ICANN83 | Foro sobre políticas – Sesión de At-Large sobre políticas con la OCTO
Jueves, 12 de junio de 2025 – 9:00 a 10:15 CEST

YEŞİM SAĞLAM: Recuerden decir su nombre para los registros, el idioma en el que hablarán si lo hacen en un idioma distinto del inglés y recuerden por favor hablar a una velocidad razonable. Muchas gracias a todos por estar aquí. Ahora le voy a dar la palabra a Jonathan Zuck, presidente del ALAC. Muchas gracias.

JONATHAN ZUCK: Muchas gracias, Yeşim. Muchas gracias a todos los aquí presentes, los que están con nosotros en forma remota. Es el día de uso indebido del DNS. Vamos a hablar sobre este tema. Vamos a estar aquí trabajando con OCTO para comprender cuáles son los planes sobre el uso indebido del DNS. Tal vez ver alguna demostración de Domain Metrica. Es el momento de ver todo esto juntos y sin más le voy a dar la palabra a nuestra colega Samaneh, que va a hablarnos sobre OCTO.

SAMANEH TAJALIZADEHKHOOB: Muchas gracias, Jonathan, por invitarnos a la sesión del día de hoy. Estamos muy contentos de estar aquí y compartir con ustedes

nuestro trabajo. Dicho esto, voy a comenzar con la presentación del día de hoy. Próxima diapositiva, por favor.

Hoy voy a darles una breve introducción sobre cuáles son los temas generales en los que SSR y OCTO trabajan, y lo que hacemos en líneas generales. Luego le voy dar la palabra a Sion, para que hable sobre Metrica. Luego Carlos y Sam, de mi equipo, van a presentar un pantallazo general sobre el otro trabajo de investigación que hacemos dentro de SSR OCTO. Por último, vamos a hablar sobre el informe INFERMAL. Próxima diapositiva, por favor.

Bien. Comencemos con la definición que hemos tenido en la comunidad por muchos años. Cuando pensamos en el DNS y en la seguridad del DNS o el uso indebido del DNS o la inseguridad hablamos sobre la infraestructura del DNS y los componentes y procesos asociados que no se encuentran asegurados en forma apropiada.

Esta terminología son puntos en los que nos vamos a concentrar luego en la presentación. Cuando estos procesos y componentes no se aseguran en forma apropiada, esto crea una oportunidad para los actores maliciosos para explotar estas vulnerabilidades o participar en forma de uso indebido del DNS, incluyendo los vínculos más débiles. Phishing, malware, botnet, spam y otros mecanismos que definimos dentro de la comunidad de la ICANN. Próxima diapositiva, por favor.

Hubo mucho que se dijo de parte de los científicos tecnológicos. Se trata de un sistema basado en la economía pero creo que la falla de

los sistemas y la falla de seguridad en general está causada por incentivos inapropiados o desalineados o por un mal diseño. Básicamente la persona que se encarga de esto no es la persona que sufre las consecuencias cuando esto falla.

¿Por qué esto es importante? Porque hablamos mucho de esto en la comunidad. A veces no conectamos unos con otros al respecto pero se habla mucho de esto. Cuando hablamos de cómo resguardar los sistemas hablamos sobre medidas de seguridad proactivas. Es decir, hacer que los sistemas sean seguros para evitar el uso indebido o malicioso a futuro. Cuando hablamos sobre aquellos que sufren las consecuencias cuando el sistema falla hablamos de las medidas de seguridad reactivas. Es decir, primero ocurre una falla. Hay víctimas de esa falla y luego aquellos que salvaguardan el sistema pueden tomar medidas, que es lo que denominamos esta seguridad reactiva pero ambas están relacionadas y a la vez no, porque una es preventiva y la otra reactiva.

Este es el área de trabajo a la que nos dedicamos y la idea es que ambos deberían estar alineados. Es decir, las medidas proactivas, los que deben salvaguardar o cuidar el sistema deben trabajar en forma alineada con aquellos que son víctimas del sistema y sabemos que en nuestro espacio no es el caso aún. Próxima diapositiva, por favor.

Luego, si los incentivos tienen que estar alineados, cómo podemos hacerlo. Esta es una de las formas sugeridas de hacerlo

basándonos en la bibliografía de economía que existe desde hace mucho tiempo. Diseñar métricas sólidas y de alta calidad. Esto independientemente del ámbito de trabajo.

¿Cuáles son estas métricas de alta calidad? Son métricas que influyen en decisiones para la performance de seguridad. Asignar recursos y también establecer prioridades, y las métricas mal diseñadas pueden traer consecuencias que no habíamos previsto.

Aquí a la izquierda vemos ambos gráficos que nos muestran la concentración de dominios que se informan como uso indebido del DNS. A la izquierda vemos la cuenta de los números de dominios que se informan en esta lista. A la izquierda se ven las cuentas o las cifras y a la derecha el porcentaje. Ambas muestran las concentraciones pero de dos formas muy distintas. A la derecha se toma en cuenta el tamaño de los registradores o los TLD.

Si vemos el punto en azul, es un ccTLD, .CC y no un TLD. Una vez que normalizamos una métrica, la posición de estos puntos cambia mucho. Es la misma idea en un punto pero se calcula de dos formas diferentes. Si observamos estos dos gráficos, alguien que deba tomar decisiones, un actor puede interpretarlas de distintas maneras. Esto muestra cuán importante es contar con una buena métrica que sea pertinente para la decisión que deseamos tomar. Próxima diapositiva, por favor.

Luego pensamos: "¿Qué es lo que hace que las métricas sean mejores?" Los datos de alta calidad, los datos con menos falsos positivos, definiciones más precisas de lo que se mide. Por

ejemplo, spam como email no solicitado comparado con spam como mecanismos de entrega para el malware. Incorporar errores de mediciones. Reconocer que hay limitaciones. Por ejemplo, si hay publicaciones sobre phishing o las tendencias del uso indebido del DNS, hay poca información sobre lo que se mide en esos informes. Qué es lo que se cuenta como actividad maliciosa. Cada una de estas definiciones puede realmente variar y traen resultados de mediciones diferentes. Si bien todas son ciertas, van a ser muy distintas unas de otras.

