SAN JUAN – Indicadores de Sanidad de Tecnologías de Identificadores
Martes, 13 de marzo de 2018 – 17:00 a 18:30 hora estándar del Atlántico (AST)
ICANN61 | San Juan, Puerto Rico

CATHY PETERSEN: Buenas tardes a todos. Comenzaremos la sesión de Indicadores de Sanidad de Tecnologías de Identificadores en unos minutos. Les daremos un par de minutos más. Gracias.

ALAIN DURAND: Buenas tardes. Esta es la sesión de ITHI que significa Indicadores de Sanidad de Tecnologías de Identificadores. Este es un proyecto que se ha iniciado hace ya un tiempo, y hoy vamos a mostrar algunos números interesantes [inaudible]. Números que serán interesantes para ustedes. Fueron interesantes para mí.

En esta sesión, contaremos con tres presentadores. El primero será Paul Wilson, actual Presidente de la NRO. Nos proporcionará información actualizada sobre lo que la NRO ha estado haciendo en esta área.

El segundo y el tercero se enfocarán en la parte principal del proyecto, las cosas que administra la ICANN. La segunda presentación estará a cargo de Christian Huitema sobre las

métricas actuales y los datos que estamos encontrando sobre las métricas actuales.

La última presentación estará a cargo de Geoff Huston sobre el conjunto propuesto de [algunas] nuevas métricas.

Así que sin más preámbulos, le cederé la palabra a Paul, que hablará sobre las actividades en el espacio de números. ¿Paul?

PAUL WILSON: Gracias, Alain, y hola, a todos. Los registros regionales de Internet que todos tenemos son cinco, todos nosotros tenemos un registro de WHOIS que utiliza un servicio de WHOIS bastante familiar. Entonces hay cinco registros diferentes que son ejecutados por los cinco RIR. Están bastante bien coordinados entre sí. Técnicamente son capaces de tener inconsistencias y, por supuesto, errores y estar incompletos y demás. Por lo tanto, los RIR trabajan conjuntamente bajo la bandera de la NRO para asegurarse de que esos registros tengan sentido entre sí y que cumplan su propósito en términos de los registros que almacenan.

Lo hemos hecho durante mucho tiempo, pero creo que las cosas han cambiado un poco en los últimos años debido a que existe un interés mucho más elevado de un grupo mucho más amplio de partes que dependen de la corrección y eficacia de las bases

de datos. Además, el ritmo de las actualizaciones también está aumentando bastante. Entonces, antes de llegar a la escasez actual de direcciones IPv4, las asignaciones eran bastante estáticas. Se hicieron para las partes que conservaron esas asignaciones y las usaron. Pero en estos días, se producen muchas transferencias. Por lo tanto, dentro de las regiones y entre los registros regionales de Internet, hay muchas transferencias en curso, que obviamente requieren actualizaciones de la base de datos. Ese es otro motivo por el cual nuestro enfoque en la corrección e integridad y demás está en aumento.

Como digo, nos hemos preocupado desde el inicio de los RIR realmente para que los registros hagan su trabajo. No hemos hablado sobre la sanidad como tal, pero la idea de la sanidad de los identificadores es algo nuevo que creo que ha llegado con el proyecto de la ICANN. Pero dicho esto, hemos mantenido este enfoque.

El otro aspecto de esto, por supuesto, es que tenemos relaciones de membresía con los operadores de red que son el primer receptor de bloques de direcciones y ASN. Por lo tanto, están obligados por sus relaciones con cada uno de los RIR a mantener sus registros al día y actualizados. Existen cuestiones de política allí en términos de cuáles son exactamente las

expectativas y las sanciones, por así decirlo, por no cumplir con esas políticas.

Estas cosas se están manejando a nivel regional en general. Entonces, los cinco RIR tienen membresías y procesos de políticas independientes, y en diferentes momentos tendrán debates sobre políticas relacionadas con WHOIS. Esos debates, como mencioné, también están aumentando en frecuencia e intensidad en estos días. Diría que en las cinco regiones, creo que cabe decir que todos nos estamos ajustando de diferentes maneras pero, en general, en la misma dirección hacia políticas más claras e implementadas de una forma más estricta.

El proyecto de ITHI realmente no fue una sorpresa para la ICANN. Obviamente, es un interés compartido de todos los que desempeñamos funciones de registro y, por lo tanto, decidimos incorporarnos a la iniciativa de ITHI de la ICANN. Creo que en realidad fue bastante útil porque, aunque hemos trabajado juntos de una forma bastante estrecha, en realidad no hemos establecido un conjunto de métricas coherentes que ahora hemos avanzado a través del proyecto de ITHI.

De hecho, dedicamos un período el año pasado, pasó por nuestro grupo de coordinación de servicios de registración que es un grupo de personal de los cinco RIR que trabajan en las áreas de servicios de registración. Ese grupo hizo parte del

trabajo sobre un conjunto preliminar de métricas para lo que nos referiríamos a la sanidad de identificadores en el espacio de números. También lanzaron una consulta pública sobre un documento preliminar.

Eso sucedió hacia fines del año pasado, y brindó la oportunidad a nuestras comunidades de realizar aportes a ese proceso. De hecho, recibimos muy pocos comentarios, por lo que ahora tenemos un documento que está casi listo para su aprobación y publicación final. Básicamente, documenta el estado del identificador en el espacio de números en términos de registros de WHOIS. A lo que hemos llegado es a nuestros tres objetivos en materia de datos, que sean: completos, actuales y correctos.

