SAN JUAN – Indicadores de integridade de tecnologias de identificadores
Terça-feira, 13 de março de 2018 – 17h às 18h30 AST
ICANN61 | San Juan, Porto Rico

CATHY PETERSEN: Boa tarde a todos.Vamos começar a sessão sobre indicadores de integridade de tecnologias de identificadores em alguns minutos.Vamos esperar só mais um pouquinho.Obrigada.

ALAIN DURAND: Boa tarde.Esta é a sessão de indicadores de integridade de tecnologias de identificadores.Esse projeto começou há algum tempo, e hoje vamos mostrar alguns números interessantes [inaudível].Vocês vão achar esses números interessantes.Eu achei.

Nesta sessão, teremos três apresentadores.O primeiro será Paul Wilson, presidente atual da NRO.Ele vai dar notícias sobre o que a NRO está fazendo nessa área.

O segundo e o terceiro vão se concentrar na parte principal do projeto, nas coisas que são gerenciadas pela ICANN.A segunda apresentação será feita por Christian Huitema, sobre as métricas atuais e os dados que estamos descobrindo com essas métricas.

A última apresentação será feita por Geoff Huston, sobre o conjunto proposto de novas métricas.

Então, sem mais, vou passar a palavra para o Paul, que vai falar sobre as atividades no espaço de números.Paul?

PAUL WILSON: Obrigado, Alain, olá a todos.Os registros regionais da Internet são cinco, cada um de nós administra um registro de WHOIS usando um serviço de WHOIS mais ou menos conhecido.Então, há cinco registros diferentes, que são administrados pelos cinco RIRs.Eles são bem coordenados entre si.Tecnicamente, eles podem ser inconsistentes e, é claro que pode haver erros e falta de conclusão, etc.Então, os RIRs trabalham juntos dentro da NRO para garantir que os registros façam sentido um em relação ao outro e que cumpram a finalidade deles em relação aos registros que armazenam.

Fazemos isso há muito tempo, mas acho que as coisas estão mudando nos últimos anos, pois há um interesse cada vez maior de um grupo muito mais amplo de partes que confiam nos bancos de dados, na correção e na eficácia desses bancos de dados.Além disso, o ritmo das atualizações também está aumentando bastante.Então, antes que a falta de endereços IPv4 acontecesse, as alocações estavam bastante estáticas.Elas eram feitas para partes que mantinham e usavam essas

alocações.Mas, nos últimos dias, estão acontecendo muitas transferências.Então, dentro das regiões e entre os Registros Regionais da Internet, há muitas transferências acontecendo, que obviamente exigem que o banco de dados seja atualizado.Esse é outro motivo pelo qual o nosso foco na correção e na completude está aumentando.

Como eu disse, estamos preocupados desde a criação dos RIRs com que os registros façam seus trabalhos.Não falamos de integridade em si, mas a ideia de integridade de identificadores é algo novo, e acho que surgiu com o projeto da ICANN.Dito isso, mantivemos esse foco.

O outro aspecto disso, é claro, é que temos relações de associação com os operadores de rede que são os primeiros destinatários dos blocos de endereços e ASNs.Eles têm a obrigação, mediante acordos com cada RIR, de manter os registros atualizados.Existem questões de políticas em relação às expectativas e penalidades exatas, por assim dizer, por não obedecer essas políticas.

No geral, essas coisas estão sendo resolvidas em nível regional.Então, os cinco RIRs têm processos de políticas e associações independentes, e farão discussões sobre políticas relacionadas ao WHOIS em diferentes momentos.Essas discussões, como eu mencionei, também estão cada vez mais

frequentes e intensas nos últimos tempos.Eu diria que, nas cinco regiões, é justo dizer que estamos todos mais apertados de diferentes maneiras, mas, no geral, vamos na mesma direção rumo a políticas mais claras e implementadas com mais vigor.

O projeto de ITHI não foi uma surpresa de parte da ICANN.É óbvio que esse é um interesse compartilhado por todos os que desempenham funções de registro, por isso decidimos aderir a essa iniciativa da ICANN. Acho que foi muito útil, pois, embora trabalhemos juntos, não definimos um conjunto de métricas consistentes, e agora, com o projeto de ITHI, estamos caminhando nessa direção.

Na verdade, no ano passado, nosso grupo de coordenação de serviços, um grupo de funcionários dos cinco RIRs que trabalham em áreas de serviços de registro,trabalhou em um conjunto preliminar de métricas para a integridade de identificadores no espaço de números.Esse grupo também lançou uma consulta pública sobre um documento preliminar.

Isso aconteceu perto do fim do ano passado e, dessa forma, nossas comunidades tiveram a chance de fazer comentários sobre esse processo.Na verdade, recebemos poucos comentários, então agora temos um documento próximo da aprovação final para publicação.Basicamente, ele documenta a integridade dos identificadores no espaço de números em

relação aos registros do WHOIS.Nossa conclusão foi que os dados devem ser abrangentes, atuais e corretos.