Dicho esto, quiénes somos. Somos el equipo de investigación de seguridad, estabilidad y flexibilidad dentro de la Oficina del Director de Tecnologías, OCTO, dentro del equipo de participación técnica e investigación. Yo lidero este equipo y mi equipo está compuesto por Sam Cheadle, Sion Lloyd y Carlos Ganan.

Quiénes somos. Somos un equipo que hace investigación y desarrollo en los identificadores de Internet. El centro de nuestro trabajo es que utilizamos datos y métodos científicos para responder preguntas que interesan a nuestra comunidad, la comunidad de la ICANN, pero también la comunidad científica.

Nuestro objetivo es aumentar la confiabilidad de las métricas que medimos, la seguridad, el mal uso y demás. Nos concentramos en áreas donde los métodos existentes y las métricas necesitan mejoras o necesitamos desarrollar un método desde cero. El motivo por el que dije lo que dije hasta ahora era para presentarnos. Básicamente para contarles por qué las cosas en las

que trabajamos son importantes. Primero contarles quiénes somos.

Ahora voy a hablar más en detalle sobre lo que hacemos, sobre nuestro trabajo. Comencemos con las trampas más comunes en las métricas del uso indebido del DNS que existen actualmente. Tenemos definiciones de uso indebido del DNS que no son consistentes o coherentes. Esto se ha debatido en la comunidad por años. Hay algunas instancias donde algunos casos de edge y falsos positivos no se tienen en cuenta. Algunas veces no son importantes pero algunos casos de uso sí lo son.

El contexto a veces se ignora y el contexto importa mucho, sobre todo si estamos hablando dependiendo del nivel de análisis, estas cuentas y cómo las analizamos importan muchísimo y pueden variar mucho. Incluso si hablamos sobre registradores de TLD o proveedores de servicios. Parece que fueran distintas entidades, entidades económicas pero sabemos que cuando analizamos un mercado operan en mercados muy distintos. Esto es importante. Cuáles son las métricas que definimos para cada uno de ellos y capturar cuáles son los incentivos dentro de ese mercado.

Algunas de las trampas más importantes en las métricas del uso indebido común, y quisiera concentrarme en esto hoy y pedirle a la comunidad que preste atención a esto, porque hay mucha falta de transparencia en la fuente de los datos de las métricas publicadas. Lo que se denomina la paradoja de la RBL y la paradoja temporal, de lo que vamos a hablar luego en más detalle.

RBL se refiere a la lista de bloqueo de reputación. Si no han escuchado hablar de este acrónimo, estas iniciales, son listas que enumeran dominios que son importantes para utilizar en distintas fuentes de actividades maliciosas. Cuando hablamos sobre RBL hablamos de actividades que están dentro del alcance de nuestro trabajo. Phishing, malware y demás.

Cuando hablamos de esta paradoja hablamos de muchos ojos pero ninguna vista. Es decir, hay muchas amenazas y distintos conjuntos de amenazas maliciosas. Lo que vimos en nuestro análisis, si no recuerdo mal, el OCTO 37, el artículo científico que publicamos en la conferencia de grupo.

Cuantas más fuentes de amenaza utilizamos para ganar claridad, a veces exponemos más desacuerdos. Es decir. Si se utilizan múltiples listas se revelan contradicciones. La RBL refleja distintas fuentes de datos, porcentajes de actualizaciones y definiciones de amenazas. Es decir, cuanto más RBL utilizamos o con cada una de ellas tenemos que tener cuidado y saber que cada una cubre un área de este escenario de amenazas. Esto se refiere a este ejemplo del ciego y el elefante. Si cada uno toca una parte del elefante, ve esa parte pero no el panorama general. Próxima diapositiva, por favor.

Lo que vimos recién es una parte del problema. Solo vemos una parte si utilizamos una cierta cantidad de listas de bloqueo de reputación. Vemos una parte del escenario posible pero luego tenemos el aspecto temporal. La ventana temporal de esta

paradoja. Cuando vemos las cuentas del uso indebido dentro de estas listas vemos que nos concentramos en la ventana de tiempo, ya sea en forma semanal, mensual, como sea que lo contemos.

A estas ventanas cortas les falta persistencia. Cuando son largas ventanas de tiempo, nos falta la posibilidad de responder rápidamente. Es decir, el uso indebido a corto plazo no nos permite las remediaciones rápidas. Es decir, distintas partes informan el uso indebido utilizando distintas ventanas de observación.

La elección de la ventana de tiempo puede manipular en forma intencional o no intencional y sesgar esta narrativa. Dependiendo del análisis que se elija para presentar los datos, la métrica puede variar mucho. Un ejemplo claro. El phishing aumentó en un 10 % desde julio del año pasado. Esto no quiere decir nada en realidad. Nadie puede reaccionar frente a esta afirmación porque todo va a depender de la ventana de tiempo en que los datos sean agregados, va a depender de la cantidad de listas que se hayan utilizado. Ahora les voy a mostrar un poco más de esto en la próxima diapositiva.

Otro aspecto temporal es que aquí, en este gráfico, vemos estos ccTLD, los 10 más importantes y que están clasificados según la cantidad de dominios maliciosos. En cada columna se agregan en forma mensual, cuatrimestral, semestral, anual y van de 1 a 10. Son los mismos. Los vemos en forma horizontal. Cuando lo analizamos en forma mensual, si lo comparamos con las otras mediciones, esto va cambiando y esto muestra que las ventanas de tiempo más

corta crean distintas métricas que aquellas que son en forma trimestral o semestral o anual. Según cómo se informe la métrica, esto puede traer resultados muy diferentes. Próxima diapositiva, por favor.

Aquí tenemos otro aspecto de las mediciones temporales. Cada una de estas barras representa un TLD. Lo que vemos es que los tres colores muestran el porcentaje de dominios activos, inactivos y los que no tienen direcciones, y estos son para los dominios observados en distintos TLD. Estos dominios observados, estos nuevos, tienen distintos porcentajes cuando nos referimos a la cantidad de dominios que están activos e inactivos.