Pero esos objetivos se dividen en cinco medidas específicas que son métricas cuantificables de nuestra base de datos que se relacionan con la integridad de la base de datos, la singularidad, la coincidencia de nuestra base de datos con otros registros oficiales externos, la efectividad de los datos involucrados para llegar realmente a las personas que están documentadas en el registro, y la actualización de los datos también. Ese documento también identifica los diversos riesgos asociados con no alcanzar nuestros objetivos en esas medidas, y analiza las causas que se asociarían con ese tipo de incumplimiento.

Ese documento será publicado en breve. Lo que todavía no tenemos, obviamente, dado que las métricas mismas todavía están en formato de borrador, todavía no tenemos lo que creo que Alain presentará en breve, que son datos reales sobre nuestro cumplimiento. Pero obviamente el punto de tener estas métricas es poder medir cosas para establecer algunos objetivos con los que esperamos cumplir y hacer un seguimiento del grado de cumplimiento durante un período de tiempo.

Entonces, creo que si están interesados en el espacio de números, miren este espacio y podremos informar a su debido tiempo sobre el estado de los identificadores de Internet en el espacio de números en los cinco RIR. Creo que eso es todo. Gracias, Alain.

ALAIN DURAND: Gracias, Paul. Me gustaría aprovechar esta oportunidad para agradecerles a ustedes y a los demás miembros de la comunidad de números por colaborar con nosotros en este proyecto. Creo que es una colaboración interesante, y estamos aprendiendo mucho en este proceso.

PAUL WILSON: Sí, estoy de acuerdo. Igualmente, Alain. Gracias.

ALAIN DURAND: Ahora estamos siguiendo el mismo proceso, como revisar el espacio [problemático] y definir las métricas y luego llegar a la medición real para que sea realmente normal. Pero aún no tienen los números, y esperamos con ansias ver en el futuro el primer lote de números cuando sea que estén listos.

Ahora vamos a cambiar de tema y pasaremos al espacio de nombres. Christian hará una presentación sobre dónde estamos.

CHRISTIAN HUITEMA: Buenas tardes. Soy Christian Huitema. He estado trabajando en la medición de los datos del DNS y su [estado] durante el último año y medio, aproximadamente, tras realizar un estudio para ver qué se podía hacer con el DNS y trabajamos con Alain en la creación de métricas reales.

Como dijo Paul, tiene que haber algunos principios cuando se establecen esas métricas. Los principios que hemos adoptado están prácticamente en esta diapositiva. En primer lugar, queríamos que esta operación de métricas fuera técnica. No queremos involucrarnos en las políticas. El propósito de las métricas es describir el estado del sistema. No es, básicamente, juzgar de una forma u otra.

Hemos estado buscando definir áreas que queremos monitorear que son potencialmente problemáticas, definir las métricas en estas áreas y definir las formas de medirlas.

Otro principio es que no queremos capturar instantáneas. Lo que queremos hacer es tener un sistema continuo que funcione durante mucho tiempo y, básicamente, que proporcione las métricas que tenemos como objetivo, cada mes publicar un nuevo valor y, por supuesto, publicar también el valor de los últimos meses para que podamos estimar tendencias. Porque creemos que las tendencias son casi tan importantes como el valor real.

Esa es la razón por la cual estamos invirtiendo mucho en automatización para hacerlo. Básicamente, estamos configurando sondas en varios lugares, y estamos realizando una constante retroalimentación y automatización. Por lo tanto, el sitio web que publica los datos está automatizado, para que las métricas se produzcan automáticamente cada mes, etc.

En las diapositivas que presentamos, vamos a darles las medidas. Y por la misma razón que no queremos involucrarnos en las políticas, queremos proporcionar las medidas tal como son. Cada vez que vean un número, y digan: "Ah, el widget X ahora está en el 29 %. ¿Por qué?" Bueno, nuestra respuesta genérica es: "No sabemos por qué". Es decir, tenemos

conjeturas, pero sus conjeturas son casi tan acertadas como las nuestras. Así que, no queremos poner esas conjeturas en la publicación de métricas. Las métricas son mediciones directas.

Otro principio es que tenemos mucho cuidado de no tener problemas de privacidad. Por lo tanto, todos los datos que publicamos son de naturaleza estadística. Todas nuestras herramientas son de código abierto y todos nuestros resultados se publican para que puedan analizarse.

Tuvimos un par de presentaciones en sesiones anteriores. Ya tuvimos la presentación de las métricas de ITHI en Abu Dhabi, por ejemplo. Se encuentran en siete categorías para nosotros. Primero, estamos observando la exactitud de los datos de WHOIS. Estamos observando el comportamiento de los servidores raíz y el nivel de uso indebido que están alcanzando en cierto grado. Disculpen, el uso indebido de nombres de dominio, el uso indebido del Sistema de Nombres de Dominio. Estamos observando el tráfico raíz del DNS.

Para todas estas métricas que se enumeran aquí, tenemos fuentes de datos. Por ejemplo, para WHOIS, estamos trabajando para el departamento de Cumplimiento de la ICANN. Para el uso indebido de nombres de dominio, estamos trabajando para el proyecto DAAR. Para la medición de tráfico raíz, servidores recursivos, registros de la IANA para parámetros del DNS y

despliegue de DNSSEC, estamos trabajando con análisis del tráfico raíz o con análisis del tráfico del resolutor recursivo. Y estamos colaborando con los resolutores recursivos para [sondear] de manera efectiva qué nos proporciona esas estadísticas.