Tudo isso foi dividido em cinco métricas específicas, que podem ser medidas, para o nosso banco de dados, em relação à completude e à exclusividade do banco de dados, à correspondência dele com outros registros oficiais externos, à eficácia dos dados em questão para alcançar as pessoas documentadas no registro e também à atualização dos dados.Esse documento também identifica os diferentes riscos associados a não alcançar as metas nessas medidas, além de analisar as causas que estariam associadas a esse tipo de falha.

Bom, esse documento será apresentado em breve.O que ainda não temos, obviamente, já que as métricas em si ainda estão em formato preliminar, é o que Alain vai apresentar logo mais, ou seja, dados reais sobre a nossa conformidade.Mas, obviamente, o objetivo de ter essas métricas é poder medir coisas para definir metas que esperamos cumprir, além de acompanhar o grau de conformidade ao longo de um período.

Então, acho que se vocês estiverem interessados no espaço de nomes, devem prestar atenção nesse espaço, poderemos informar sobre a integridade dos identificadores de Internet no espaço de números nos cinco RIRs.Acho que isso é tudo.Obrigado, Alan.

ALAIN DURAND: Obrigado, Paul.Quero aproveitar essa oportunidade para agradecer você e os outros membros da comunidade de números por trabalhar conosco nesse projeto.Acho que é uma colaboração interessante, e estamos aprendendo muito durante esse processo.

PAUL WILSON: Sim, eu concordo.Agradeço a você também, Alain.Obrigado.

ALAIN DURAND: Agora, estamos seguindo o mesmo processo, como analisar o espaço do [problema] e definir as métricas, depois fazer a medição real, então isso é normal.Mas vocês ainda não têm os números, e no futuro queremos ver o primeiro lote de números quando eles estiverem prontos.

Agora vamos passar para o espaço de nomes.Christian vai fazer uma apresentação sobre como estamos.

CHRISTIAN HUITEMA: Boa tarde.Sou Christian Huitema.Comecei a trabalhar na medição dos dados e do [status] do DNS há um ano e meio, depois de fazer um estudo para ver o que poderia ser feito com o DNS, trabalhando com Alain nas métricas em si.

Como Paul disse, precisa haver alguns princípios na hora de definir essas métricas.Os princípios que adotamos basicamente são os que estão neste slide.Primeiro, queríamos muito que essa operação de métricas fosse técnica.Não queremos nos envolver em políticas.A finalidade das métricas é descrever o estado do sistema.Não é fazer julgamentos de uma forma ou de outra.

Estamos querendo definir áreas que queremos acompanhar que podem ser problemáticas, definir as métricas nessas áreas e definir maneiras de medi-las.

Outro princípio é que não queremos imagens pontuais.Queremos um sistema contínuo que funcione por muito tempo e, basicamente ofereça as métricas necessárias todos os meses, publique um novo valor e, é claro, publique o valor dos últimos meses também para que possamos estimar tendências.Porque acreditamos que as tendências são quase tão importantes quanto os valores em si.

É por isso que estamos investindo tanto em automação.Basicamente, estamos fazendo testes em vários lugares, com feedback e automação constantes.De forma que o site que publica os dados seja automatizado, que as métricas sejam produzidas automaticamente todos os meses, etc.

Nos slides que vamos apresentar, vamos mostrar as medidas.Pelo mesmo motivo que não queremos nos envolver

em políticas, queremos mostrar as medidas como são.Sempre que vemos um número, dizemos: “ah, X agora está em 29%.Por quê?”Bom, nossa resposta genérica é: “Não sabemos por quê”.Ou seja, temos opiniões, mas elas valem tanto quanto as de vocês.Não queremos colocar essas opiniões na publicação das métricas.As métricas são medidas diretas.

Outro princípio é que somos muito cuidadosos para não ter problemas de privacidade.Então, todos os dados que vamos publicar são de natureza estatística.Todas as nossas ferramentas são de código aberto, e todos os nossos resultados são publicados para que possam ser analisados.

Fizemos algumas apresentações nas sessões anteriores.Já fizemos a apresentação das métricas de ITHI em Abu Dhabi, por exemplo.Elas estão em sete categorias para nós.Primeiro, analisamos a precisão dos dados de WHOIS.Em seguida, analisamos o comportamento dos servidores raiz e o nível de abuso que estão sofrendo.Desculpem, o abuso de nomes de domínio, o abuso do sistema de nomes de domínio.Analisamos o tráfego na raiz do DNS.

Para todas as métricas indicadas aqui, temos fontes de dados.Por exemplo, no caso do WHOIS, estamos trabalhando para o departamento de conformidade da ICANN.No caso do abuso de nomes de domínio, estamos trabalhando para o

projeto DAAR.Para medir o tráfego na raiz, servidores recursivos, registros da IANA para parâmetros de DNS, e implementação de DNSSEC, estamos trabalhando com o escaneamento do tráfego na raiz ou do tráfego de resolvedores recursivos.Além disso, estamos colaborando com resolvedores recursivos para [testar] efetivamente e conseguir essas estatísticas.

