arrivé avec le projet de l'ICANN. Mais cela dit, nous avons conservé cette direction.

L'autre aspect de cette situation est, bien sûr, le fait que nous avons des relations d'adhésion avec les opérateurs de réseau qui sont les premiers récipiendaires des blocs d'adresse et des ASN. De sorte qu'ils sont tenus, en vertu de leurs relations avec chacun des RIR de tenir leurs registres à jour et bien tenus. Il y a des questions de politique sur ce que sont exactement les

attentes et les pénalités, pour ainsi dire, pour ne pas se conformer à ces politiques.

Ces choses sont traitées en général à un niveau régional. De sorte que les cinq RIR ont des processus indépendants de politiques et d'adhésion et auront à des moments différents des discussions sur les politiques connexes au WHOIS. Ces discussions, comme je l'ai mentionné, connaissent également une augmentation de fréquence et d'intensité de nos jours. Je dirais qu'avec ces cinq régions, je pense qu'il est juste de dire que nous travaillons tous à resserrer les liens de différentes façons, mais généralement dans la même direction, vers une mise en œuvre plus claire et plus ferme des politiques.

Le projet ITHI n'a pas vraiment étonné du côté de l'ICANN. Il s'agit de toute évidence d'un intérêt partagé de tous ceux d'entre nous qui exercent des rôles de registre, et nous avons donc décidé de nous inscrire dans le cadre de l'initiative ITHI de l'ICANN. Je pense qu'elle a été en réalité très utile parce que, même si nous avons collaboré très étroitement, nous n'avons pu nous entendre sur un ensemble d'indicateurs cohérents alors que nous avons pu progresser via le projet ITHI.

Nous l'avons durant une période de l'année précédente examiné avec notre groupe de coordination des services d'enregistrement, c'est un groupe du personnel de l'ensemble

des cinq RIR qui travaillent dans les domaines des services d'enregistrement. Ce groupe a effectué des travaux auxquels nous ferons référence sur un projet d'ensemble d'indicateurs en ce qui concerne la santé des identificateurs dans l'espace de numéros. Il a aussi lancé une consultation publique sur un papier préliminaire.

Cela s'est produit vers la fin de l'année dernière, et cela a donné une opportunité à nos communautés d'alimenter ce processus. Nous avons en fait reçu très peu de commentaires, et nous avons donc maintenant un document qui est pratiquement prêt pour l'approbation finale et sa publication. Fondamentalement il documente la santé des identificateurs dans l'espace des numéros en termes d'enregistrements WHOIS. Ce que nous avons obtenu est les trois C : des données complètes, correctes et courantes (actuelles), ce qui était notre objectif.

Mais ceux-ci sont ventilés en cinq mesures spécifiques qui sont des indicateurs mesurables de notre base de données qui se rapportent à l'intégralité de la base de données, son unicité, la correspondance de notre base de données avec d'autres documents officiels externes, l'efficacité des données concernées à effectivement contacter les gens qui sont inscrits dans le registre, et leurs capacités à être mises à jour aussi. Ce document identifie également les divers risques associés à l'absence d'atteinte de nos objectifs dans ces mesures, et il

analyse les causes qui seraient associées à ce genre de défaillance.

Donc, ce document sera publié sous peu. Ce que nous n'avons pas encore, évidemment, puisque les indicateurs eux-mêmes ne sont encore que sous forme de projet, nous n'avons pas encore ce qu'Alain, je pense, va présenter sous peu, ce sont des données réelles de notre conformité. Mais il est évident que le but de ces indicateurs est d'être capables de mesurer des choses permettant de définir des objectifs avec lesquels nous espérons nous conformer et surveiller le degré de conformité sur une période de temps.

Je pense donc que si vous êtes intéressés par l'espace des numéros, alors, surveillez cet espace et nous serons en mesure de faire un rapport en temps opportun sur la santé des identificateurs Internet dans l'espace des numéros dans les cinq RIR. Je pense que c'est tout. Merci, Alain.

ALAIN DURAND : Merci, Paul. Je voudrais profiter de cette occasion pour vous remercier, vous et les autres membres de la communauté des numéros pour avoir collaboré avec nous sur ce projet. Je pense que c'est une collaboration intéressante, et nous apprenons beaucoup grâce à ce processus.

PAUL WILSON : Oui, je suis d'accord. Je vous retourne le compliment, Alain. Merci.

ALAIN DURAND : Maintenant, nous suivons le même processus pour regarder l'espace [problème] et définir des indicateurs et ensuite obtenir la réelle mesure pour qu'elle soit normale. Mais vous n'avez pas encore les chiffres, et nous avons hâte de voir à l'avenir le premier lot de chiffres lorsque ceux-ci seront prêts.

Maintenant nous allons passer à la vitesse supérieure et passer dans l'espace de noms. Christian va maintenant faire une présentation pour indiquer où nous en sommes.

CHRISTIAN HUITEMA : Bonjour. Je suis Christian Huitema. J'ai travaillé sur la mesure des données du DNS et de leur [statut] depuis environ un an et demi, après avoir fait une étude pour voir ce qui pourrait être fait avec le DNS en travaillant avec Alain à calculer les indicateurs actuels.sur faisant les chiffres réels.

Comme Paul l'a dit, lorsque vous configurez ces indicateurs, il faut respecter certains principes. Les principes que nous avons adoptés sont précisés sur cette diapositive. Nous souhaitions

d'abord réellement que cette opération pour trouver ces indicateurs soit technique. Nous ne souhaitons pas faire intervenir les politiques. Le but des indicateurs est de décrire l'état du système. Il n'est pas de rendre un jugement dans un sens ou dans l'autre.