Cronograma, hemos estado trabajando el año pasado en la definición de las métricas. Lo que tenemos ahora es una presentación de los primeros datos. En los últimos dos meses, hemos configurado las capturas iniciales, y hemos podido obtener datos para M1, M2, M3 y M7. Geoff Huston presentará los datos de M5 en la próxima charla. También hemos podido colaborar de manera anticipada para obtener un conjunto inicial de datos para las métricas de M4 y M6, que tratan sobre el uso del DNS por parte del cliente.

Entonces, integramos M5 a medida que se desarrolle. Vamos a construir la tubería y obtener más sondeos para que tengamos datos que sean más ricos a partir de las métricas de M4 y M6. Vamos a enriquecer eso y publicarlo en el sitio web de ITHI de la ICANN.

Primeras métricas, M1. M1 está monitoreando la exactitud de los datos de WHOIS. Lo estamos haciendo mediante el uso de un proxy para la exactitud, que representa la mayoría de los reclamos. No tomamos el número [real] de reclamos. Tomamos

la cantidad de reclamos que han sido validados por el departamento de Cumplimiento de la ICANN. En este momento, ese número se encuentra en un poco menos de 6 por millón. Esos son nuestros primeros datos, por lo tanto, aún no tenemos una tendencia. Pero vamos a monitorear esa tendencia con el transcurso del tiempo.

Con todos estos datos, lo que vemos es que el promedio no cuenta la historia. Si les digo que hay 6 reclamos por cada millón de dominios [números] registrados en promedio, bueno, eso es tan solo un promedio. Hemos trazado aquí una curva que es la [inaudible] frecuencia de reclamos. Básicamente, el total de reclamos en el eje Y, y luego en el eje X, el número de registradores clasificados desde el que tiene más reclamos hasta el que tiene menos.

Lo que vemos allí es que la distribución no es [inaudible]. Si cada registrador tiene tantos reclamos, verá una línea recta en la diagonal. Eso no es lo que ustedes ven. Lo que ven es que la línea es muy curva, muy inclinada hacia el eje Y. De hecho, se requieren seis registradores para representar al menos el 50 % de los reclamos. Es decir, no es un número par, seis registradores representan un poco más del 50 %. Se necesitan 44 registradores para representar el 90 % de los reclamos. Eso es en un total de casi 2000 registradores. Por lo tanto, hay una distribución muy [sesgada] allí.

Como dije, estos son [inaudibles] números. No es un juicio o un razonamiento de la causa. Sino que eso es lo que observamos.

CATHY PETERSEN: Discúlpeme, Christian. Tenemos una pregunta en línea.

CHRISTIAN HUITEMA: ¿Es así?

CATHY PETERSEN: De Kathy Kleiman, "¿Cómo saben que los reclamos de WHOIS son válidos? Comprendemos que algunos se hacen con fines de acoso".

ALAIN DURAND: Responderé esta pregunta. Hemos estado trabajando estrechamente con el departamento de Cumplimiento de la ICANN. No estamos observando todos los reclamos. Solo estamos examinando los reclamos relacionados con la exactitud de los datos. Hay muchos otros tipos de reclamos que no estamos teniendo en cuenta.

El departamento de Cumplimiento de la ICANN tiene un proceso en el que observan esos reclamos y los evalúan. Si consideran que existen fundamentos suficientes para dichos reclamos, entonces envían lo que denominan una primera notificación. Si

no hay respuesta, pasan a una segunda notificación y a una tercera notificación, y luego potencialmente a un [incumplimiento]. Entonces, ese es un proceso que está muy bien definido, que está bien documentado en el departamento de Cumplimiento de la ICANN.

Para responder a esta pregunta nuevamente, solo estamos examinando los reclamos relacionados con la exactitud de la base de datos de WHOIS de una registración y los reclamos que han sido validados, que han pasado por la primera etapa de notificación.

CHRISTIAN HUITEMA: Gracias, Alain. Esa es la métrica M1 sobre la exactitud de los datos de WHOIS. La serie de métricas M2 están relacionadas con el uso indebido del Sistema de Nombres de Dominio, y estamos trabajando con el proyecto DAAR para eso. Están monitoreando cuatro tipos de uso indebido: la cantidad de sitios web utilizados por los dominios de phishing, la cantidad utilizada por los dominios de malware, la cantidad de comandos y controles de botnet, y la cantidad de dominios de spam. La métrica se define como la cantidad de dominios abusados para 10 000 nombres de dominio.

Vemos allí los promedios globales, que son básicamente del orden de 4 o 3 para los primeros tres tipos de uso indebido y de

un valor mucho mayor para los dominios de correo no deseado porque el spam es una actividad ampliamente distribuida.

Ahora también, de la misma manera que clasificamos para M1, también vemos que esos promedios no cuentan la historia. Si observamos la distribución por TLD, vemos que, por ejemplo, cuando se trata de phishing, un solo gTLD representa más del 50 % de todos los dominios de phishing. Y tan solo se necesitan 11 gTLD para representar a todos los dominios de phishing. Vemos el mismo tipo de distribución sesgada para los otros dominios. Entonces, eso es claramente una indicación de la estructura del problema.

Hemos intentado hacer la misma medición para los registradores, pero no queremos dedicar demasiado tiempo a los datos del registrador porque nuestros datos de registrador deben evaluarse con el proceso de WHOIS y están sujetos a todas las restricciones del uso de los datos de WHOIS como en la limitación en el tiempo y todo eso. Así que solo publicamos seriamente cuando obtenemos datos que realmente podemos verificar que podemos confiar, y hoy es algo un tanto preliminar.