Quanto ao cronograma, estamos trabalhando há um ano na definição das métricas.O que temos por enquanto é uma apresentação dos primeiros dados.Nos últimos dois meses, fizemos as capturas iniciais e conseguimos dados para M1, M2, M3 e M7.Geoff Huston vai apresentar os dados da M5 na próxima apresentação.Também conseguimos, com colaboração desde o início, um conjunto de dados inicial para as métricas M4 e M6, que são sobre o uso do DNS pelos clientes.

Então, vamos integrar a M5 quando ele for desenvolvido.Vamos desenvolver o caminho e fazer mais testes para ter dados mais completos a partir das métricas M4 e M6.Vamos complementar esses dados e publicá-los no site de ITHI da ICANN.

Primeiras métricas, M1.A M1 avalia a precisão dos dados de WHOIS.Vamos fazer isso usando um proxy de precisão, que é o número de denúncias.Não usamos o número [real] de denúncias.Usamos o número de denúncias validadas pelo departamento de conformidade da ICANN.No momento, esse

número é um pouco menos de 6 por milhão.Esses são os nossos primeiros dados, então ainda não temos uma tendência.Mas vamos acompanhar essa tendência ao longo do tempo.

Com todos esses dados, o que vemos é que a média não conta a história.Se eu disser que há 6 denúncias por milhões de [números] de domínios registrados em média, bom, essa é só uma média.Criamos uma curva aqui que é a [inaudível] frequência de denúncias.Basicamente, o total de denúncias no eixo Y e, depois, no eixo X, o número de registradores classificados a partir do que tem mais denúncias até o que tem menos.

O que vemos aqui é que a distribuição não é [inaudível].Se todos os registradores tivessem o mesmo número de denúncias, vocês veriam uma linha reta em diagonal.Isso não é o que vocês estão vendo.O que acontece aqui é que a linha está bastante curva, um pouco inclinada em direção ao eixo Y.Na verdade, são necessários seis registradores para responder por pelo menos 50% das denúncias.Ou seja, não é um número homogêneo, seis registradores respondem por um pouco mais de 50%.São necessários 44 registradores para responder por pelo menos 90% das denúncias.No total são quase 2.000 registradores.Então, a distribuição aqui é muito [enviesada].

Como eu disse, esses números são [inaudível].Não é uma opinião nem um racionamento sobre o motivo.Mas é o que observamos.

CATHY PETERSEN: Desculpe, Christian.Temos uma pergunta online.

CHRISTIAN HUITEMA: Sim?

CATHY PETERSEN: É de Kathy Kleiman, "Como vocês sabem que as denúncias sobre o WHOIS são válidas?Sabemos que algumas delas são feitas por perseguição".

ALAIN DURAND: Vou responder essa pergunta.Estamos trabalhando com o departamento de conformidade da ICANN.Não analisamos todas as denúncias.Analisamos apenas as denúncias relacionadas à precisão dos dados.Existem muitos outros tipos de denúncias que não estamos levando em consideração.

O departamento de conformidade da ICANN tem um processo de análise e avaliação dessas denúncias.Se eles acham que as denúncias têm fundamentos, então eles enviam o chamado primeiro aviso.Se não houver resposta, então eles enviam o

segundo aviso e depois o terceiro e, possivelmente, seguem todo o caminho até definir a [violação].É um processo bem definido, bem documentado no departamento de conformidade da ICANN.

Então, para responder a essa pergunta mais uma vez, consideramos apenas as denúncias relacionadas à precisão do banco de dados do WHOIS de um registro e as denúncias que foram validadas, que passaram pela etapa do primeiro aviso.

CHRISTIAN HUITEMA: Obrigado, Alain.Então, essa é a métrica M1 sobre a precisão do WHOIS.A série de métricas M2 é sobre o abuso do sistema de nomes de domínio, e estamos trabalhando com o projeto DAAR para isso.Eles estão acompanhando quatro tipos de abuso: o número de sites usados pelos domínios de phishing, o número usado por domínios de malware, o número de comandos e controles de botnets, e o número de domínios de spam.A métrica é definida como o número de domínios que sofreram abuso a cada 10 mil domínios.

Vemos aqui as médias globais, que basicamente estão na ordem de 4 ou 3 para os primeiros três tipos de abuso, e um valor muito maior para os domínios de spam, pois o spam é uma atividade muito distribuída.

Assim como com a M1, também vemos aqui que essas métricas não explicam tudo.Analisando a distribuição por TLDs, vemos que, por exemplo, em relação ao phishing, um só TLD responde por mais de 50% de todos os domínios de phishing.São necessários apenas 11 gTLDs para responder por todos os domínios de phishing.Vemos o mesmo tipo de distribuição enviesada para os outros domínios.Isso indica claramente a estrutura do problema.

Estamos tentando fazer a mesma medida para os registradores, mas não queremos passar muito tempo com os dados de registradores aqui porque esses dados precisam ser avaliados com o processo do WHOIS e estão sujeitos a todas as restrições de usar dados de WHOIS, como a limitação de tempo e tudo isso.Então, publicamos apenas quando recebemos dados que podemos comprovar e confiar, e o que temos hoje são informações preliminares.