Nous avons cherché à définir des domaines que nous voulions suivre, des domaines qui seraient susceptibles de poser des problèmes et ensuite, définir des indicateurs dans ces domaines, et définir des moyens de les mesurer.

Un autre principe est que nous ne voulons pas prendre d'instantanés. Ce que nous voulons avoir est un système continu qui fonctionne pour longtemps et fondamentalement donne les indicateurs que nous recherchons tous les mois, qui publie une nouvelle valeur et, bien sûr, publie la valeur des derniers mois ainsi afin que nous puissions estimer des tendances. Parce que nous croyons que les tendances sont presque aussi importantes que la valeur réelle.

C'est la raison pour laquelle, ce faisant, nous investissons beaucoup dans l'automatisation. Fondamentalement, nous sommes en train de mettre en place des sondes à différents endroits, et nous faisons une rétroaction constante et une automatisation. Donc, le site Web qui publie des données est

automatisé, de même que les indicateurs sont automatiquement produits chaque mois, etc.

Dans les diapositives que nous vous présentons, nous allons vous donner les mesures. Et pour la même raison que nous ne voulons pas être impliqués dans les politiques, nous voulons vous donner les mesures telles qu'elles sont. Chaque fois que vous voyez un nombre, vous dites, « Oh, le widget X est maintenant à 29 %. Pourquoi ça ? » Eh bien, notre réponse générique est, « nous ne savons pas pourquoi. » Je veux dire, nous avons bien certaines idées, mais vos idées sont à peu près aussi valables que les nôtres. Et donc, nous ne voulons pas indiquer nos idées avec la publication des indicateurs. Les indicateurs sont uniquement des mesures.

Un de nos autres principes est que nous sommes très prudents et ne voulons pas avoir de problèmes de confidentialité. Donc, toutes les données que nous publions sont de nature statistique. Tous nos outils sont en source ouverte, et tous nos résultats sont publiés afin qu'ils puissent être analysés.

Nous avons eu quelques présentations dans les séances précédentes. Nous avons déjà eu la présentation des indicateurs ITHI à Abu Dhabi, par exemple. Pour nous, ils rentrent dans sept catégories. Un, nous nous penchons sur l'exactitude des données WHOIS. Nous regardons ensuite le comportement des

serveurs racine et le niveau d'utilisation malveillante qu'ils subissent dans une certaine mesure. Excusez-moi, l'utilisation malveillante des noms de domaine, l'utilisation malveillante du système des noms de domaine. Nous étudions le trafic des zones racine du DNS.

Pour tous les indicateurs qui sont énumérés ici, nous avons des sources de données. Par exemple, pour le service WHOIS, nous travaillons pour le département de la conformité de l'ICANN. Pour les utilisations malveillantes des noms de domaine, nous travaillons pour le projet DAAR. Pour mesurer le trafic de la zone racine, les serveurs récursifs, les registres IANA pour les paramètres DNS, et du déploiement DNSSEC, nous travaillons avec des scans du trafic de la zone racine ou avec des scans de trafic de résolveurs récursifs. Et nous collaborons avec les résolveurs récursifs pour [tester] efficacement la façon dont nous obtenons ces statistiques.

La chronologie, nous avons travaillé au cours de l'année dernière sur la définition des indicateurs. Ce que nous avons maintenant est une présentation des premières données. Au cours des deux derniers mois, nous avons mis en place les premières saisies, et nous avons été en mesure d'obtenir des données pour M1, M2, M3 et M7. Geoff Huston présentera les données pour M5 dans la prochaine présentation. Nous avons également été en mesure grâce à une collaboration rapide

d'obtenir un ensemble de données initiales pour les indicateurs M4 et M6, qui portent sur l'utilisation du DNS par les clients.

Ainsi, nous intégrons M5 au fur et à mesure de son développement. Nous allons construire le pipeline et obtenir un plus grand nombre de sondes de sorte que nous disposerons de données qui seront plus riches des indicateurs M4 et M6. Nous allons les enrichir et les publier sur le site Web de l'ITHI de l'ICANN.

Premier indicateur, M1. M1 suit l'exactitude des données du WHOIS. C'est ce que nous faisons en utilisant pour plus de précision un proxy qui est le nombre de plaintes. Nous ne prenons pas le [réel] nombre de plaintes. Nous prenons le nombre de plaintes qui ont été validées par le département de la conformité de l'ICANN. Actuellement, ce nombre s'élève à un peu moins de 6 par million. C'est notre première donnée, donc nous n'avons pas encore de tendance. Mais nous allons suivre cette tendance au fil du temps.

Avec toutes ces données, nous constatons que la moyenne ne fournit pas tant d'indications. Si je vous dis, il y a 6 plaintes par million de noms de domaine enregistrés en moyenne, eh bien, ce [nombre] n'est qu'une moyenne. Nous avons tracé une courbe qui est la [inaudible] fréquence des plaintes.

Essentiellement, le nombre total de plaintes sur l'axe Y et sur l'axe X le nombre de bureaux d'enregistrement classés de celui ayant le plus de plaintes à celui qui en a le moins.

Ce que nous voyons là est que la distribution n'est pas [inaudible]. Si tous les bureaux d'enregistrement avaient chacun autant de plaintes, vous pourriez voir une ligne droite sur la diagonale. Ce n'est pas ce que vous voyez. Ce que vous voyez est une ligne très incurvée, très inclinée vers l'axe Y. En fait, il faut six bureaux d'enregistrement pour compter au moins 50 % des plaintes. Je veux dire, ce n'est pas le nombre juste, six des bureaux d'enregistrement comptent pour un peu plus de 50 %. Il faut 44 bureaux d'enregistrement pour compter 90 % des plaintes. Ceci sur un total de près de 2 000 bureaux d'enregistrement. Il existe donc ici une distribution très [asymétrique].