Pero tenemos la intención de presentar este sesgo de los datos en una especie de tabla como esta que dice, está bien, cuántos gTLD se necesitan para representar al menos el 50 % de los dominios de phishing, de los dominios de malware, etc. Y

entonces, cuántos se necesitan para contabilizar al menos el 90 % de esas variaciones. Como dije, estamos en el negocio de medir cosas. No hacemos ninguna interpretación, y no hacemos ningún razonamiento sobre por qué es así.

Los datos de M1 y M2 son generados por el departamento de Cumplimiento de la ICANN y por el proyecto DAAR, y se refieren a la calidad de los datos. Los datos de M3 y M4 que veremos más adelante están relacionados con el tráfico real del DNS, lo que vemos allí. M3 aborda el tráfico raíz. Medimos el tráfico raíz instrumentando la Raíz L. Básicamente, estamos realizando un muestreo por día por cada servidor Raíz L. Esas muestras se toman en momentos aleatorios, por lo que representan todas las variaciones de tiempo cuando las acumulamos estadísticamente. Luego, recibimos todas esas muestras y las resumimos todos los meses y recibimos esas mediciones.

Lo que ves allí es la primera métrica: ¿cuántas de las consultas raíz obtienen una respuesta de "no existe dicho dominio"? Y es una cantidad bastante grande. Es efectivamente casi dos tercios del tráfico raíz, consultas que no tienen ningún valor particular. Luego, en las consultas restantes, observamos cuántas de esas consultas podrían haber sido almacenadas en caché por el resolutor. De nuevo, vemos que es una proporción equitativa, casi un 30 %. Las consultas que no sabemos si podrían haber sido almacenadas en caché, probablemente no podrían ser, es

del orden del 6 al 6,5 %. Hacemos un seguimiento de eso todos los meses. Aquí pueden ver el valor actual y el promedio y el gráfico circular que muestra cómo se dividen los dominios.

Para las consultas de "no existe dicho dominio", que es esta gran parte del gráfico en este círculo, tratamos de dividir el gráfico en componentes. ¿Qué causa eso? Hemos encontrado que estamos viendo cuatro componentes: los nombres reservados, los nombres que han sido reservados por el IETF como .local, por ejemplo, y hay cinco o seis de ellos que representan aproximadamente el 3,4 % del tráfico; las cadenas de caracteres frecuentemente filtradas, como por ejemplo .home, que representan el 9,3 % del tráfico; y los patrones frecuentes, vemos un patrón en los datos. No son cadenas de caracteres frecuentes. Cada nombre aparece solo en una fracción muy pequeña del tiempo. Hay muchos, muchos nombres diferentes, pero siguen patrones e intentamos representar esos patrones. Y luego todo lo demás. Hay aproximadamente un 10 % que no podemos explicar directamente mediante ninguno de esos procesos.

Para los nombres de uso especial definidos en la RFC 6761, lo que vemos es que la mayor parte del uso es con el dominio .local. Se trata de un 2,77 % del tráfico en la raíz hoy. Otros dominios reservados están presentes pero en números mucho más pequeños: .localhost está bastante presente, .invalid está

bastante presente y luego los otros dominios son realmente vestigios.

En las cadenas de caracteres filtradas con frecuencia, lo que estamos haciendo aquí es que estamos obteniendo las cadenas de caracteres que son más frecuentes en la raíz y en la variación actual en la implementación actual solo estamos observando cadenas de caracteres que ocurren al menos el 0,01 % del tiempo.

En esta diapositiva en particular, solo incluyo esas cadenas de caracteres que ocurren al menos el 0,02 % de las veces porque cuanto menor sea el número, menos seguros estaremos de los resultados. Y también porque haría que el PowerPoint fuera muy difícil de leer.

Una vez más, vemos que hay un nombre que domina eso, que es .home, que representa el 3,5 % de las solicitudes que ve la raíz. Luego hay otra serie de nombres. La moraleja es que podemos medir absolutamente la filtración de esos nombres en las raíces, y podemos monitorearla mes a mes, y sabemos qué nombres están siendo utilizados y qué nombres son recurrentes. Podemos ver que cambia un poco de mes a mes. Algunos nombres van a desaparecer, pero podemos ver que hay un núcleo de nombres bien utilizados que aparecen todo el tiempo.

Les mencioné que varios de los nombres que vemos en este tráfico raíz no son dominios de uso especial y no se corresponden con cadenas de caracteres usadas frecuentemente. Son simplemente nombres aleatorios. De hecho, si miran aquí en esta distribución, hicimos una distribución de esos nombres por longitud. Vemos que la mayoría de esos nombres tienen entre 7 y 15 caracteres de longitud. Los nombres más largos no graficamos porque hay muy, muy pocos.

Muchos de esos nombres con una longitud entre 7 y 15 cuando se observan mediante un muestreo aleatorio, parecen cosas que han sido generadas aleatoriamente por computadoras. No son todos así. En realidad, es muy difícil distinguir lo que se genera aleatoriamente en una computadora y lo que es simplemente un tipo de plan de numeración en alguna parte de una red Wi-Fi, por ejemplo. Pero es algo que queremos monitorear y queremos ir más allá y analizarlo más a fondo.