Mas a ideia é apresentar esse viés dos dados em algum tipo de tabela, como por exemplo, quantos gTLDs respondem por pelo menos 50% dos domínios de phishing, dos domínios de malware, etc. Depois, quantos respondem por pelo menos 90% dessas variações.Como eu disse, nosso trabalho é medir as coisas.Não fazemos interpretações, nem considerações sobre os motivos.

Os dados da M1 e da M2 são produzidos pelo departamento de conformidade da ICANN e pelo projeto DAAR, e são sobre a qualidade dos dados.Os dados da M3 e da M4 que veremos mais tarde são sobre o tráfego do DNS em si, o que vemos aqui.A M3 é sobre o tráfego na raiz.Para medir o tráfego na raiz, instrumentamos o servidor de raiz "L".Basicamente, fazemos uma amostragem por dia por servidor de raiz “L”.Essas amostragens são feitas em horários aleatórios, então respondem por todas as variações de horários quando são agregados estatisticamente.Depois, pegamos todas essas amostras e fazemos um resumo todos os meses para conseguir essas métricas.

O que vocês veem aqui é a primeira métrica: que porcentagem das consultas à raiz recebem a resposta “esse domínio não existe”?O número é bem grande.Efetivamente são quase dois terços do tráfego da raiz, consultas sem valor específico.Depois, sobre as consultas restantes, analisamos quantas delas poderiam ter sido armazenadas em cache pelo resolvedor.Mais uma vez, vemos que é um número bastante grande, 30%.As consultas que não sabemos se poderiam ter sido armazenadas em cache e provavelmente não poderiam, são por volta de 6 a 6,5%.Fazemos o acompanhamento todos os meses.Vocês podem ver aqui o valor atual e a média, além do gráfico de pizza que mostra como os domínios são divididos.

Em relação às consultas “esse domínio não existe”, que é uma parte importante da pizza aqui, tentamos dividir essa questão em componentes.O que provoca isso?Descobrimos que precisamos analisar quatro componentes: os nomes reservados, os nomes que foram reservados pela IETF, como por exemplo .local (existem cinco ou seis deles, que respondem por aproximadamente 3,4% do tráfego), as cadeias de caracteres vazadas com frequência, por exemplo .home, que respondem por 9,3% do tráfego, e os padrões frequentes, quando vemos um padrão nos dados.Não são cadeias de caracteres frequentes.Cada nome aparece apenas uma pequena fração do tempo.Existem muitos nomes diferentes, mas eles seguem padrões, e tentamos analisar esses padrões.Depois as outras coisas.Existem aproximadamente 10% que não podemos explicar diretamente por nenhum desses processos.

No caso dos nomes de uso especial definidos na RFC 6761, o que vemos é que o maior volume de uso é do domínio .local.É aproximadamente 2,77% do tráfego na raiz hoje.Existem outros domínios reservados, mas eles estão presentes em números muito menores:.localhost está bastante presente, .invalid aparece bastante, depois os outros domínios são apenas vestígios.

Em relação aos domínios vazados com frequência, o que fazemos aqui é pegar as cadeias de caracteres mais frequentes

na raiz e na variação atual na implementação atual, analisando apenas as cadeias de caracteres que aparecem pelo menos ,01% das vezes.

Neste slide específico, temos apenas cadeias de caracteres que aparecem pelo menos ,02% das vezes porque, quanto menor o número, menos certeza podemos ter sobre os resultados.Isso também deixaria o PowerPoint muito difícil de ler.

Mais uma vez, vemos que existe um nome dominante, .home, que responde por 3,5% dessas solicitações que a raiz recebe.Depois, temos uma série de outros nomes.Com o tema de .home, podemos medir de forma absoluta o vazamento desses nomes na raiz, e podemos fazer o rastreamento mês a mês, e sabemos quais nomes estão sendo usados e quais são recorrentes.Podemos ver que há mudanças mês a mês.Alguns nomes vão desaparecer, mas podemos ver que há um núcleo de nomes bem usados que aparecem o tempo todo.

Eu já disse que vários nomes que vemos nesse tráfego da raiz não são domínios de uso especial e não correspondem às cadeias de caracteres frequentemente usadas.São apenas nomes aleatórios.Na verdade, vocês veem aqui nesta distribuição, fizemos uma distribuição desses nomes por comprimento.Vemos que a maior parte desses nomes têm de 7 a

15 caracteres.Não incluímos os nomes mais longos porque são muito poucos.

Muitos desses nomes com 7 a 15 caracteres, quando faço uma amostragem aleatória, parecem nomes gerados aleatoriamente por computadores.Nem todos são assim.Na verdade, é bem difícil diferenciar o que é gerado aleatoriamente por um computador e o que são apenas planos de numeração de redes Wi-Fi, por exemplo.Mas queremos rastrear isso e ir mais além, fazer mais análises.