Comme je l'ai dit, il s'agit d'un nombre de [inaudible]. Il ne s'agit pas d'un jugement ou d'un raisonnement sur la cause de ce nombre. Mais c'est ça que nous observons.

CATHY PETERSON : Excusez-moi, Christian. Nous avons une question en ligne.

CHRISTIAN HUITEMA : Oui ?

CATHY PETERSON : De Kathy Kleiman, « Comment savez-vous que les plaintes concernant le WHOIS sont valides ? Nous savons que certaines sont uniquement une forme de harcèlement. »

ALAIN DURAND : Je vais répondre à cette question. Nous avons travaillé en étroite collaboration avec le département de la conformité de l'ICANN. Nous n'examinons pas toutes les plaintes. Nous considérons seulement les plaintes qui sont liées à l'exactitude des données. Il existe de nombreux autres types de plaintes que nous ne prenons pas en considération.

Le département de la conformité de l'ICANN a un processus où ces plaintes sont examinées et évaluées. S'ils pensent qu'elles sont suffisamment étayées, ils envoient ce qu'ils appellent un premier avis. S'il n'y a pas de réponse, ils envoient alors un deuxième avis, puis ils passent au troisième avis, puis éventuellement jusqu'à la [rupture]. C'est un processus qui est très bien défini, qui est bien documenté au sein du département de la conformité de l'ICANN.

Donc pour répondre à nouveau à cette question, nous ne prenons en compte que les plaintes qui sont liées à l'exactitude d'un enregistrement dans la base de données WHOIS et les

plaintes qui ont été validées, qui ont fait l'objet d'un premier avis.

CHRISTIAN HUITEMA : Merci Alain. Donc, c'est l'indicateur M1 qui porte sur la précision du WHOIS. Les séries d'indicateurs M2 parlent de l'utilisation malveillante du système des noms de domaine, et pour cela, nous travaillons avec le projet DAAR. Ils font le suivi des quatre types d'utilisation malveillante : le nombre de sites Web utilisés par des domaines d'hameçonnage, le nombre utilisé par des logiciels malveillants, le nombre de commande et contrôle de réseau zombie, et le nombre de domaines de spam. L'indicateur est défini comme le nombre de domaines avec une utilisation malveillante pour 10 000 noms de domaine.

Nous voyons ici les moyennes mondiales, qui sont fondamentalement sur l'ordre de 4 ou 3 pour les trois premiers types d'abus et d'une valeur beaucoup plus grande pour les domaines de spam parce que le spam est une activité très largement distribuée.

Maintenant de la même façon que nous l'avons vue pour M1, nous voyons aussi que ces moyennes ne nous disent pas tout. Si je regarde la distribution par TLD, nous voyons que par exemple lorsqu'il s'agit d'hameçonnage, un seul TLD générique représente plus de 50 % de tous les domaines d'hameçonnage.

Et il ne faut que 11 TLD génériques pour prendre en compte tous les domaines d'hameçonnage. Nous voyons le même genre de distribution asymétrique pour les autres domaines. C'est donc clairement une indication de la structure du problème.

Nous avons essayé de faire la même mesure pour les bureaux d'enregistrement, mais nous ne voulons pas consacrer trop de temps aux données des bureaux d'enregistrement, car nos données de bureaux d'enregistrement doivent être évaluées avec le processus de WHOIS et sont soumises à toutes les restrictions liées à l'utilisation des données WHOIS comme la limitation dans le temps et tout cela. Donc nous ne publions sérieusement que lorsque nous recevons des données que nous pouvons vérifier, en qui nous pouvons avoir confiance, et c'est aujourd'hui encore un peu trop tôt.

Mais nous avons l'intention de présenter cette asymétrie des données dans quelques tableaux comme celui-ci qui dit, d'accord, combien de gTLD prend-il en compte pour au moins 50 % des domaines d'hameçonnage, de domaines de logiciel malveillant, etc., puis combien nous faut-il prendre en considération pour au moins 90 % de ces variations. Comme je l'ai dit, nous sommes ici pour mesurer des choses. Nous ne faisons aucune interprétation, et nous ne faisons pas de raisonnement pour savoir pourquoi il en est ainsi.

Les données M1 et M2 sont produites par le service de la conformité de l'ICANN et par le projet DAAR, et ils concernent la qualité des données. Les données M3 et M4 que nous allons voir plus tard portent sur le trafic réel du DNS, ce que nous y voyons. M3 parle du trafic racine. Nous mesurons le trafic racine en instrumentalisant la racine L. Nous faisons simplement un échantillonnage par jour par serveur racine L. Ces échantillons sont prélevés de manière aléatoire, pour qu'ils rendent compte de toutes les variations horaires lorsque nous les agrégeons statistiquement. Puis nous recevons tous ces échantillons et les résumons tous les mois pour obtenir ces indicateurs.

Ce que vous voyez là est le premier indicateur : combien des demandes de la racine reçoivent une réponse « domaine inexistant » ? Et c'est assez important. C'est effectivement pratiquement les deux tiers du trafic de la racine, les demandes qui n'ont pas de valeur particulière. Puis pour les demandes restantes, nous examinons comment bon nombre de ces demandes pourraient avoir été mises en cache par le résolveur. Encore une fois, nous voyons que c'est une bonne part, près de 30 %. Les demandes dont nous ne savons pas si elles pourraient avoir été mises en cache, elles ne le pourraient probablement pas, c'est de l'ordre de 6-6.5 %. Nous suivons cela chaque mois. Vous voyez ici la valeur actuelle et la moyenne et le camembert qui montre comment elles se répartissent dans les domaines.