Ese es el tráfico en la raíz. Ahora, cuando hicimos ese primer estudio el año pasado, realizamos algunos experimentos y llegamos rápidamente a la conclusión de que la raíz no era necesariamente representativa de todo el tráfico de los usuarios. Si comprenden la arquitectura del DNS, saben que lo que se ve como raíz ya ha sido filtrado por los resolutores del DNS. De hecho, si los resolutores del DNS aplicaran toda la

tecnología moderna definida por el IETF, se vería muy poco tráfico en la raíz. Guardarían en caché los buenos resultados. Guardarían en caché los resultados [insatisfactorios]. Entonces no veríamos nada de eso. Por lo tanto, gran parte del tráfico en la raíz corresponde a comportamientos anómalos.

Si queremos ver lo que los usuarios realmente están haciendo, queremos estar cerca de los clientes. Es por eso que hemos estado trabajando con resolutores recursivos para poner sondeos en resolutores recursivos y tratar de ver lo que está sucediendo allí. ¿Cuántas de las consultas emitidas por clientes van a TLD registrados en lugar de todas estas cadenas de caracteres que vemos en la raíz? ¿Cuántas van a estos nombres reservados del IETF? ¿Cuántas van a las cadenas de caracteres usadas con frecuencia que vemos allí y qué más?

Ahora recuerdan que cuando observamos el tráfico raíz, vemos que estas consultas a TLD inexistentes representan casi dos tercios del tráfico. Aquí, en el único sondeo que tenemos, y debo matizar que en nuestros datos tenemos solo un punto de medición hoy. Estamos aumentando para obtener más. En este punto de medición, estos TLD inexistentes representan tan solo el 1 % del tráfico, mucho menos.

Las tendencias también son diferentes. En los nombres reservados donde vemos una pequeña cantidad de tráfico con .localhost, .local y casi nada para los otros nombres.

En las cadenas de caracteres frecuentemente usadas, eso nos tomó un poco por sorpresa. En realidad, está dominado por nombres locales, como nombres de host que las personas intentan resolver y no hacen sus consultas correctamente y terminan enviando la [consulta] poniendo el nombre de host como un nombre de [identificador] único que podría confundirse con un dominio de alto nivel. No publicamos el valor de esos nombres porque existen cuestiones de privacidad. Normalmente son nombres en la infraestructura local de las personas que están [proporcionando] los sondeos, por lo que los incluimos en una categoría global de "nombres de host locales".

Si vamos más allá, lo que vemos es muy poco tráfico para este tipo de nombres que vemos en la raíz. Vemos algo de tráfico para los grandes nombres como .home, pero vemos tráfico para nombres como .dns, .internal o .unifi que, en ese caso, representa la red Wi-Fi que utilizan. Entonces esa fue una de las lecciones que aprendimos. En ese punto, queremos tener muchos más sondeos antes de poder hacer declaraciones definitivas, pero vemos que hay una diferencia entre el tráfico en los clientes y el tráfico en la raíz.

¿Deseaba intervenir?

ALAIN DURAND: Quisiera agregar un pequeño punto a lo que Christian acaba de decir. Hasta ahora, hemos estado trabajando con varias organizaciones pequeñas y ya tenemos dos organizaciones que aceptaron participar y ya están aportando datos. Me gustaría agradecer su aporte aquí.

Una de ellas es la Universidad de Cape Coast en Ghana, y otra es la Universidad de La Plata en Argentina. También estamos trabajando con una tercera organización, Nawala, que es [una especie de] proveedor de servicios en Indonesia. Anoche, nos levantamos muy tarde tratando de ayudarlos a instalar las herramientas para realizar todas esas mediciones.

Nos estamos acercando a otros socios potenciales, y nuestro objetivo es lograr que haya más participantes en esto. Si pudiéramos obtener tal vez cinco, seis, hasta diez quizás para fin de año, estaremos bastante contentos. Nos gustaría obtener diferentes tipos de actores, algunos que sean académicos, otros pueden ser más industriales, otros pueden ser proveedores de servicios, algunos pueden ser pequeños o grandes o muy grandes.

Pero estamos comenzando con un enfoque central [inaudible]. Comenzamos de a poco. Eso nos ha permitido comprender cómo funcionan realmente las cosas para perfeccionar las herramientas que tenemos. Ahora estamos desarrollando un proceso para hacer esto más automático. Podemos acudir a participantes más grandes y, con suerte, incluso en algún momento a participantes mucho más grandes.

Así que quería agradecer a Christian por [escribir] la herramienta y ayudar a todos a implementarla.

CHRISTIAN HUITEMA: Gracias. Bueno, de hecho, Alain dedicó mucho tiempo a implementarla también. La cuestión sobre esta infraestructura mundial es que uno se pasa mucho tiempo haciendo llamadas telefónicas o chats en la computadora en medio de la noche. Pero eso es parte del territorio, diría yo.

Entonces, M3 y M4 son análisis de dos partes del tráfico. ¿Qué tipo de tráfico del DNS vemos en la raíz y en el lado del cliente? Con los datos de M4, también queríamos ver qué tan buenos y útiles son todos estos registros de la IANA que estamos haciendo para el IETF. No podemos rastrear todos los registros de la IANA porque solo tenemos datos [relacionados] del DNS. Pero lo que hicimos fue para que las tablas [relacionadas] del DNS, observen los parámetros que son parte de los registros. Por ejemplo, los

tipos [r] o las clases [r code], pero también los parámetros utilizados por DNSSEC o los parámetros utilizados por DANE.

Para esos parámetros, queríamos responder dos preguntas. Una es, ¿las personas realmente usan los datos que están registrados? Básicamente lo que hicimos, dijimos fue: "Está bien, si una tabla define diez valores, ¿cuántos de esos valores realmente vemos al menos una vez en nuestro conjunto de datos?" Para el caso de algunas tablas, la respuesta es cero. Hay unas cuantas tablas como esas.