Estou falando do tráfego na raiz.Bom, quando fizemos esse primeiro estudo no ano passado, fizemos algumas experiências e rapidamente chegamos à conclusão de que a raiz não era necessariamente representativa de todo o tráfego de usuários.Quem entende a arquitetura do DNS sabe que o que é considerado raiz já foi filtrado pelos resolvedores do DNS.Na verdade, se os resolvedores do DNS aplicassem todas as tecnologias modernas definidas pela IETF, haveria pouquíssimo tráfego na raiz.Os bons resultados seriam armazenados em cache.Os resultados [insatisfatórios] seriam armazenados em cache.Não veríamos nada disso.Então, grande parte do tráfego na raiz corresponde a comportamentos anômalos.

Para analisar o que os usuários estão fazendo de verdade, precisamos estar próximos aos clientes,Por isso estamos

trabalhando com resolvedores recursivos para fazer testes e tentar analisar o que está acontecendo lá.Quantas das consultas feitas pelos clientes vão para TLDs registrados em vez de todas essas cadeias de caracteres que vemos na raiz?Quantas vão para os nomes reservados da IETF?Quantas vão para as cadeias de caracteres usadas com frequência que vemos aqui e o que mais?

Bom, vocês lembram que quando analisamos o tráfego na raiz, vemos que as consultas a TLDs não existentes representam quase dois terços do tráfego.Aqui, no único teste que temos... preciso qualificar nossos dados, temos apenas um ponto de medida hoje.Estamos trabalhando para conseguir mais.Nesse ponto de medida, esses TLDs não existentes representam apenas 1% do tráfego, muito menos.

As tendências também são diferentes.Nos nomes reservados, vemos um menor volume de tráfego com .localhost, .local e praticamente nada para os outros nomes.

Nas cadeias de caracteres usadas com frequência, ficamos um pouco surpresos.O domínio é dos nomes locais, como nomes de host que as pessoas tentam resolver, não fazem as consultas corretamente e terminam enviando a [consulta] colocando o nome de host como um [token] que poderia ser confundido com um domínio de primeiro nível.Não publicamos os valores desses

nomes porque existem questões de privacidade.Costumam ser nomes na infraestrutura local das pessoas que [fornecem] os testes, então colocamos todos eles em uma categoria global de "nomes de host locais".

Além disso, o que vemos é pouco tráfego para esses tipos de nomes que vemos na raiz.Vemos algum tráfego para os grandes nomes, como .home, e o tráfego para nomes como .dns, .internal ou .unifi, que no caso representa a rede Wi-Fi usada.Essa foi das lições que aprendemos.No momento, queremos fazer muito mais testes para poder fazer declarações definitivas, mas vemos que existe uma diferença entre o tráfego nos clientes e na raiz.

Quer dizer alguma coisa?

ALAIN DURAND: Quero complementar o que o Christian acabou de dizer.Até agora, estamos trabalhando com algumas organizações pequenas, duas delas já concordaram em participar e estão contribuindo com dados.Quero agradecê-las aqui.

Uma delas é a Universidade de Cape Coast em Gana, outra é a Universidade de La Plata na Argentina.Agora, também estamos trabalhando com mais uma organização, Nawala, que é [mais ou menos] um provedor de serviços na Indonésia.Ontem à noite,

ficamos acordados até tarde tentando ajudá-los a instalar as ferramentas para fazer todas essas medições.

Vamos conversar com outros possíveis parceiros, e nosso objetivo é conseguir mais participantes.Se conseguirmos cinco, seis talvez até dez até o fim do ano, ficaremos muito felizes.Queremos conseguir diferentes tipos de participantes, alguns acadêmicos, alguns mais industriais, alguns provedores de serviços, alguns pequenos, outros grandes ou até muito grandes.

Mas vamos começar com uma abordagem [inaudível].Começamos pequeno.Por isso, conseguimos entender como as coisas funcionaram para ajustar as ferramentas que temos.Agora, estamos desenvolvendo um processo para que isso seja mais automático.Podemos entrar em contato com participantes maiores, até mesmo muito maiores.

Então, queria agradecer ao Christian por [desenvolver] a ferramenta e ajudar na implementação.

CHRISTIAN HUITEMA: Obrigado.Bom, o Alain também passou muito tempo implementando essas ferramentas.Essa infraestrutura mundial

faz a gente passar muito tempo em ligações telefônicas ou bate-papos no meio da noite.Mas são ossos do ofício, eu diria.

Então a M3 e a M4 são análises de duas partes do tráfego.Que tipo de tráfego de DNS vemos na raiz e no cliente?Com os dados da M4, também queríamos ver a utilidade de todos esses registros da IANA que estamos fazendo para a IETF.Não podemos rastrear todos os registros da IANA porque temos apenas dados [relacionados] ao DNS.Mas o que fizemos no caso das tabelas [relacionadas] ao DNS foi analisar os parâmetros que fazem parte dos registros.Por exemplo, os tipos [r] das classes [código r], mas também os parâmetros usados pelas DNSSEC ou os parâmetros usados pelo DANE.

No caso desses parâmetros, queríamos responder a duas perguntas.A primeira é: as pessoas realmente usam os dados registrados?Basicamente, o que fizemos foi dizer: “bom, se uma tabela define dez valores, quantos deles realmente vemos pelo menos uma vez no conjunto de dados?”No caso de algumas tabelas, a resposta é zero.Existem algumas tabelas assim.