Tenemos una definición de eso, y esto no es del alcance de esta presentación, pero básicamente aquellos que tienen contenido y pueden ser activos en Internet. Depende de cómo se muestren estos datos, la interpretación de las métricas pueden ser muy distintas. Próxima diapositiva, por favor.

Es muy importante cómo analizamos los datos y cómo definimos la métrica. Esto es muy importante en nuestra investigación. En general, estamos interesados en la seguridad y el uso indebido de los identificadores. Nos referimos a nombres y números. Analizamos también datos y métricas y tratamos de encontrar áreas, si existen, cómo podemos mejorarlas y, si no existen, cómo podemos crear métricas que creen incentivos para una seguridad mayor.

Algunas de las instancias de nuestro trabajo, aquí las están viendo en rosa. Trabajamos en identificar las registraciones en lote, clasificar los subtipos. También consideramos las concentraciones de uso indebido y los tipos de dominios, el tiempo de respuesta de cada uno de estos dominios.

Esta área destacada aquí, en el círculo, es la idea núcleo en la plataforma Domain Metrica. Se trata de una plataforma que se va a centrar en datos y métricas, en mejorarlos, presentarlos, etc. Con esto termino mi parte y quisiera darle la palabra a Sion para que nos cuente sobre los detalles de la plataforma de Domain Metrica.

SION LLOYD

Gracias, Samaneh. Soy Sion Lloyd. Les voy a hablar de Domain Metrica. Voy a hacer primero una introducción y después les voy a mostrar un estudio de caso. Un ejemplo de una investigación con la que pueden utilizar la plataforma. Domain Metrica es una plataforma que hemos construido aquí. La idea es que recolectamos la mayor cantidad de datos posibles, distintas informaciones que tienen que ver con los identificadores de Internet.

Esos datos los reunimos de manera que podamos incorporarlos, agregarlos, combinarlos de distintas maneras y producir una métrica útil a partir de los datos con los que comenzamos. Producir datos es solo la mitad de nuestra historia. Es el aspecto más útil y

más interesante, que es compartir esos datos para poder visualizarlos e interactuar.

Tenemos entonces dos maneras de lograrlos. El front end de la página web y luego tenemos una API donde se pueden obtener datos programáticamente. Es decir, se puede quizá combinarlo con un conjunto de datos que podemos tener dentro de la organización y hacer cosas que quizá no hacemos porque no tenemos otros conjuntos de datos.

En cuanto a la disponibilidad, Domain Metrica está abierto a todos los que tengan una cuenta de ICANN. Es decir, la cuenta que han utilizado para registrarse para esta reunión les va a dar acceso a Domain Metrica. Además, las partes contratadas, es decir, registros y registradores reciben más información con respecto a sus propios datos.

Hay otras plataformas que hacen cosas muy parecidas pero tratamos de construir aquello que nos sirva para el nicho que tenemos y que también resulte útil para otros. Tratamos entonces de generar la mayor cantidad de datos posibles para tener muchas listas de reputación. Esta es nuestra manera de enfrentar esta paradoja que se mencionó antes.

Hay distintos datos, distintos aspectos y lo que queremos es ver cuál es nuestra confianza en poder ver un panorama lo más amplio posible. Nadie puede ver el panorama total pero lo que sí tratamos

es ver la mayor cantidad posible y saber dónde tenemos limitaciones.

Vamos a hablar ahora de algunos otros aspectos y la forma de poder interactuar con los datos. Lo vamos a ver un poco más adelante también. En la medida de lo posible compartimos todos los datos que tenemos. En algunos casos hay limitaciones contractuales pero en la medida de lo posible compartimos los datos, somos transparentes respecto de esos datos y también tratamos de expandir para tener nuevos conjuntos de datos y tener nuevos usos de los datos, nuevas formas de visualizarlos. Básicamente hacer evolucionar la plataforma para que sea lo más útil e interesante posible para distintas áreas de la comunidad. Siguiente diapositiva, por favor.

Con respecto a la paradoja del tiempo, básicamente lo que hacemos es permitir que el usuario pueda elegir el rango de tiempo que quiere analizar. Si queremos ver los datos de un año, por ejemplo, un mes o una semana, todo esto es posible. Vamos a mostrar los datos y los vamos a escalar dentro del periodo de tiempo que va cambiando.

En la medida de lo posible también filtramos aquellos dominios que creemos que no deben estar presentes y este es un proceso que vamos a ir mejorando y ajustando con el tiempo. Actualmente quitamos los dominios que sabemos que no existen. Tenemos los ABR y también tenemos las zonas de archivos de los gTLD. Cualquier entrada que no tiene un archivo de zona raíz lo quitamos

de nuestros datos porque es un dominio que puede no haber existido. Pueden ser un falso positivo. Por eso no lo contamos. Siguiente diapositiva.

Esta es una introducción básica y lo que nos han pedido aquí es que presentemos un caso de estudio. Hay un artículo de blog que ha emitido el departamento de cumplimiento contractual sobre el gTLD .TOP, que había recibido una notificación de incumplimiento. Ahí está el link para leer el artículo, si quieren verlo.

Creo que podría ser interesante utilizar Metrica para analizar esta historia y ver cuáles son los datos que tenemos y cómo podemos analizar los datos y ver qué podría significar todo esto. Aquí tenemos varias diapositivas que nos van a mostrar el front end y cómo podemos utilizar este punto de datos.

Esta es la pantalla que ustedes van a ver cuando se conecten a Domain Metrica. Hay varias cosas que van a ir sucediendo aquí. Básicamente, lo que mostramos en esta página son estadísticas de titular. Es decir, una gama muy amplia de números y la división de estos números aparece en vistas no filtradas de todos los datos que tenemos en nuestras métricas. Es decir, hay una cantidad de zonas a las que pertenecen estos dominios, aquellas zonas donde ha habido un uso indebido informado y luego tenemos una división por los distintos tipos de uso indebido, phishing, malware comando y control. Seguimos teniendo algunos errores en el sistema. No podemos llegar a cero. Hay muchas partes que se mueven para tener todos los conjuntos de datos en el lugar

adecuado y estamos teniendo un pequeño problema con algunos de ellos en este momento.