Pour les demandes « Domaine inexistant » qui sont la grande partie du camembert sur le cercle, nous avons essayé de découper le camembert en composants. Quelles en sont les causes ? Nous avons constaté que nous examinions quatre composantes : les noms réservés, les noms qui ont été réservés par l'IETF, par exemple .local, et il y a cinq ou six de ceux-là, qui compte pour environ 3,4 % du trafic ; les chaînes faisant le plus fréquemment objet de fuite par exemple .home qui compte pour 9,3 % du trafic ; et le les tendances fréquentes, nous voyons une tendance dans les données. Ce ne sont pas des chaînes fréquentes. Chaque nom n'apparaît qu'une très petite fraction du temps. Il y a beaucoup, beaucoup de noms différents, mais ils suivent des tendances et nous essayons de tenir compte de ces tendances. Et puis tout le reste. Il y en a environ 10 % que nous ne pouvons pas expliquer directement par l'un ou l'autre de ces processus.

Pour les noms d'utilisation particulière comme définis dans RFC 6761, ce que nous constatons est que l'essentiel de l'utilisation se passe dans le domaine .local. C'est environ 2,77 % du trafic sur la racine d'aujourd'hui. D'autres domaines réservés sont présents, mais qui sont présents en nombre beaucoup plus restreint : .Localhost est assez présent, .invalid est aussi présent et les autres domaines montrent simplement des traces.

Sur les chaînes faisant fréquemment l'objet de fuites, ce que nous faisons ici est que nous recevons les chaînes qui sont les plus fréquentes à la racine et dans les variations actuelles dans la mise en œuvre actuelle nous considérons seulement les chaînes qui se produisent au moins 0,01 % des fois.

Sur cette diapositive particulière, je ne donne que des chaînes qui se produisent au moins 0,02 % des fois, parce que plus le nombre est faible, moins vous êtes sûrs des résultats. Et aussi parce que cela rendrait le PowerPoint très difficile à lire.

Encore une fois, nous voyons qu'il y a un nom qui domine cela, .home qui représente 3,5 % de ces demandes qui sont vues par la racine. Il y a ensuite une série d'autres noms. Le truc à emporter à la maison est que nous pouvons absolument mesurer la fuite de ces noms à la racine, et que nous pouvons la suivre, mois après mois, et nous savons quels noms sont utilisés et quels noms sont récurrents. Nous pourrons peut-être voir si cela se modifie un peu de mois en mois. Certains noms sont appelés à disparaître, mais nous pouvons voir qu'il y a un noyau de noms bien utilisés qui apparaissent tout le temps.

Je vous ai dit qu'un certain nombre de noms que nous voyons dans le trafic de la racine ne sont pas des domaines d'utilisation spéciale et ne correspondent pas à des chaînes fréquemment utilisées. Ce sont juste des noms au hasard. En fait, si vous

pouvez voir ici dans cette distribution, nous avons une distribution de ces noms par longueur. Nous voyons que la plupart de ces noms sont de 7 à 15 caractères de long. Nous n'avons pas inscrit les noms les plus longs parce qu'il y en a très, très très peu.

Beaucoup de ces noms d'une longueur entre 7 et 15 caractères, lorsque je les regarde en effectuant un échantillonnage aléatoire, ils ont l'apparence à des choses qui ont été générées de manière aléatoire par ordinateurs. Ils ne sont pas tous comme cela. Il est réellement très difficile de distinguer ce qui est généré de façon aléatoire par ordinateur et ce qui n'était qu'une sorte de plan de numérotage quelque part dans un réseau Wi-Fi, par exemple. Mais c'est quelque chose que nous voulons suivre et nous voulons aller plus loin et analyser plus encore.

C'est le trafic au niveau de la racine. Maintenant, quand nous avons fait la première étude l'an dernier, nous avons fait certaines expériences et nous sommes très rapidement arrivés à la conclusion que la racine n'est pas nécessairement représentative de l'ensemble du trafic par les utilisateurs. Si vous comprenez l'architecture du DNS, vous savez que ce qui est considéré comme la racine a déjà été filtré par les résolveurs DNS. En fait, si les résolveurs DNS appliquaient toute la technologie moderne définie par l'IETF, vous verriez très peu de

trafic à la racine. Ils mettraient en cache les bons résultats. Ils mettraient en cache les résultats [non satisfaisants]. Donc nous ne verrions rien de tout cela. Et donc une bonne partie de la circulation à la racine correspond à des comportements anormaux.

Si nous voulons nous pencher sur ce que les utilisateurs sont en train de faire, nous voulons être proches des clients. C'est pourquoi nous avons travaillé avec les résolveurs récursifs pour y mettre des sondes et essayer de regarder ce qui s'y passe. Combien de demandes émises par des clients vont plutôt vers des TLD enregistrés qu'à ces chaînes que nous voyons à la racine ? Combien vont vers ces noms réservés IETF ? Combien vont vers des chaînes fréquemment utilisées que nous voyons ici et quoi d'autre ?

Maintenant vous vous souvenez que lorsque nous examinons le trafic racine, nous voyons ces demandes vers des TLD inexistants représentent presque les deux tiers du trafic. Ici dans l'une des sondes que nous avons — et je dois valider nos données que nous n'en avons qu'en un seul point de mesure aujourd'hui. Nous sommes sur la voie d'en obtenir plus. Dans ce point de mesure particulier, ces TLD inexistants ne représentent que 1 % du trafic, beaucoup moins.

Les tendances sont également différentes. Dans les noms réservés où nous voyons une petite quantité de trafic avec .localhost, .local, et presque rien d'autre pour les autres noms.