No caso das tabelas clássicas, como as classes do DNS ou os números de algoritmo, a resposta está entre 20% e 70%.Alguns valores são usados raramente.Por exemplo, no caso dos algoritmos de segurança, alguns algoritmos de segurança estão obsoletos e não são mais usados.Mas podemos ver isso.Dessa

forma, temos uma ideia é a confiança de que o que a IANA está fazendo é útil.

Outra coisa que queríamos ver é se as pessoas estavam ignorando o registro da IANA e criando diretamente os próprios valores.Nesse conjunto, vemos isso apenas no caso dos códigos de opção do DNS, o código de opção EDNS0 do DNS, embora haja certo uso dos valores experimentais que vemos no total.Globalmente, as coisas estão assim.

Agora, gostaria de fazer uma observação sobre o uso do certificado TLSA e do certificado gerado pelo DANE.Eles não estão nos meus dados.Tive uma longa conversa com Victor Dukhovny sobre isso.Ele me disse que era normal porque a maior parte do uso do DANE é entre servidores de e-mail e servidores autoritativos.O servidor de e-mail consulta diretamente o servidor autoritativo.Então, o tráfego não é registrado nos nossos pontos de teste.Estou trabalhando com ele para conseguir informações diretas sobre o tráfego que ele tem nas medições do DANE, para que possamos avaliar adequadamente o uso das tabelas do DANE.

Essa é basicamente a forma em que podemos usar essas medidas para acompanhar a IANA.Não acompanhamos apenas uma tabela.Apresentei os dados de quatro ou cinco tabelas no slide anterior.Aqui, temos a lista completa que estamos

montando, podemos até adicionar mais tabelas quando descobrirmos como analisar e extrair os dados.

A última métrica, M7, é sobre a implementação das DNSSEC.Começamos essa avaliação da implementação das DNSSEC analisando a zona raiz para ver quantos TLDs estavam fornecendo chaves de DNS.Esse número é bastante estável, aproximadamente 90%.Mas esperamos que isso mude ao longo do tempo e chegue a 100%, acontece que as mudanças são muito lentas.

Bom, analisando os dados da M4, percebemos que vemos grande parte do tráfego que na verdade era tráfego de segurança do DNS.Vemos isso porque podemos perceber quando um cliente usa a segurança do DNS e coloca a parte DO nas consultas que [inaudível] na resposta.Então, podemos medir a fração das consultas que têm essa parte e dizer: “Ah, se pudermos descobrir quais clientes estão fazendo isso, saberemos quantos clientes estão usando as DNSSEC”.

Dessa forma, podemos adicionar a esses dados o que queremos fazer na M7.2, que é a porcentagem de consultas sobre DNSSEC de clientes que estejam usando as DNSSEC.Sendo ambiciosos, também podemos ver a porcentagem de consultas de resolvedores recursivos que usam as DNSSEC e, algo muito interessante, a porcentagem de respostas de servidores

autoritativos que fornecem as respostas das DNSSEC.Acho que, fazendo isso, vamos ter uma ideia sobre o uso real das DNSSEC e podermos responder à pergunta: qual é o volume de uso das DNSSEC?Acho que isso seria interessante para a comunidade.

Então, já vimos seis das nossas sete métricas.Geoff Huston vai apresentar a métrica 5 depois de mim.A M7, como eu disse, é muito estável, então esse gráfico não diz muito agora.

Quero agradecer a atenção de vocês e responder perguntas agora, se houver alguma.

RUBENS KUHL: Rubens Kuhl, .br.Só queria comentar que o servidor recursivo, o DNS recursivo e as métricas recursivas se baseiam em três servidores recursivos.E atualmente temos 50 mil [inaudível] sistemas na Internet, então publicar esses resultados antes de chegar a pelo menos 5 mil servidores recursivos não é uma boa ideia, pois não existe relevância estatística.É como colocar um [inaudível] no microscópio e deduzir que todo o tecido do mundo se baseia nisso.

[Fiquei espantado] que a ICANN tenha publicado essa métrica, especialmente em uma área em que a ICANN não tem dados diretos, diferente do caso dos dados da raiz autoritativa, em que a organização gerencia um dos sistemas de servidores raiz mais

abrangentes porque é a instância [inaudível].Então, um servidor raiz tem uma boa significância estatística entre as consultas à raiz.Mas quanto às consultas recursivas, não devemos publicar esses dados até que eles atinjam um limite mínimo de relevância estatística.

CHRISTIAN HUITEMA: É uma boa observação.Estamos usando muitas limitações nessa conversa para explicar que temos apenas um ponto de medida agora e explicar que [inaudível] queremos. Está claro que queremos mais de um ponto de medição.Não sabemos se precisamos de 5 mil.Seria ótimo ter 5 mil, mas não sabemos se precisamos de 5 mil.

Minha ideia é comparar os dados de muitos sites que decidam participar para ver as diferenças e semelhanças.A ideia é saber que existem diferenças.Existem diferenças de horário, como de manhã e à noite.No fim de semana e em dias de semana.Sabemos que existem diferenças geográficas.As pessoas não pedem o mesmo tráfego na China e nos Estados Unidos.Sabemos que existem diferenças no tipo de uso.As pessoas não fazem as mesmas consultas em ambientes acadêmicos, governamentais, empresariais, em redes privadas ou em redes móveis.Então, claramente, queremos ter uma representação de tudo isso.