Lo otro que mostramos en la página principal es una tendencia de 12 meses. Es decir, cómo han evolucionado con el tiempo, qué es lo que estamos viendo como la dirección en la cual van avanzando todas estas cuestiones. Hoy podemos utilizar una búsqueda para gTLD. También podemos buscar en los registradores y podemos buscar dominios específicos.

La búsqueda de un registrador nos da un dato similar a lo que vamos a ver en este TLD, la búsqueda de dominio nos va a dar datos distintos porque no tiene sentido si no los agregamos pero tratamos de mostrar metadatos sobre la entidad que estamos buscando.

Esta es la sección superior que vemos ahí. Es información de contexto, el tamaño de las zonas, si es un registrador. Va a mostrar el tamaño de la cartera. Luego vemos el uso indebido informado actual para esa entidad. Está dividido en distintos tipos de uso indebido. Mostramos también el porcentaje del tamaño. Es decir, lo normalizamos hacia el tamaño de la zona raíz, etc. Siguiente diapositiva.

Después tenemos algunas tendencias y la proporción de uso indebido en cada una de las categorías. Estos gráficos son muy interactivos. Podemos hacer clic en cada uno de ellos. Vamos a hacer clic para mostrar el informe de uso indebido. Estas son las

tendencias mensuales a 30 días que muestran los conteos diarios del uso indebido informado y cómo esto evoluciona con el tiempo.

Hay una cierta cantidad de cuestiones que podemos hacer ahora que estamos interactuando con estos gráficos. Por ejemplo, si vamos a la próxima, podemos cambiar el periodo de tiempo. Aquí vamos a principios de abril y, como tenemos un periodo más amplio, ahora en lugar de mostrar los datos diarios mostramos los datos semanales para no tener demasiados puntos que si no, no se pueden ver. Este es el conteo promedio por semana. Esta es la manera en la que agregamos las semanas en un solo punto con ese periodo de siete días.

Al considerar la entidad por sí misma vemos cómo estaba cambiando pero no nos dice demasiado porque hay otras cosas que pueden estar ocurriendo en líneas generales. Otra cosa que permitimos es agregar otros TLD en la comparación. Aquí entonces tenemos el TLD .PRO. Podemos ver que .TOP tiene una baja bastante gradual. En el mismo periodo .PRO sigue avanzando en el mismo sentido. El conteo absoluto es menos. De todos modos, sigue una tendencia ascendente y no hacia abajo.

Uno de los problemas de comparar las distintas entidades es que son de un tamaño muy distinto. La zona .TOP es mucho más grande que la de .PRO. Entonces, lo que permitimos es que en lugar de hacer un conteo absoluto incluimos los porcentajes. Es decir, seguimos tomando el wikimedia pero normalizamos el tamaño de la zona. Esto nos muestra una imagen bastante diferente.

Si consideramos la concentración o la cantidad proporcionar del uso indebido, lo que vemos ahora es que .TOP va cayendo y .PRO, como proporción de la zona raíz, en realidad ha cambiado de lugar con .TOP. Podemos tener más TLD en la comparación.

Aquí hemos agregado el TLD .WIN y esto nos muestra una actividad a corto plazo que podemos ver. No sabemos muy bien cómo va a evolucionar pero definitivamente hay un pico hacia mediados de abril. Pareciera que va descendiendo. Quizá hay una campaña de ese TLD o quizá hay algo más sistemático que podemos ver a lo largo del tiempo que hace que sea un poco más preocupante.

Aquí entonces vemos tres TLD diferentes y sus distintas conductas o comportamientos. Estamos por presentar una nueva funcionalidad. Esta es una noticia que les damos en exclusiva. Tenemos la opción de comparar con todos los otros TLD. Lo que hace esto es tomar el agregado de todos los otros datos que no mostramos en forma individual. Es decir, todos los TLD, excepto .PRO y .WIN, que ya ploteamos en este gráfico. Está mostrando cómo han evolucionado en el mismo periodo.

Esta es la línea con tamaño de diamante que ven hacia abajo. Este es el contexto de la situación. Lo que está ocurriendo es que puede haber algo extraño que está ocurriendo en el conjunto de datos o quizá se trata de un TLD en particular. Vemos entonces que en general, en el mismo periodo de tiempo, el panorama es bastante plano. Es decir, en esta escala, es bastante invisible.

Creo que lo último que se puede decir del front end es que todos los datos que están en estos gráficos se pueden descargar. Hay algunos links aquí, en estos gráficos. Pueden descargarlo en formato CSV o JSON. Quizá pueden plotearlos de distintas maneras o quizá tienen sus propios conjuntos de datos con los que los pueden hacer. Pueden sacar los datos del sistema, colocarlos en los sistemas propios de ustedes y utilizarlos de la manera que les resulte más conveniente.

La otra manera de extraer los datos es a través de las API. Si pueden escribir un script para poder extraer los datos o si quieren utilizar los datos que ya están en el sistema y simplemente extraerlos en el propio warehouse de datos pueden utilizar la API para eso. Tiene todos los mismos datos que están disponibles y son programáticos.

Esta es una captura de pantalla de la documentación API que tenemos. Está todo allí. Pueden buscar, pueden poner periodos de tiempo, todo lo mismo que se puede hacer en el front end. Esto es lo que van a recibir. Una estructura JSON que es muy fácil de importar a Python o plotearlo. Pueden combinarlos con sus propios datos.

Para cerrar el tema de Domain Metrica, no está terminado. Lo que tenemos hoy lo pensamos como el inicio. Ya hemos recibido opiniones de los usuarios, tenemos nuestros propios casos de uso, nuestras propias listas que queremos ingresar en el sistema y lo

pensamos como un cimiento sobre el cual vamos a empezar a ir construyendo.