Dans les chaînes fréquemment utilisées, cela a été pour nous un peu surprenant. C'est en fait dominé par des noms locaux, comme les noms de l'hôte que les gens essaient de résoudre et qu'ils ne font pas leurs demandes correctement et qu'ils finissent par envoyer leur demande en mettant le nom d'hôte comme un seul nom d'[objet] qui pourrait être pris par erreur pour un domaine de premier niveau. Nous ne publions pas la valeur de ces noms parce qu'ils sont liés à des questions de confidentialité. Ce sont généralement des noms dans l'infrastructure locale des personnes qui [fournissent] les sondes, donc nous les mettrons tous dans une seule catégorie globale de « noms de l'hôte local. »

Si nous allons au-delà, ce que nous voyons est un très faible trafic pour ces types de noms que nous voyons à la racine. Nous voyons une partie du trafic pour les grands noms comme .home, mais nous voyons le trafic pour des noms comme .dns, .internal, ou .unifi, qui dans ce cas représente le réseau Wi-Fi qu'ils utilisent. Cela a été pour nous une des leçons. Actuellement, nous voulons avoir beaucoup plus de sondes avant que nous puissions formuler des déclarations définitives, mais nous

voyons qu'il y a une différence entre le trafic au niveau de la clientèle et le trafic au niveau de la racine.

Vous vouliez intervenir ?

ALAIN DURAND : J'aimerais ajouter un petit point à ce que Christian vient de dire. Jusqu'ici, nous avons travaillé avec un certain nombre de petites organisations et nous avons déjà deux organisations qui ont accepté de participer et fournissent déjà des données. Je voudrais reconnaître leur travail ici.

L'une est l'Université de Cape Coast au Ghana, et une autre est l'Université de La Plata en Argentine. Nous travaillons aussi maintenant avec une troisième organisation Nawala qui est [plus ou moins] un fournisseur de service en Indonésie. Hier soir, nous avons travaillé très tardivement pour essayer de les aider à installer les outils pour faire toutes ces mesures.

Nous tendons la main à d'autres partenaires potentiels, et notre objectif est d'en obtenir davantage. Si nous pouvions en obtenir peut-être cinq, six, dix peut-être d'ici la fin de l'année, nous serions très heureux. Nous aimerions obtenir différents types d'acteurs, certains qui feraient partie de domaines académiques, d'autres qui pourraient être dans l'industrie,

quelques-uns des fournisseurs de services et certains pourraient être petits, grands ou très grands.

Mais nous commençons par une approche de noyau [inaudible]. Nous avons commencé petit. Cela nous a permis de comprendre comment les choses fonctionnaient vraiment pour perfectionner les outils que nous avons. Maintenant nous élaborons un processus pour rendre cette démarche plus automatique. Nous pouvons passer à de plus grands joueurs et, je l'espère, à un certain moment vers de plus grands acteurs encore.

Alors, je voulais remercier Christian pour [avoir écrit] l'outil et avoir aidé tout le monde à le déployer.

CHRISTIAN HUITEMA : Merci. Eh bien, Alain a réellement passé beaucoup de temps à les déployer, lui aussi. L'idée globale de cette infrastructure mondiale est que vous passez beaucoup de temps en appels téléphoniques ou en discussions sur l'ordinateur au milieu de la nuit. Mais cela fait partie du boulot, je dirais.

Ainsi donc, M3 et M4 sont les analyses de deux parties du trafic. Quel type de trafic du DNS voyons-nous à la racine et du côté client ? Avec les données M4, ce que nous voulions aussi voir était la valeur et l'utilité de tous ces registres IANA que nous faisons pour l'IETF ? Nous ne pouvons pas suivre tous les

registres IANA, car nous avons uniquement des données [liées] au DNS. Mais ce que nous avons fait, pour les tableaux [liés] au DNS, nous avons regardé les paramètres qui font partie des registres. Par exemple, le type [r] ou les classes de [code r], mais aussi des paramètres utilisés par les DNSSEC ou les paramètres utilisés par la DANE.

Pour ces paramètres, nous avons voulu répondre à deux questions. D'une part, les gens utilisent-ils vraiment les données qui sont enregistrées ? Fondamentalement, ce que nous avons fait, nous nous sommes dit : « d'accord, si un tableau définit dix valeurs, combien de ces valeurs sont-elles visibles au moins une fois dans notre ensemble de données ? » Pour certains tableaux, la réponse est zéro. Il y a quelques tableaux comme cela.

Pour les tableaux classiques comme les classes du DNS ou l'algorithme de nombres, la réponse est entre 20 % et 70 %. Certaines valeurs sont rarement utilisées. Par exemple, pour les algorithmes de sécurité, certains de ces algorithmes de sécurité sont obsolètes et les gens ne les utilisent plus. Mais nous pouvons le voir. Cela nous donne une idée et la confiance que ce que l'IANA est en train de faire est utile.

L'autre chose que nous voulions voir est de savoir si les gens contournaient les enregistrements de l'IANA et prenaient directement leurs propres valeurs. Dans cet ensemble, nous

voyons cela uniquement que pour les codes de l'option DNS, le code de l'option DNS0 EDNS, bien qu'il y ait quelques utilisations des valeurs expérimentales que nous voyons dans le tas. Donc, globalement, c'est comme ça.

J'aimerais maintenant faire une remarque sur l'utilisation de certificat TLSA et les certificats générés DANE. Dans mes données, je ne les vois pas. Donc j'ai eu une longue conversation avec Victor Dukhovny à ce sujet. Il m'a dit que c'était normal, car la plupart des utilisations de DANE se passe entre un serveur de messagerie et les serveurs faisant autorité. Le serveur de messagerie fera ses demandes directement au serveur faisant autorité. Alors, le trafic ne sera pas saisi par nos points de sonde. Je suis en train de travailler avec lui pour obtenir une alimentation directe du trafic qu'il a dans ses mesures de DANE pour que nous puissions évaluer correctement l'utilisation des tableaux DANE.