Para las tablas clásicas como las clases del DNS o los números de algoritmo, la respuesta es entre 20 % y 70 %. Algunos valores rara vez se usan. Por ejemplo, en los algoritmos de seguridad, algunos algoritmos de seguridad están obsoletos y las personas ya no los usan. Pero podemos ver eso. Eso nos da una idea y una confianza de que lo que la IANA está haciendo es útil.

Otra cosa que queríamos ver es si las personas eludían el registro de la IANA y creaban directamente sus propios valores. En ese conjunto, solo vemos eso en el caso de los códigos de opción del DNS, el código de opción del DNS EDNS0, si bien hay algún uso de valores experimentales que vemos en su estado natural. Así que globalmente eso es lo que es.

Ahora, me gustaría agregar un comentario allí sobre el uso del certificado de TLSA y el certificado de DANE que se genera. En

mis datos, no los veo. Así que tuve una larga conversación con Victor Dukhovny al respecto. Me dijo que eso era normal porque la mayor parte del uso de DANE es entre un servidor de correo y servidores autoritativos. El servidor de correo consultará directamente al servidor autoritativo. Para que el tráfico no quede atrapado en nuestros puntos de sondeo. Estoy trabajando con él para obtener una alimentación directa del tráfico que tiene en sus mediciones de DANE, de modo que realmente podamos evaluar adecuadamente el uso de las tablas de DANE.

Básicamente, esta es la forma en que podemos utilizar esas mediciones para monitorear a la IANA. No monitoreamos tan solo una tabla. Proporcioné los datos para cuatro o cinco tablas en la diapositiva anterior. Aquí está la lista completa que estamos haciendo allí, y podríamos agregar con el tiempo más tablas a la lista cuando descubramos cómo analizar los datos y extraerlos.

La métrica final, M7, trata sobre la implementación de DNSSEC. Comenzamos esta evaluación de la implementación de DNSSEC analizando la zona raíz para ver cuántos TLD estaban proporcionando una clave de DNS. Esa cifra es bastante estable en algo así como el 90 %. Pero esperamos que cambie con el tiempo y llegue al 100 %, pero cambia muy lentamente.

Ahora, cuando analizamos los datos de M4, nos dimos cuenta de que estábamos viendo una gran parte del tráfico que en realidad era tráfico de seguridad del DNS. Vemos que, como podemos observar cuando un cliente utiliza la seguridad del DNS, coloca un bit DO en las consultas que [inaudible] en la respuesta. Entonces, podemos medir la fracción de consultas que tienen ese bit y decir: "Oigan, si podemos encontrar al cliente que hace eso, sabemos que muchos clientes están usando DNSSEC".

De modo que podemos agregar a estos datos lo que queremos hacer en M7.2, que es el porcentaje de consultas de DNSSEC de clientes que utilizan DNSSEC. Si somos realmente ambiciosos, también vemos con certeza el porcentaje de consultas de los resolutores recursivos que utilizan DNSSEC y, curiosamente, el porcentaje de respuestas de los servidores autoritativos que proporcionan respuestas de DNSSEC. Creo que al hacer eso, obtendremos un control sobre el uso real de DNSSEC y podremos responder a la pregunta: ¿qué cantidad de DNSSEC se utiliza hoy en día? Creo que eso es algo que resultaría interesante para la comunidad.

Así que he estado examinando seis de nuestras siete métricas. Geoff Huston presentará la Métrica 5 después de mí. M7, como digo, es muy estable, por lo que este tipo de gráfico no nos dice mucho ahora.

Quisiera agradecerles por su atención y responderé cualquier pregunta que tengan ahora si es que tiene alguna pregunta.

RUBENS KUHL: Rubens Kuhl, .br. Quisiera comentar que el servidor recursivo, el DNS recursivo y las métricas recursivas se basan en tres servidores recursivos. Y actualmente tenemos 50 000 sistemas [inaudible] en Internet, por lo que publicar esos resultados hasta que obtengamos al menos 5000 servidores recursivos del DNS probablemente no sea lo que deberíamos hacer, ya que no tiene relevancia estadística alguna. Es como poner un [inaudible] en el microscopio y deducir todos los tejidos del mundo en función de eso.

[Entonces, me divierte] que la ICANN publique dicha métrica, especialmente en un área donde la ICANN no tiene datos directos diferentes a los datos raíz autoritativos cuando ejecuta uno de los sistemas de servidor raíz más completos porque es una instancia [inaudible] . Entonces, una raíz tiene una muy buena significación estadística entre las consultas de raíz. Pero en cuanto a las consultas recursivas, no deberíamos publicarlas hasta que crucen un umbral realmente adecuado de relevancia estadística.

CHRISTIAN HUITEMA: De hecho, es un muy buen punto. Hemos estado utilizando todo tipo de salvedades en esta charla para explicar que ahora tenemos un solo punto de medición y explicamos el [inaudible] que queremos realizar. Está claro que queremos más que ese punto. No sé si necesitamos 5000. Me gustaría tener 5000, pero no sé si necesitamos 5000.

Lo que planifico hacer es comparar los datos de los diversos sitios cuando se registran para ver cómo difieren y qué tienen en común. La idea es que sabemos que existen diferencias. Existen diferencias en el tiempo, como en la mañana y la tarde que no son lo mismo. En el fin de semana y en la jornada laboral no son lo mismo. Sabemos que existen diferencias en la geografía. La gente no solicita el mismo tráfico en China y en América. Sabemos que existen diferencias en el tipo de ocupación. Las personas no hacen la misma consulta en un sector académico y gubernamental o en una empresa o en una red privada o en una red móvil. Entonces, claramente, queremos tener una representación de todos ellos.