Se ha mencionado esto en el chat pero evaluamos los datos fuente que tenemos para tratar de garantizar que tenemos el mejor conjunto de datos disponibles porque el contraargumento es que el ploteo genera tendencias. No podemos cambiar los datos con mucha frecuencia. Tenemos que tener cuidado. Tenemos que tomar nota de cuando haya un cambio. Los cambios en las tendencias pueden atribuirse a algo que ocurre sin los datos.

Finalmente, hay algunos links. Una página que les da más información. También a las preguntas frecuentes, a la documentación y a la plataforma en sí. Como dije, si se conectan con su propia cuenta de la ICANN pueden interactuar con el sistema. No sé si ahora vamos a recibir preguntas sobre Domain Metrica o si vamos a esperar hasta el final hasta que terminen todas las otras presentaciones. Ahora si quieren puedo responder.

JONATHAN ZUCK Jonathan Zuck, presidente de ALAC. Muchas gracias por esta presentación. Estoy viendo en línea y aquí en la sala si hay alguna pregunta. Yo tengo una pregunta breve. Has mencionado la comparación entre los nombres de dominio en las RBL con los archivos de zona y analizar los falsos positivos y eliminar aquellos dominios.

Si existieron históricamente y no están eliminados, ¿esto significa que se traen datos nuevos? La pregunta más amplia sería. Varios de estos ataques en un periodo corto de tiempo puede ser difícil de medir cuando algo se ha registrado en forma maliciosa porque es como si estuviera estacionado por la mayor parte del tiempo excepto por el breve periodo que se ha utilizado. Es decir, antes de que pase a los datos históricos se elimina de los datos actuales. ¿Se entendió la pregunta?

SION LLOYD Creo que sí. O sea, hay una diferencia entre la visión histórica y la actual. Lo que hacemos es que cada día analizamos la situación como la vemos ese día y luego tal vez al día siguiente el dominio sigue estando en la RBL pero ya no está en el archivo de zona. Lo vamos a contar hoy pero tal vez no lo contemos mañana si eso responde a la pregunta. Sé que no es un sistema perfecto. Tenemos otras ideas y formas que vamos a intentar perfeccionar. Las listas de los dominios y los que creemos que pueden ser abusivos en algún momento en el tiempo. Este sería como el primer paso para limpiar esas RBL y los dominios que creemos que deberían estar en las estadísticas ese día. Tal vez el día siguiente pueden cambiar o la semana siguiente también. Hay un momento en el que pueden coexistir.

JONATHAN ZUCK Gracias. Hubo algunas veces en las que cuando se hizo el estudio, cuando se estudiaron los nombres de dominio en la época de la

COVID, la mayoría estaba estacionado o eso parecía. La presunción era que estaban estacionados por objetivos comerciales. ¿Cómo se hace esa determinación si lo comparamos con aquellos que están estacionados durante periodos de tiempo y lo comparamos con los objetivos maliciosos?

SION LLOYD Es muy difícil tener una visión completa que atraviese el tiempo. Podemos ver solo lo que está presente actualmente. Hacemos lo mejor que podemos para dar un panorama de lo que creemos que es la situación actual. En forma realista, si lo analizamos, esto puede cambiar en media hora. No podemos permitir esto. Podemos intentar mejorarlo cuando hay una incoherencia, cuando medimos algo y cuando alguien analiza esos datos para ver si es pertinente o no.

JONATHAN ZUCK Gracias. Veo que Rowena levantó la mano.

ROWENA SCHOO Rowena Schoo, de NetBeacon. Muchas gracias a Samaneh y a Sion por esta presentación que es un trabajo muy interesante. Me gustaron mucho estos gráficos que permiten la comparación de los TLD. Me pareció muy interesante. En primer lugar, estoy totalmente de acuerdo en el hecho de que normalizar los porcentajes permite comparar el porcentaje de uso indebido. Tenemos la opción de analizarlos. Me pregunto, analizando la

interfaz, si también se puede ver la cantidad de dominios porque a veces nos parece que eso puede ser pertinente. A veces puede ser difícil reflejar esto en los datos, sobre todo si tienen zonas reducidas o pequeñas. Mi segunda pregunta es la siguiente. Cuando se compara con todos los TLD, ¿son los TLD juntos o se hacen exclusiones en términos del uso indebido?

SION LLOYD Gracias, Rowena. Sí, se obtiene más información en hover. Se obtienen los números. Lo otro que intentamos introducir es la posibilidad de desactivar algunas líneas en esas tablas porque si uno se interesa solamente en el malware es difícil de verlas porque tal vez se superponen con las de phishing en ese eje. Todos los demás TLD son todos los gTLD que tenemos en el sistema. Es decir, es el total de cómo se ven en un mismo periodo de tiempo. Si solamente tenemos un TLD en el gráfico, va a ser todo lo demás con excepción de ese TLD.

ROWENA SCHOO Gracias. Tiene sentido. Algo que veo mucho en los datos subyacentes es que hay muchos que no tienen números elevados o cifras elevadas de uso indebido. Por eso me parece que está muy bien tener este tipo de información, de tablas. Muchas gracias a tu comunidad por ese trabajo.

JONATHAN ZUCK Tenemos aquí una serie de preguntas. Justine Chew, ¿desea hacer uso de la palabra?

JUSTINE CHEW Muchas gracias por la presentación, Sion. Tengo algunas preguntas básicas. Mencionaste que tu equipo hace una limpieza, por así decirlo, de datos para eliminar los falsos positivos y demás. ¿Hacen un seguimiento en una línea de tiempo estricta para introducir datos nuevos o se hace en forma automática?

Lo pregunto porque podemos tener una imagen de hoy con los datos que tenemos y, si lo hacemos mañana, ¿esa base de datos va a ser la misma? Mi segunda pregunta es la siguiente. ¿Podríamos recordarme por favor si va a haber iniciativas para introducir datos de ccTLD?