C'est fondamentalement la façon dont nous pouvons utiliser ces mesures pour suivre l'IANA. Nous ne suivons pas qu'un seul tableau. J'ai fourni les données concernant quatre ou cinq tableaux dans la diapositive précédente. Ici, voici l'ensemble de la liste de ce que nous faisons là et nous pourrions rajouter éventuellement plusieurs tableaux à la liste lorsque nous déterminerons comment analyser les données et les extraire.

Le dernier indicateur, M7, concerne le déploiement des DNSSEC. Nous avons commencé cette évaluation de déploiement des DNSSEC par l'analyse de la zone racine pour voir combien de TLD ont fourni une clé DNS. Ce nombre reste relativement stable aux environs de 90 %. Mais nous espérons que cela va changer au fil du temps et atteindre 100 %, mais le changement est très lent.

Maintenant, en analysant les données M4, nous nous sommes rendu compte que nous observions une grande partie du trafic qui était en fait le trafic de la sécurité du DNS. Nous voyons que parce que nous pouvons remarquer que lorsqu'un client utilise la sécurité du DNS, ils placent des morceaux de DO dans les demandes qui [inaudible] dans la réponse. Nous pouvons donc mesurer la fraction des demandes qui ont ce morceau et dire, « Hé, si nous pouvons trouver le client qui fait ça, nous saurons que tant de clients utilisent les DNSSEC. »

Donc nous pouvons ajouter à ces données ce que nous voulons faire en M7.2, qui est le pourcentage des demandes DNSSEC provenant de clients qui utilisent les DNSSEC. Si nous sommes vraiment ambitieux, nous pourrons certainement également voir le pourcentage de demandes des résolveurs récursifs qui utilisent les DNSSEC et fait intéressant, le pourcentage de réponses des serveurs faisant autorité qui offrent des réponses DNSSEC. Je pense qu'en faisant cela, nous aurons une bonne

idée de l'utilisation réelle des DNSSEC et serons en mesure de répondre à cette question, combien de DNSSEC sont utilisées aujourd'hui ? Je pense que c'est quelque chose qui serait très intéressant pour la communauté.

Je viens donc de présenter six de nos sept indicateurs. Geoff Huston viendra présenter l'indicateur 5 après moi. M7, comme je l'ai dit, est très stable, de sorte que ce type de graphique ne nous indique pas grand-chose aujourd'hui.

Je tiens à vous remercier pour votre attention et je vais maintenant prendre toutes les questions que vous avez, si vous en avez.

RUBENS KUHL : Rubens Kuhl, .br. J'aimerais faire des commentaires sur le fait que le serveur récursif, le DNS récursif, les indicateurs récursifs sont basés sur trois serveurs récursifs. Et nous avons actuellement 50 000 systèmes [inaudible] sur l'Internet, de sorte que publier les résultats jusqu'à ce que nous obtenions au moins 5 000 serveurs récursifs du DNS n'est probablement pas quelque chose à faire puisque cela n'a aucune pertinence statistique de quelque nature que ce soit. C'est comme mettre un [inaudible] sou le microscope et déduire tous les tissus dans le monde l'ont comme base.

[Donc cela m'amuse] qu'un tel indicateur puisse être publié par l'ICANN, en particulier dans un domaine pour lequel l'ICANN n'a pas de données directes sur des données différentes de la racine qui fait autorité alors qu'elle fait fonctionner l'un des systèmes de serveur racine le plus complet parce que c'est [inaudible] instance. Alors une racine possède une signification statistique excellente parmi les demandes de racine. Mais comme pour les demandes récursives, nous ne devrions pas les publier du tout jusqu'à ce qu'elles dépassent un véritable seuil de pertinence statistique.

CHRISTIAN HUITEMA : C'est vraiment un très bon commentaire. Nous avons eu recours à toutes sortes de mises en garde dans cette présentation pour expliquer que nous n'avions actuellement qu'un seul point de mesure et ce que nous voulions faire [inaudible]. Il est clair que nous voulons plus de points. Je ne sais pas si nous en avons besoin de 5 000. Je serais heureux d'en avoir 5 000, mais je ne sais pas si nous en avons besoin de 5 000.

Ce que j'ai l'intention de le faire est de comparer les données provenant de nombreux sites quand ils s'enregistrent pour voir ce qu'ont de différent et en commun. L'idée étant que nous savons qu'il existe des différences. Il existe des différences dans le temps, comme leurs différences entre la matinée et le soir. Ils

sont différents entre le week-end et les jours ouvrables. Nous savons qu'il existe des différences selon leur géographie. Les gens ne demandent pas vraiment le même trafic en Chine et en Amérique. Nous savons qu'il existe des différences selon le type de profession. Les gens ne peuvent pas demander la même demande dans le secteur universitaire et gouvernemental ou dans une entreprise ou un réseau privé ou un réseau mobile. Il est donc évident que nous voulons avoir des données représentatives de tous ces cas.

La signification statistique est quelque chose que nous allons gérer. C'est certainement quelque chose que nous voulons faire. Mais il faut bien commencer quelque part, nous n'en sommes donc vraiment qu'à la collecte de données. Et nous comparerons les sources pour que nous puissions répondre à votre question.

RUBENS KUHL : Oui, mais j'aimerais répondre à cela. Bien que ces mises en garde soient présentes, elles le sont habituellement en toutes petites lettres. Ainsi toute personne qui lit ce qu'on trouve dans le rapport le répétera et le publiera dans la presse et les médias sociaux, mais ne reproduira pas la mise en garde indiquant que cette donnée est en fait dénuée de sens. Donc, en réalité, publier

ces chiffres, c'est rendre un mauvais service à la communauté. C'est ce dont je voulais vous faire part.