La significación estadística es algo que abordaremos. Es definitivamente algo que queremos hacer. Pero tenemos que comenzar en alguna parte, así que estamos recopilando datos de manera eficaz. Y estaremos comparando las fuentes para que podamos responder su pregunta.

RUBENS KUHL: Sí, pero me gustaría responder a eso. Si bien esas salvedades se describen, por lo general van como letras muy pequeñas. Entonces, cualquiera que lea lo que está en el informe lo repetirá y lo publicará en prensa y redes sociales y no reproducirá las salvedades de que estos datos en realidad carecen de sentido. De hecho, publicar eso es un perjuicio para la comunidad. Ese es mi punto de vista al respecto.

Un comentario que tenía sobre otro tema es que se mencionó que algunas de las consultas del registrador se vieron afectadas por las limitaciones en WHOIS y demás. Hay datos que puedo recopilar de todos los registros [amplios], que es el BRDA, que es el Acceso Masivo a los Datos de Registración Acotados. Esos datos de registración acotados ya contienen el registrador al cual está asociado ese dominio. Entonces, no hay necesidad de hacer consultas de WHOIS. Ya existen datos dentro de la ICANN que brindan esa información con una precisión del 100 %, así que es posible que deseen analizarlos también.

CHRISTIAN HUITEMA: Ese es un punto interesante En realidad, es un buen punto. Me gustaría abordarlo diciendo en primer lugar que el proyecto de ITHI es un cliente del proyecto DAAR. Estamos obteniendo datos de DAAR así que cualquier decisión que DAAR haya estado

tomando, la estamos heredando. Entonces, primero, me gustaría sugerirle que redirija su pregunta a las personas que administran el DAAR.

Lo segundo [y de alguna manera yo] intentaré canalizarlos. Por lo que tengo entendido, querían que el estudio se pudiera replicar, lo que significaba que alguien de afuera y no la ICANN pudiera replicar exactamente el mismo estudio, metodología abierta y datos que sean accesibles. Los datos que usted mencionó pueden ser o no accesibles desde el exterior, y eso colocaría a la ICANN en una posición singular para que sea el único que pueda realizar este estudio. Su elección fue no ir en esa dirección. Pueden cambiar su enfoque en algún momento, y tal vez John Crain pueda decir algo al respecto, pero hasta ahora esa es la dirección que hemos tomado. Y como cliente para ellos, estamos heredando esta decisión.

Entonces, la pregunta es, ¿por qué tenemos que depender de WHOIS para atribuir a un registrador en lugar de utilizar datos internos de la ICANN?

JOHN CRAIN: Si tenemos todos esos datos disponibles internamente, no los he encontrado internamente. Eso sería formidable. Pero una de las cosas que intentamos hacer es que sea replicable por otras personas, lo que significa utilizar fuentes externas. Lo único que

necesitamos de WHOIS es el ID del registrador. En realidad, hablamos antes sobre dónde pueden haber fuentes internas, por lo que puede que tengamos que cambiar a eso porque creo que el WHOIS puede no ser práctico.

RUBENS KUHL: La fuente, de hecho, se llama BRDA, por lo que es posible que deseen investigarlo o piratear esos servidores y obtener los datos de ellos. Pero incluso si utilizan esa información, eso hace que las cosas se puedan reproducir, pero hace que sea más complicado para que otras personas utilicen realmente las consultas de WHOIS. Pero pueden reproducirse utilizando consultas de WHOIS porque es la misma información. Por lo tanto, no es información privilegiada de ninguna manera.

JOHN CRAIN: Sí, entendido. Cuando comenzamos el proyecto, estábamos pensando que todo debería ser exactamente como lo haría la gente de afuera. Y tiene razón acerca de la reproducibilidad, así que lo estamos reconsiderando.

RUBENS KUHL: Bien.

ALAIN DURAND: Bien. ¿Hay alguna pregunta en la sala de chat, Cathy?

CATHY PETERSEN: No.

ALAIN DURAND: ¿Ninguna pregunta? Bien. Entonces, gracias, Christian.

CHRISTIAN HUITEMA: Gracias.

ALAIN DURAND: Muchas gracias por mostrar los números por primera vez aquí. Ahora, me gustaría invitar a Geoff Huston. ¿Geoff está en la sala?

CATHY PETERSEN: Aquí hay una pregunta.

ALAIN DURAND: Ah, tenemos una pregunta.

SEBASTIEN BACHOLLET: Como Geoff no está en la sala, simplemente quería hacer una pregunta y proviene de alguien que no tiene conocimientos técnicos. ¿Existe algún vínculo entre lo que están haciendo y

algunas de las preguntas sobre el traspaso de claves y los datos que necesitan para comprender lo que está sucediendo? Realmente lo siento, es una pregunta [inaudible] de [inaudible]. Gracias.