SION LLOYD En cuanto a la primera pregunta, lo reevaluamos. El plan es hacerlo en forma anual. Domain Metrica todavía no ha existido ni por un año pero la intención es hacerlo en forma anual y cualquier cambio que hagamos va a anunciarse y se anunciará en forma prudente para que todos sepan lo que está sucediendo. Seguramente pongamos una notificación en el sistema en sí. No vamos a cambiar los datos y lo vamos a tener en forma invisible, si se quiere, para el usuario.

La limpieza de datos que hacemos se suele basar en el archivo de zona y al menos conocer el tamaño de la zona nos permite

normalizarlo para ver cómo lo abordamos. Potencialmente no tener acceso al archivo de zona pero tener el TLD presentado en forma tal que sea comparable con los gTLD que tenemos.

JONATHAN ZUCK: Hadia.

HADIA ELMINIAWI: Muchas gracias por esta presentación. Quería preguntar cómo Domain Metrica se relaciona con el DAAR e incluye los ccTLD. Si DAAR sigue existiendo, si hay una diferencia entre ambas plataformas y también en cuanto a la accesibilidad. ¿Quién puede acceder a Domain Metrica? ¿Quién puede utilizar Domain Metrica? La comunidad, por ejemplo, ¿puede utilizar esta plataforma? Gracias.

SAMANEH TAJALIZADEHKHOOB: Muchas gracias, Hadia, por tu pregunta. DAAR sigue existiendo en la plataforma funciona pero estamos trabajando para ir usándola cada vez menos. La idea de Metrica es que Metrica es modular e intentamos agregar un módulo en cada periodo de tiempo. No soy más específica porque todavía no está claro pero el primer módulo fue sobre el uso indebido del DNS.

La plataforma no necesita presentar datos sobre uso indebido. Estamos trabajando para presentar datos de metadatos relacionados con los dominios, las ISP y la seguridad. Metrica va

más allá de DAAR y el primer módulo que fue sobre el uso indebido del DNS y sus métricas va más allá de DAAR.

Acerca de la interacción de ccTLD, hemos tenido unos 40 ccTLD que participaron en DAAR. La intención es agregar un endpoint de ccTLD a Metrica. Sion y el equipo han tenido sesiones con el grupo de trabajo de uso indebido del DNS de la ccNSO para ver cómo se podían incluir estos ccTLD. Estamos analizando este feedback y viendo cómo podemos implementarlos y a futuro vamos a tener ccTLD que puedan loguearse a sus propias credenciales y ver sus propios datos además de todos los demás datos públicos.

Tenemos el consentimiento para los que participaron en DAAR y que pasaron automáticamente a Metrica y quien desee unirse a Metrica nos puede contactar y en cuanto a los proveedores de zonas, los ccTLD también pueden formar parte de Metrica si no quieren proveer sus zonas. Vamos a tener en cuenta los tamaños de las zonas para calcular las métricas. Vamos a escuchar más de todo este tema a futuro. Vamos a seguir trabajando en esto.

Espero haber respondido a tu pregunta. Dicho esto, quisiera pedir si podemos continuar con las presentaciones para poder respetar el tiempo que se nos ha asignado, salvo que haya alguna otra pregunta.

SAMUEL CHEADLE Hola. Soy Sam Cheadle. Voy a presentar un proyecto que actualmente está en progreso sobre las registraciones en lote. Nos

vamos a concentrar en intentar identificar las registraciones relacionadas con esto. Hay un debate en la comunidad actualmente sobre las distintas formas de intentar identificar estos dominios relacionados en el momento de la registración.

Uno de los principales objetivos de esto es poder tomar estas fuentes de datos centralizadas a las que la ICANN tiene acceso y testear la posibilidad de identificar los grupos de dominios relacionados, teniendo en cuenta que estas fuentes de datos centralizados no contienen datos personales. Es decir, no vamos a poder taggear dominios basados en los registratarios o los que tienen las cuentas.

Según una serie de estudios e investigaciones anteriores, quienes atacan en general registran dominios en lote o en forma masiva. Por ejemplo, sabemos que las API se utilizan comúnmente en el informe INFERMAL reciente se confirma esto. Estos puntos que vemos en pantalla describen esta hipótesis pero creo que es un poco más sólido que esto.

Tenemos distintas pruebas. Las API, por supuesto, se utilizan para objetivos legítimos pero a menudo quienes cometen los ataques hacen registraciones en lote o en forma masiva. Nos estamos concentrando particularmente en las registraciones en lote que son aquellas que están atadas a un límite de tiempo, los grupos de registraciones de dominios, que son muy similares a las registraciones masivas y que cubren un tiempo relativamente más

largo. Nosotros nos estamos concentrando en los lotes de un periodo de tiempo más corto.

Una de las motivaciones principales es intentar seleccionar estos grupos de dominios rápidamente para permitir un enfoque más proactivo y la mitigación de las amenazas para exponer estos resultados a la comunidad. Como dije antes, este trabajo se concentra en fuentes de datos centralizados con unas características de registración limitadas, concentrado en conjuntos de datos de registración. En este caso el registrador y los servidores de nombre autoritativo y los datos de creación.

El método principal es separar todas las registraciones de dominios recientes basándonos en el registrador y en el servidor de nombre autoritativo. Una vez que tenemos estos subgrupos podemos ejecutar un algoritmo de agrupación, DBSCAN, que es un algoritmo de agrupación basado en la densidad de aplicaciones. Este método depende en cierta medida de la popularidad o la frecuencia de las registraciones dentro de cada registrador y su agrupación de servidor de nombre.

Hay muchos casos, muchos dominios que se registran en periodos cortos de tiempo. A menudo estos se asocian con algún defecto o algún servidor de nombre compartido. Estos se excluyen de esta fase de análisis actual. Próxima diapositiva, por favor.

Como mencioné, estos datos centralizados de la ICANN, estamos analizando los datos BRDA, o acceso masivo a datos de registración. gTLD solamente, no ccTLD. Estamos haciendo una

verificación cruzada con las listas de bloqueo de reputación como vemos en la lista que vemos en la diapositiva.