Vamos abordar a significância estatística.Definitivamente queremos fazer isso, mas precisamos começar em algum lugar, então vamos coletar os dados.Além disso, vamos comparar as fontes para poder responder à sua pergunta.

RUBENS KUHL: Sim, mas gostaria de responder a isso.Essas limitações são explicadas, mas normalmente elas não têm muito destaque.Então, qualquer pessoa que leia o relatório pode repetir os dados e publicá-los na imprensa e nas rede sociais sem reproduzir as limitações e dizer que esses dados na verdade não são significativos.Por isso, publicar esses dados é um desserviço para a comunidade.Essa é minha opinião sobre isso.

Também tenho um comentário sobre outro assunto mencionado, que algumas consultas de registradores foram afetadas por limitações do WHOIS, etc.Posso coletar dados de todos os registros [Thick] do BRDA, ou seja, os dados de registro thin em massa.Esses dados de registro thin já indicam quais registradores estão associados a cada domínio.Por isso, não é necessário fazer a consulta ao WHOIS.Já existem dados na ICANN que fornecem essas informações com 100% de precisão, então acho que vocês também deveriam analisar isso.

CHRISTIAN HUITEMA: É uma boa observação.Muito boa.Quero responder primeiro dizendo que o projeto ITHI é cliente do projeto DAAR.Estamos recebendo dados do DAAR, então herdamos todas as decisões deles.Por isso, primeiro quero sugerir que você redirecione a pergunta aos administradores do DAAR.

Em segundo lugar, [vou tentar] representá-los.Sob o meu ponto de vista, eles queriam que o estudo fosse replicável, ou seja, que uma pessoa fora da ICANN possa replicar o mesmo estudo de forma exata, com metodologia aberta e dados acessíveis.Os dados que você mencionou podem ser ou não acessíveis por pessoas externas, dessa forma a ICANN seria a única que poderia realizar esse estudo.Eles não queriam ir nessa direção.Eles podem mudar a abordagem em algum momento, e talvez John Crain possa falar um pouco sobre isso, mas até agora é o caminho que eles escolheram.Como clientes deles, estamos herdando essa decisão.

Então, a pergunta é, por que precisamos depender do WHOIS para fazer a atribuição a um registrador em vez de usar dados internos da ICANN?

JOHN CRAIN: Se tivermos todos os dados disponíveis internamente, não encontrei esses dados internamente.Seria ótimo.Mas uma das coisas que estávamos tentando fazer era possibilitar a

replicação por outras pessoas, o que significa usar fontes externas.A única coisa que precisamos do WHOIS é o ID do registrador.Estávamos falando antes sobre onde poderiam estar as fontes internas, então talvez passemos a usá-las porque acho que o WHOIS não será prático.

RUBENS KUHL: A fonte se chama BRDA, então vocês podem analisar isso ou entrar nesses servidores e pegar os dados.Mas mesmo se vocês usarem os dados, seria possível reproduzir isso, só seria um pouco mais complicado para as outras pessoas usarem as consultas ao WHOIS.Mas elas podem reproduzir usando as consultas ao WHOIS porque são as mesmas informações.Ou seja, as informações não são privilegiadas de qualquer forma.

JOHN CRAIN: Sim, entendi.Quando começamos o projeto, pensamos que tudo deveria ser exatamente como as pessoas de fora fariam.Você tem razão quanto à capacidade de reprodução, então vamos reconsiderar.

RUBENS KUHL: Certo.

ALAIN DURAND: Muito bem.Temos alguma pergunta na sala de bate-papo, Cathy?

CATHY PETERSEN: Não.

ALAIN DURAND: Sem perguntas?Certo.Então, obrigado, Christian.

CHRISTIAN HUITEMA: Obrigado.

ALAIN DURAND: Muito obrigado por apresentar os números pela primeira vez aqui.Agora, quero convidar Geoff Huston.Geoff está aqui?

CATHY PETERSEN: Temos uma pergunta aqui.

ALAIN DURAND: Ah, temos uma pergunta.

SEBASTIEN BACHOLLET: Já que Geoff não está na sala, só queria fazer uma pergunta, como não tenho experiência técnica.Existe algum vínculo entre

o que vocês estão fazendo e algumas perguntas sobre a renovação da chave e os dados necessários para entender o que está acontecendo?Desculpe, é [inaudível] pergunta de [inaudível].Obrigado.