CHRISTIAN HUITEMA: A partir de hoy, la respuesta es no. No tenemos conexión. Esta no es una medida que consideramos inicialmente. Ahora, como en el futuro podría haber más traspasos y pueden ser más o menos frecuentes, podría ser algo que nos gustaría monitorear. Así que hoy hablamos sobre siete métricas. Creemos que las entendemos lo suficientemente bien como para poder medirlas. Estamos pensando ahora en la segunda fase en la que agregaremos más métricas, que veremos otros tipos de problemas. Y esa puede ser una de las áreas que debemos analizar y agregar a lo que estamos haciendo ahora. [responde en francés también]

SEBASTIEN BACHOLLET: Entiendo lo que quiere decir. Simplemente quiero asegurarme de que mi pregunta estuviera bien explicada, y lo siento por eso. Parece que hoy nos faltan datos para asegurarnos de que es el momento adecuado para realizar el traspaso de claves. No es el hecho de que cuando haremos un traspaso de claves, cada año podrán recopilar datos [como pregunta] que si con los datos de

su proyecto hay datos que pueden ser utilizados por las personas que necesitan decidir cuándo hacer el traspaso de claves.

CHRISTIAN HUITEMA: A partir de hoy, no tenemos datos que los ayudarán.

SEBASTIEN BACHOLLET: Gracias.

PAT KANE: Hola. Pat Kane con Verisign. Tan solo un comentario de la pregunta de Sebastien. Hoy temprano, el CTO de la ICANN correlacionó una disminución en las consultas de DNSSEC con el impulso del traspaso de KSK desde el otoño pasado hasta el próximo mes de octubre. Entonces, creo que es importante que comprendamos cómo se relaciona eso con el uso de DNSSEC para informar esa decisión porque muchos de los datos en términos de personas con resolutores que no tienen ambos pares de claves están empeorando más de lo que estaban a fines del año pasado. Por lo tanto, sería muy bueno enviar esa información a David lo antes posible. Gracias.

ALAIN DURAND: Gracias. Es un muy buen punto. Como hemos visto, Christian estaba hablando sobre una nueva métrica M7.2 que realmente monitoreará la cantidad de consultas que tienen configurado el bit DO. Eso puede ayudar en esa dirección con otras métricas que estamos tratando de diseñar en ese espacio en particular. Así que tal vez podamos tener una conversación fuera de línea si tiene ideas específicas sobre lo que deberíamos monitorear.

¿Geoff volvió a la sala? De acuerdo, entonces me disculpo. Perdimos a uno de nuestros oradores. Puedo hablar brevemente sobre lo que tenemos planificado hacer.

La métrica M5 inicialmente era una de las métricas que también observaba a los resolutores pero más desde la perspectiva del cliente. Le hemos pedido a Geoff que mire esto, y Geoff tiene un sistema de medición basado en Google Ads que es muy conocido y lo hemos estado usando en otros contextos. Le hemos pedido que utilice el sistema para explorar qué se puede hacer desde la perspectiva de un cliente, los resolutores stub.

Una de las cosas que nos gustaría ver es que los resolutores realmente almacenan en caché las cosas. A veces creemos que lo hacen. A veces creemos que no pueden o que pueden almacenar en caché por un tiempo más corto o que pueden almacenar en caché por un tiempo más largo. Por lo tanto, creemos que podemos obtener alguna medida de eso.

También podemos ver algunas de las distribuciones DNSSEC e IPv6 para determinar si el resolutor está configurado con DNSSEC o no, o si es capaz de usar IPv6 o no. También podríamos encontrar los resolutores más utilizados. Los resolutores más utilizados deberían calificar por ojos porque el sistema depende de Google Ads, por lo tanto, esto es utilizado por los usuarios físicos y no por las máquinas. Por lo tanto, no capturará la comunicación máquina a máquina, sino que la comunicación de usuario a equipo.

Este podría ser algún proyecto de medición [inaudible] que también nos podría informar acerca del cambio de clave y cuántos resolutores son realmente necesarios para cubrir el 95 % o cualquier porcentaje de la población que quisiéramos.

Este es un trabajo en curso. [Esas] son nuevas métricas que Geoff quisiera proponer. Con el mismo espíritu que Christian describió anteriormente, queremos que esto sea automático para que podamos recopilar la medición y monitorear esto con el transcurso del tiempo durante varios años.

Entonces, en pocas palabras, ese es el proyecto que nos gustaría hacer con Geoff.

Me disculpo porque él no pudo estar aquí, pero debe haber habido alguna circunstancia externa.

Si no hay más preguntas, cerraremos esta sesión temprano. Ah, pregunta ahora.

RUBENS KUHL: En realidad, es más una respuesta al comentario de Pat Kane. ¿Por qué el número de informes de resolutores de DNSSEC ha aumentado el número de informes de [inaudible] 2010 KSK? Todavía no sabemos si esos son resolutor de validación o no. Entonces, podría ser alguien que realmente solo tenga la clave raíz pero que no esté validando. Por lo tanto, no es un problema posible cuando se traspasa la clave. Entonces, si hacemos un estudio como ese en una métrica, probablemente deberíamos analizar resolutores de validación con claves antiguas, no solo los resolutores que informan claves antiguas. Porque eso no es algo que mida nada que pueda predecir lo que sucederá cuando hagamos el traspaso de la clave raíz.

ALAIN DURAND: Ese es un muy buen punto, pero agregaré lo siguiente. Deberíamos de alguna manera sopesar esto por la cantidad de usuarios que están detrás de ese resolutor. Si es solo algo que se usa en el sótano de alguien y se enciende durante cinco minutos, puede no tener la misma importancia que un resolutor que sirve a millones de clientes.

RUBENS KUHL: De acuerdo.

ALAIN DURAND: Entonces, si no hay más comentarios, cerraremos esta sesión. La próxima reunión de la ICANN es una reunión de políticas, por lo que no habrá sesiones técnicas así que no nos reuniremos. Pero los veremos a todos en Barcelona.

[FIN DE LA TRANSCRIPCIÓN]