Une observation que j'avais sur un autre sujet est qu'il avait été mentionné que certaines des demandes de bureaux d'enregistrement ont été touchées par les limitations dans le service WHOIS et ainsi de suite. Il y a des données que je peux recueillir de tous les registres [détaillés] qui forment le BRDA qui sont les données d'enregistrement résumées en masse. Ces données d'enregistrement résumées contiennent l'indication du bureau d'enregistrement associé à ce domaine. Donc il n'y a pas besoin de faire une demande au WHOIS. Il y a déjà des données à l'intérieur de l'ICANN qui fournissent cette information avec 100 % de précision, vous devriez donc avoir envie de vous occuper de cela également.

CHRISTIAN HUITEMA : C'est une suggestion intéressante. En fait c'est un bon point. J'aimerais d'abord y répondre en disant que le projet ITHI est un client du projet DAAR. Nous obtenons les données du DAAR donc, quelle que soit la décision prise par le DAAR, nous en héritons. Alors, premièrement, j'aimerais vous suggérer de rediriger votre question aux personnes faisant tourner le DAAR.

Deuxièmement [et je vais quelque peu] essayer de parler pour eux. Ma compréhension est qu'ils voulaient une étude qui soit

reproductible, ce qui signifie que quelqu'un en dehors de l'ICANN pourrait effectivement reproduire exactement la même étude, une méthodologie ouverte et des données qui sont accessibles. Les données que vous avez mentionnées peuvent être accessibles ou non depuis l'extérieur, et cela mettrait l'ICANN dans une position unique en étant le seul à être en mesure de faire cette étude. Leur choix a été de ne pas aller dans cette direction. Ils peuvent modifier leur approche dans un proche avenir et peut-être que John Crain pourrait dire un mot à ce sujet, mais jusqu'ici, c'est la direction que nous avons prise. Et en tant que leur client, nous avons hérité de cette décision.

La question est donc, pourquoi devrions-nous nous fier au WHOIS pour attribuer ceci à un bureau d'enregistrement en opposition à l'utilisation de données internes à l'ICANN ?

JOHN CRAIN : Si nous avons toutes ces données disponibles en interne, je ne les ai pas trouvées. Ce serait génial. Mais une des choses que nous avons essayé de faire, nous avons voulu rendre cela reproductible par d'autres personnes, ce qui signifie utiliser des sources externes. La seule chose dont nous avons besoin d'obtenir du WHOIS est l'ID du bureau d'enregistrement. Nous avons effectivement parlé précédemment de sources qui pourraient être internes, donc, il se peut que nous devions

réellement changer parce que je pense que le WHOIS puisse simplement ne pas être pratique.

RUBENS KUHL : Cette source s'appelle vraiment BRDA, vous pourriez avoir envie de regarder ou de pirater ces serveurs et de récupérer leurs données. Mais même si vous utilisez ces données, cela rend toujours les choses reproductibles, mais cela rend juste plus compliquée pour d'autres personnes d'utiliser vraiment les demandes WHOIS. Mais ils peuvent les reproduire en utilisant des demandes du WHOIS parce que c'est la même information. Ce n'est en aucun cas une information privilégiée.

JOHN CRAIN : Ouais, c'est entendu. Lorsque nous avons commencé le projet, nous pensions vraiment le faire exactement comme une personne extérieure pourrait le faire. Et vous avez raison à propos de la reproductibilité, donc, nous réexaminons la situation.

RUBENS KUHL : D'accord.

ALAIN DURAND : Très bien. Y a-t-il des questions dans le tchat, Cathy ?

CATHY PETERSON : Non.

ALAIN DURAND : Pas de question ? D'accord. Alors je vous remercie, Christian.

CHRISTIAN HUITEMA : Merci.

ALAIN DURAND : Je vous remercie beaucoup d'avoir montré ces chiffres ici pour la première fois. J'aimerais maintenant inviter Geoff Huston à prendre la parole. Geoff est-il dans la salle ?

CATHY PETERSON : Il y a une question par là.

ALAIN DURAND : Oh, nous avons une question.

SEBASTIEN BACHOLLET : Comme Geoff n'est pas dans la salle, je voulais juste poser une question, et ce, de quelqu'un sans antécédents techniques. Existe-t-il un lien entre ce que vous faites et certaines des questions sur le roulement des clés et les données dont ils ont

besoin pour comprendre ce qui se passe ? Vraiment désolé, c'est une question [inaudible] de [inaudible]. Merci.

CHRISTIAN HUITEMA : Aujourd'hui, la réponse est non. Nous n'avons aucun lien. Ce n'est pas un indicateur auquel nous avons pensé au départ. Maintenant, dans l'avenir il y aura peut-être plus de roulements et ils pourront être plus ou moins fréquents, cela pourrait devenir quelque chose que nous souhaiterons suivre. Donc, aujourd'hui, nous avons parlé de sept indicateurs. Nous pensons que nous les comprenons suffisamment bien pour être en mesure de les mesurer. Nous pensons maintenant à la deuxième phase où nous allons ajouter d'autres indicateurs, que nous allons examiner d'autres types de problèmes. Et ce domaine peut être l'un des domaines que nous devrons examiner et ajouter à ce que nous faisons maintenant. [Réponse en français aussi]

SEBASTIEN BACHOLLET : Je vois où vous voulez en venir. Je veux seulement être sûr que ma question a été bien expliquée, et désolé pour cela. Il semble aujourd'hui qu'il nous manque des données permettant d'être sûr que c'est le bon moment pour faire le roulement des clés. Il n'est pas le fait du moment où nous ferons un roulement des clés, c'est chaque année, vous serez en mesure de recueillir des

données [comme une question] qui, avec vos données de projet, pourront être utilisées par les personnes qui en ont besoin pour décider du moment pour faire le roulement des clés.