CHRISTIAN HUITEMA: Hoje, a resposta é não. Não há vínculos.Não consideramos essa métrica inicialmente.Mas no futuro, pode haver mais renovações, que podem ser mais ou menos frequentes, talvez seja necessário rastreá-las.Hoje, falamos de sete métricas.Nós as entendemos bem o suficiente para medi-las.Agora, estamos pensando na segunda fase, em que vamos adicionar mais métricas e analisar outros tipos de problemas.Essa pode ser uma das áreas que precisamos analisar e agregar ao que estamos fazendo agora.[responde também em francês]

SEBASTIEN BACHOLLET: Entendi.Só quero garantir que minha pergunta seja bem explicada, peço desculpas.Hoje, parece que faltam dados para ter certeza de que é o momento certo para fazer a renovação da chave.Não é o fato de quando faremos a renovação da chave, é se a cada ano vocês conseguirão coletar dados [como pergunta], se com os dados do projeto de vocês haverá dados que poderão ser usados pelas pessoas que precisam decidir quando fazer a renovação da chave.

CHRISTIAN HUITEMA: Hoje, não temos dados para ajudar com isso.

SEBASTIEN BACHOLLET: Obrigado.

PAT KANE: Olá.Pat Kane da Verisign.Só um complemento da pergunta do Sebastien.Hoje, mais cedo, o CTO da ICANN relacionou a diminuição das consultas às DNSSEC à renovação da KSK, que começou o ano passado e vai até outubro deste ano.Então, acho que é importante entender, a partir do uso das DNSSEC, qual é a relação disso, para embasar essa decisão, pois muitos dos dados em relação de pessoas com resolvedores que não têm os dois pares de chaves estão piores que no ano passado.Então, seria muito bom fornecer essas informações ao David o mais cedo possível.Obrigado.

ALAIN DURAND: Obrigado.É uma boa observação.Como vimos, Christian estava falando sobre uma nova métrica, M7.2, que vai acompanhar o número de consultas que têm o conjunto DO.Isso pode ajudar com esse tema com outras métricas que estamos tentando desenvolver nesse espaço específico.Então, talvez possamos ter

uma conversa offline se vocês tiverem ideias específicas sobre o que devemos rastrear.

O Geoff voltou para a sala?OK, então peço desculpas.Perdemos um dos nossos apresentadores.Posso falar rapidamente sobre o que estamos planejando.

Inicialmente, a métrica M5 era uma das que também analisava os resolvedores, mas sob o ponto de vista dos clientes.Pedimos para o Geoff analisar isso, e ele tem um sistema de medida baseado em Google Ads, que é bem conhecido e já foi usado em outros contextos.Pedimos para ela usar o sistema para explorar o que pode ser feito sob o ponto de vista de um cliente, os resolvedores de stub.

Uma das coisas que queremos analisar é: os resolvedores realmente estão percebendo as coisas?Às vezes, achamos que sim.Às vezes, achamos que talvez não ou que talvez eles possam armazenar em cache por menos ou mais tempo.Achamos que podemos conseguir uma medida disso.

Também podemos analisar a distribuição das DNSSEC e do IPv6 para descobrir se o resolvedor está configurado com DNSSEC ou não, ou se é capaz de usar IPv6 ou não.Possivelmente, também podemos encontrar os resolvedores mais usados.Eu deveria poder qualificar os resolvedores mais usados a olho, pois o sistema usa o Google Ads, e eles são usados por usuários físicos,

não máquinas.Então, ele não registra a comunicação de máquina a máquina, mas sim de usuário a máquina.

Esse pode ser um projeto [inaudível] de medida que também poderia nos instruir sobre a renovação da chave e quantos resolvedores são necessários para cobrir 95% ou a porcentagem que seja da população.

É um trabalho em andamento.[Essas] são as novas métricas que o Geoff gostaria de propor.Seguindo com o que o Christian disse antes, queremos automatizar tudo isso para poder coletar as medidas e rastreá-las ao longo do tempo durante vários anos.

Então, resumindo, esse é o projeto que gostaríamos de fazer com o Geoff.

Peço desculpas pela ausência dele, deve ter sido por alguma circunstância externa.

Se não houver mais perguntas, vamos encerrar a sessão mais cedo.Ah, temos uma pergunta.

RUBENS KUHL: Na verdade, é uma resposta ao comentário de Pat Kane.Por que o número de resolvedores das DNSSEC informados aumentaram o número [inaudível] KSK de 2010?Não sabemos ainda se esses resolvedores fazem a validação ou não.Então, poderia ser uma

pessoa que só tem a chave da raiz, mas que não valida.Então, não será um problema possível quando a chave for implementada.Então, se fizermos um estudo assim em uma métrica, deveríamos analisar resolvedores de validação com chaves antigas, não apenas resolvedores que informam chaves antigas.Porque não é algo que mede algo que poderia prever o que acontecerá quando implementemos a chave da raiz.

ALAIN DURAND: É uma ótima observação, mas vou complementar.De alguma forma, deveríamos pesar isso pelo número de usuários por trás desse resolvedor.Se é apenas uma coisa usada no porão de alguém, ligada por cinco minutos, talvez não tenha a mesma importância que um resolvedor que atende a milhões de clientes.

RUBENS KUHL: Concordo.

ALAIN DURAND: Então, se não houver mais comentários, vamos encerrar a sessão.O próximo encontro da ICANN será sobre políticas, não haverá sessões técnicas, por isso não vamos nos reunir.Mas veremos todos vocês em Barcelona.

[FIM DA TRANSCRIÇÃO]