Aquí tenemos una visualización de alto nivel de un periodo de duración de un mes luego de ejecutar este algoritmo de agrupación para tratar de seleccionar los lotes de registración. Aquí es donde vemos en el eje Y 25 registradores con registraciones del lote y su más alto porcentaje. En el eje X vemos los puntos. Cada uno tenemos azul o rojo, que muestra el volumen de estas registraciones en lote y cuantas más registraciones de lote tenemos por hora, se reflejan unos 4.000 dominios registrados en lote por ahora y en el volumen menor unos 10.

No todos los registradores tienen la misma proporción de dominios registrados en forma indebida o maliciosa. Podemos ver que algunos registradores tienen un volumen amplio de registraciones pero en general son relativamente poco frecuentes y esto puede ser indicativo de las campañas. Próxima diapositiva, por favor.

Este es un resumen de los puntos clave del análisis que hemos hecho hasta ahora y de los resultados principales. Es decir, de extender un poco más este periodo. Estamos mirando el primer trimestre de 2025. Si consideramos todos los dominios gTLD registrados estamos hablando de alrededor de 16 millones de dominios.

El número total de los dominios registrados en lote estamos en alrededor de [4.200.000]. El 25 % de todas las registraciones. Tengan en cuenta que los lotes pueden ser de cualquier tamaño. Es

decir, estamos incluyendo lotes pequeños de dos o tres dominios. Este es un punto a tener en cuenta.

En general, los análisis de registraciones masivas consideran los grupos de dominios más grandes. Si consideramos los dominios con uso indebido, tenemos los dominios con RBL dentro de este periodo y ahí estamos hablando de 1.500.000 de dominios maliciosos y de esos alrededor de 776.000 dominios los estamos etiquetando con algoritmos dentro de este lote como se espera. Esto es una proporción más alta que la frecuencia general de registraciones en lote. Están divididas por tipo de amenaza. Tenemos el spam, que es el tipo de uso indebido que tiene la proporción más alta. Comando y control, malware, todos están por encima del 25 %. Siguiente diapositiva.

Esto es para darles una imagen de la composición de los TLD. De nuevo, aquí solamente estamos considerando los gTLD. Ningún ccTLD. Pero si dividimos los lotes entre los maliciosos y todo el resto, es decir, que no hay ninguna evidencia de amenaza, es decir, son lotes desconocidos, ahí vemos cómo es la composición. Están los cinco TLD más frecuentes. La cantidad de gTLD que vemos tienen unas tasas de uso indebido bastante altas con respecto a lo que hemos visto en el pasado que muestran unas tasas de registraciones en masa bastante altas. Siguiente diapositiva.

Si continuamos con este punto y consideramos la relación por registrador en este gráfico cada uno de estos puntos está mostrando un registratario único. En el eje X tenemos la tasa de

registración. Esta métrica que hemos estado analizando en las diapositivas anteriores. Ahí está el porcentaje de dominios.

Lo que demuestra esta diapositiva es que hay una clara relación. Los registradores que tienen una tasa de registración más alta, tienen también una tasa de uso indebido más alta. Los puntos que se ven en este gráfico, el tamaño de estos puntos, muestran el volumen de las registraciones masivas y la cantidad de dominios. Hay una correlación muy significativa.

Este fue entonces una visión general para mencionar los puntos centrales de un trabajo que se encuentra en marcha y estamos muy interesados en recibir sus opiniones sobre todo esto. También estamos en la fase de validación y estamos muy interesados en trabajar con los registradores para tratar de evaluar la confiabilidad de todos estos resultados. Es decir, de los conjuntos y de las metodologías que hemos descrito. Tenemos confianza de que a nivel general estamos considerando lotes que son válidos. Puede haber falsos positivos. Ese es el problema que estamos investigando en este momento. Necesitamos la ayuda de los registradores para poder acceder al registratario o al titular de la cuenta. Este es un trabajo en marcha. No queremos solamente considerar estos lotes restringidos en el tiempo. Idealmente querríamos poder hacerlo en forma continua para poder considerar estas nuevas registraciones y nuevos lotes rápidamente. Puedo ahora responder las preguntas. Puede ocurrir después. Creo que nos queda poco tiempo. Por favor, pónganse en

contacto por email o después de la reunión. Muchas gracias. Le voy a dar la palabra ahora a Carlos.

CARLOS GANAN

Muchas gracias. Buenos días a todos. Lamentablemente no tendremos tiempo para ver mi investigación. Mi nombre es Carlos Ganan y me focalizo en el tiempo de uso indebido. Voy a presentarles la idea del proyecto INFERMAL. Como podemos ver, la mayor parte de las métricas nos cuentan cuántos dominios malos hay. En mi investigación, nosotros preguntamos cuán rápido se responde al uso indebido. De eso viene este tiempo.

¿Por qué importa entonces? Porque pasamos de mirar el atacante a mirar el defensor y cuánto tiempo está online. Hay distintas maneras. Hay medidas activas y pasivas. El resultado muestra que estos dominios son los que están la menor cantidad de tiempo online. El spam y otros son los que más tiempo están online. En este proyecto, medimos las métricas de tiempo porque de esa manera se puede hacer que los operadores puedan rendir cuentas. Con esto le doy la palabra a Samaneh.

SAMANEH TAJALIZADEHKHOOB

Gracias, Carlos. Esta seguramente fue una de las presentaciones más rápidas. Gracias al equipo por las lindas presentaciones. La razón por la cual me dan la palabra a mí es que sé que hubo un pedido específico de ALAC de presentar con respecto al INFERMAL. Queremos dejar tiempo ahí. Esperamos poder tener tiempo al final

o si no, vamos a responder offline. Mi equipo también va a responder en el chat.

Muy bien. Hablemos de INFERMAL. INFERMAL significa un análisis inferencial de dominios registrados maliciosos. INFERMAL es un proceso que fundamos, OCTO y la organización de la ICANN financiaron el proyecto. Está liderado por Maciej Korczyriski, de KOR Labs. Hemos colaborado con ellos. El trabajo va a presentarse en una conferencia sobre alta seguridad. Somos parte de ese artículo científico.