CHRISTIAN HUITEMA : Aujourd'hui, nous n'avons pas de données permettant de les aider.

SEBASTIEN BACHOLLET : Merci.

PAT KANE : Bonjour. Pat Kane de Verisign. Juste un suivi de la question de Sébastien. Plus tôt dans la journée, le CTO de l'ICANN a corrélé une diminution des demandes des DNSSEC à une poussée des roulements KSK de l'automne dernier jusqu'en octobre prochain. Je pense donc qu'il est important que nous comprenions la façon dont l'utilisation des DNSSEC se rattache à cela pour éclairer cette décision, car il y a beaucoup de données en termes de personnes avec les résolveurs qui n'ont pas les deux paires de clés, qui sont pires qu'elles ne l'étaient à la fin de l'année dernière. Il serait donc vraiment bien d'obtenir cette information de David plutôt tôt que tard. Merci.

ALAIN DURAND : Merci. C'est un très bon point. Comme nous l'avons vu, Christian parlait d'un nouvel indicateur M7.2 qui suivra vraiment le nombre de demandes requêtes qui ont l'ensemble de morceaux de DO. Cela peut aider dans ce sens avec d'autres indicateurs que nous essayons de concevoir dans cet espace particulier. Nous pourrions peut-être avoir une conversation hors ligne si vous avez des idées précises sur ce que nous devrions suivre.

Est-ce que Geoff est de retour dans la salle ? OK, je m'en excuse. Nous avons perdu un de nos intervenants. Je peux seulement vous parler brièvement de ce que nous comptons faire.

L'indicateur M5 a été initialement un des indicateurs examinant aussi les résolveurs, mais davantage d'un point de vue du client. Nous avons demandé à Geoff d'examiner cela, et Geoff a un système de mesure basé sur Google Ads qui est bien connu et nous l'avons déjà utilisé dans d'autres contextes. Nous lui avons demandé d'utiliser le système pour explorer ce qui peut être de la perspective d'un client, les résolveurs minimums.

Une des choses que nous voudrions examiner est de savoir si les résolveurs font effectivement de la mise en cache des choses ? Parfois, nous pensons qu'ils le font. Parfois nous pensons qu'ils ne le peuvent pas ou qu'ils ne peuvent mettre en cache que pour une période plus courte ou qu'ils peuvent mettre en cache pour

une plus longue période de temps. Nous pensons donc que nous pouvons avoir des mesures de tout cela.

Nous pouvons également examiner certaines des distributions des DNSSEC et IPv6 pour découvrir si le résolveur est configuré avec des DNSSEC ou non ou s'il est capable d'utiliser IPv6 ou pas. Nous pourrions également potentiellement trouver les résolveurs les plus utilisés. Les résolveurs les plus utilisés, il faut nuancer parce que le système s'appuie sur Google Ads, donc, ceci est utilisé par des utilisateurs physiques et non des machines. De sorte qu'il ne va pas saisir des communications de machine à machine, mais d'utilisateur à machine.

Cela pourrait être un projet [inaudible] de mesure qui pourrait également nous renseigner sur les roulements de clés et combien de résolveurs sont vraiment nécessaires pour couvrir 95 % ou quelque soit le pourcentage de la population que nous souhaiterons couvrir.

Il s'agit de travaux en cours. [Ce] sont de nouveaux indicateurs que Geoff voudrait proposer. Dans le même esprit que celui que Christian a décrit précédemment, nous voulons rendre cela automatique de sorte que nous puissions recevoir les mesures et les suivre au fil du temps pendant plusieurs années.

Donc en résumé, c'est le projet que nous aimerions faire avec Geoff.

Je vous prie de l'excuser de ne pas présent, il doit y avoir eu des circonstances extérieures.

S'il n'y a plus de questions, alors nous allons tout simplement terminer cette séance plus tôt. Oh, une question maintenant.

RUBENS KUHL : En fait, c'est plus une réponse au commentaire de Pat Kane. Pourquoi le nombre de résolveurs des DNSSEC signalés a augmenté comparé au nombre de KSK 2010 signalés [inaudible] ? Nous ne savons pas encore si ce sont des résolveurs validés ou non. Cela pourrait donc être quelqu'un qui en fait n'a que la clé racine, mais n'effectue pas la validation. Ce n'est donc pas un problème possible lorsque la clé roule. Donc, si nous faisons une étude de ce genre dans un indicateur, nous devrions probablement étudier la validation des résolveurs avec les anciennes clés et pas seulement les résolveurs signalant de vieilles clés. Parce que ce n'est pas quelque chose qui mesure quoi que ce soit qui pourrait prédire ce qui se passera lorsque nous faisons rouler la clé racine.

ALAIN DURAND : C'est un très bon point, mais j'ajouterai à cela. Nous devrions nous efforcer de bien peser ceci par le nombre d'utilisateurs qui sont derrière ce résolveur. Si c'est seulement quelque chose qui

est utilisé dans le sous-sol de quelqu'un et mis en marche pendant cinq minutes, cela peut ne pas avoir la même importance qu'un résolveur qui dessert des millions de clients.

RUBENS KUHL : D'accord.

ALAIN DURAND : Donc, s'il n'y a pas d'autres commentaires, nous allons clore cette séance. La prochaine réunion de l'ICANN est une réunion de politiques, il n'y aura donc pas de séances techniques, et nous ne nous réunirons pas. Mais nous allons vous voir tous à Barcelone.

[FIN DE LA TRANSCRIPTION]