Debo ir muy rápido. Les voy a saltear muy buena parte de lo que tenemos aquí. No tengo que hablar de la motivación porque todos sabemos que hay mucho uso indebido y este trabajo específico se focaliza en los atacantes y cuáles son los factores que influyen en lo que los atacantes eligen en términos de registros y TLD.

Básicamente, la idea de esta investigación es hablar de las preferencias de los atacantes. Los conjuntos de características son recolectados. Los atributos de registración, qué tipo de atributos vemos en los procesos de registración. La verificación proactiva que ocurre y las prácticas de seguridad reactivas.

Recuerden que al principio de esta sesión hablamos de las medidas proactivas y reactivas para poder hacer un análisis preciso de la relación entre los factores y las preferencias de los atacantes. Hay unas revisiones de regresión. No los voy a aburrir con esos detalles. Siguiente diapositiva, por favor.

Hay varias zonas de archivo y listas de bloqueo que se utilizan aquí, logs de transparencia y muchas otras más pero este estudio se focaliza en el phishing particularmente. A partir de aquí les voy a mostrar las características seleccionadas para los atributos de registración. Hay características como API gratuita, métodos de pago, precios, descuentos. Si son gratuitos los servicios o no.

Para las medidas proactivas y la validación operativa de los contactos, las restricciones en los procesos de validación, las advertencias. Para las medidas reactivas, consideramos la cantidad de tiempo en que el dominio está activo y cuánto tiempo está en la lista de bloqueo, y luego verificar esto periódicamente.

Algunas descripciones de lo seleccionado. Van desde 0,78 a 69 dólares, el rango de precio. Estos son los ejemplos de los dominios caros: usps.bar o dhlcenter.net. Como ven, son números más altos. También vemos si podemos hacer clic en Next, que hay descuentos durante la registración. Cuando comenzamos de un precio específico, terminamos con un precio diferente porque hay un descuento por varias razones. Esta es la distribución de los distintos servicios gratuitos que se agregaron a la registración como DNS, email, etc.

Con respecto a los resultados, básicamente los resultados sugieren que si consideramos la parte de los incentivos económicos vemos que el precio y el descuento importan, y el precio es un punto menos importante que los descuentos que se obtienen durante el proceso.

Esto es para las registraciones maliciosas. Este estudio se focaliza en registraciones maliciosas y no las no comprometidas. Al ver las registraciones benignas vemos que el precio promedio es más alto. Es 8.6, que es más alto que el de las maliciosas, que es 4.7. Básicamente entonces el precio ya sea el inicial o el final, ambos son más importantes para los usuarios benignos.

Cuando consideramos la API para registrar los dominios involucrados o para crear una cuenta o para hacer la búsqueda vemos una relación bastante positiva que se utiliza para la creación y gestión de la cuenta. Cuando consideramos los paquetes gratuitos, si bien tienen una relación significativa positiva vemos que no solamente son atractivos para los usuarios maliciosos sino también para los legítimos, lo cual sugiere que a fin de cuentas no importan tanto.

En lo que se refiere a las verificaciones previas proactivas, vemos una relación negativa. Hay una gran diferencia entre las Selecciones maliciosas y las de los usuarios benignos. Esto importa mucho cuando hablamos de registros maliciosos. Este es un punto donde la mitigación puede generar una mejora.

Hemos presentado los resultados muy rápido. Esta es la parte más importante de la presentación. Hemos publicado un blog sobre esta parte específica de este estudio. Hay muchas consideraciones y contexto en este estudio, que nos parece que es importante que nuestra comunidad considere. El blog también habla

específicamente sobre las aclaraciones, la interpretación de los estudios INFERMAL.

Lo más importante es que hay un conjunto de factores en este estudio que son seleccionados. Es importante prestar atención al hecho de que estos factores son válidos dentro del modelo que se tiene en cuenta junto con los otros factores. ¿A qué me refiero aquí? Recuerden cuando dije que API para la creación y gestión de cuentas tiene un impacto positivo significativo en el uso indebido para el [400 %]. Esto significa que el 400 % no existe por sí solo sino que impacta la API a la vez que todos los otros factores tienen una constante.

Si agregamos o quitamos una variable del modelo, este sería un número completamente diferente. Quiere decir que 400 % es un número de las cosas que medimos en relación con las cosas que medimos. Hay muchas otras cuestiones que no se miden en este estudio. En la práctica no todo es constante. Hay muchas cosas que son dinámicas. Se puede concluir que este es un factor importante donde la cantidad de importancia no se puede tomar por sí sola en aislamiento.

Mucho de lo usado en este estudio es agregado o combinado. Por ejemplo, las restricciones para la consolidación de pagos son un agregado de muchas variables. Criptomonedas, transferencias bancarias, billeteras digitales y esto es lo que miramos también cuando consideramos los resultados.

Por último, lo último a tener en cuenta es que las implicaciones económicas de los resultados de este estudio son diferentes del uso real. Es decir, miramos las preferencias de los atacantes y no solamente aquello que ellos traducen directamente hacia una política o una acción económica. Hay una brecha entre lo que prefieren los atacantes y lo que ocurre en la práctica. Queremos invitar a la comunidad a que tome en cuenta esta brecha al hablar de los resultados o al traducir los resultados a situaciones.

Con esto termina mi presentación. Terminamos el trabajo de hoy. Le agradezco a todo mi equipo y a todos los que trabajaron en el proyecto INFERMAL y otros investigadores. Si tienen preguntas pueden contactarnos a mí o al líder del proyecto, a Maciej. También tenemos un email para nuestro equipo: ssr-research@icann,org, que acabo de poner en el chat. Gracias, Jonathan.

JONATHAN ZUCK: Gracias. Muchas gracias a todos. Les vamos a enviar las preguntas después. Les agradecemos por su tiempo. Tenemos que continuar con nuestro cronograma aquí en Praga. Les agradecemos por venir y deseamos que disfruten del café. Gracias.

[FIN DE LA TRANSCRIPCIÓN]