Сан-Хуан – Индикаторы работоспособности технологий идентификаторов
Вторник, 13 марта 2018 года, 17:00 – 18:30 по AST
ICANN61 | Сан-Хуан, Пуэрто-Рико

КАТИ ПЕТЕРСЕН (CATHY PETERSEN): Добрый день! Мы начинаем заседание, посвященное индикаторам работоспособности технологий идентификаторов через несколько минут. Давайте подождем еще пару минут. Спасибо.

АЛАН ДЮРАН (ALAIN DURAND): Добрый день. Это заседание по индикаторам работоспособности технологий идентификаторов. Этот проект работает уже длительное время и сегодня мы собираемся рассмотреть несколько интересных цифр [неразборчиво]. Цифры, которые могут представлять интерес для вас. Для меня они представляют интерес.

На этом заседании у нас будет 3 докладчика. Первый — Пол Уилсон, действующий председатель Организации ресурсов нумерации. Он сделает доклад о том, чем занималась организация ресурсов нумерации в этой области.

Второй и третий докладчики рассмотрят основную часть проекта: вещи, которыми управляет ICANN. Второй доклад сделает Кристиан Хайтема (Christian Huitema), она будет

касаться текущих показателей и данных, которые мы можем получить из текущих показателей.

Последний доклад будет касаться предложенных наборов [некоторых] новых показателей. Его сделает Джефф Хьюстон (Geoff Huston).

Поэтому без лишних слов передам слово Полу, который расскажет о деятельности в пространстве номеров. Пол?

ПОЛ УИЛСОН (PAUL WILSON): Спасибо и привет всем. Региональные интернет-регистратуры управляют пятью регистратурами WHOIS, использующим достаточно знакомую службу WHOIS. Итак, имеются пять различных регистратур, управление которыми осуществляют пять RIR. Координация между ними достаточно плотная. Технически между ними могут возникать расхождения и, конечно же, ошибки и незавершенность и т.д. Поэтому RIR работают совместно под юрисдикцией NRO, чтобы обеспечить отсутствие конфликтов между этими регистратурами и что они выполняют свои задачи в отношении записей, которые в них хранятся.

Мы занимаемся этим уже очень длительное время, но я считаю, что все несколько изменилось за последние несколько лет и на данный момент имеется намного более значительная заинтересованность намного более обширной

группы сторон, использующих базы данных, в правильности и эффективности этих баз данных. Также темп обновлений растет достаточно быстро. Перед тем, как мы столкнулись с дефицитом адресов IPv4, распределение адресов было практически статичным. Адреса были предоставлены в распоряжение сторон, которые их поддерживали и использовали. Однако сегодня у нас много переносов. В регионах и между региональными интернет-регистратурами происходит много переносов, требующих обновления базы данных. Это еще одна причина, почему мы постоянно говорим о правильности и полноте.

Как я сказал, с момента создания RIR мы всегда прилагали много усилий, чтобы регистратуры хорошо выполняли свою работу. Мы не говорили о работоспособности как таковой, однако идея работоспособности идентификаторов — это нечто новое, что, как мне кажется, возникло вместе с проектом ICANN. С учетом сказанного мы поддерживали такой подход.

Другим аспектом этого, конечно же, является то, что у нас есть членские отношения с сетевыми операторами, являющимися первыми получателями адресов и ASN. Поэтому, согласно своим обязательствам перед каждой RIR, они должны поддерживать актуальность и обновлять записи. Имеются проблемы политики в отношении того, какие именно

ожидания и штрафы, так сказать, должны быть за невыполнение этих политик.

Управление всем этим, в основном, осуществляется на региональном уровне. Таким образом, у RIR имеются независимые процессы политик и членства, и они в разное время проводят обсуждения политик, касающихся WHOIS. Эти обсуждения, как я упоминал, проводятся все чаще и интенсивнее в последнее время. В отношении этих пяти регионов, думаю, будет честно сказать, что для всех нас условия станут несколько жестче в различных отношениях, но в целом мы движемся в одном направлении к более ясным и более твердо реализованным политикам.

Источником проекта ITHI стала ICANN. И это неудивительно. Это, очевидно, представляет совместные интересы всех нас, выполняющих роли регистратур, и поэтому мы приняли решение включиться в инициативу ITHI ICANN. Я считаю, что это было достаточно полезно, потому что, несмотря на то, что мы работали вместе достаточно тесно, у нас не было набора надежных показателей, которые мы сейчас продвигаем в проекте ITHI.

Фактически, мы проработали весь прошлый год, прошли через обсуждение в координационной группе по регистрационным услугам, которая является группой персонала среди всех 5 RIR, работающих в области регистрационных услуг. Эта

группа провела определенную работу над проектом набора показателей, которые мы назвали бы работоспособностью идентификаторов в пространстве номеров. Они также провели консультации с общественностью по поводу документа проекта.

Это произошло ближе к концу прошлого года и позволило нашим сообществам внести свои комментарии в этот процесс. Фактически, мы получили очень мало комментариев, поэтому сейчас мы располагаем документом, который уже почти готов к окончательному утверждению и публикации. В целом, в этом документе описана работоспособность идентификаторов в пространстве номеров в отношении записей WHOIS. Мы пришли к следующим требованиям: исчерпывающие, актуальные и правильные данные как цель.

Однако эти цели разбиты на пять конкретных измеряемых показателей, которые представляют измеряемые показатели нашей базы данных, и относятся к полноте базы данных, уникальности, соответствия нашей базы данных внешним официальным записям, эффективности соответствующих данных в отношении возможности связаться с людьми, указанными в регистратуре, а также актуальности данных. В этом документе также определены различные риски, связанные с недостижением наших целей в отношении этих

показателей, а также приведен анализ причин, сопоставляемых с подобными неудачами.

Этот документ будет опубликован уже совсем скоро. Чего у нас еще нет, так как показатели сами по себе находятся все еще в форме проекта, у нас еще нет того, о чем Алан вскоре сделает доклад, — это фактические данные о выполнении нами наших обязательств. Но очевидно, что наличие этих показателей может дать возможность измерить различные показатели, чтобы установить некоторые цели, которых мы надеемся достичь, и отслеживать выполнение с течением времени.

Поэтому я думаю, что если пространство номеров вам интересно, то наблюдайте за ним, и мы сможем предоставить своевременный отчет о работоспособности идентификаторов интернета в пространстве номеров по всем пяти RIR. По-моему, это все. Благодарю, Алан.

АЛАН ДЮРАН: Спасибо, Пол. Хотел бы воспользоваться возможностью поблагодарить вас и других членов сообщества номеров за сотрудничество с нами в этом проекте. Я думаю, что это очень интересное сотрудничество и мы многому учимся в ходе этого проекта.

ПОЛ УИЛСОН: Не возражаю. Взаимно, Алан. Спасибо.

АЛАН ДЮРАН: Мы следуем тому же процессу, что и при рассмотрении [проблемы] пространства и определения показателей, а затем переходим к фактическим измерениям, это, в принципе, нормально. Однако у вас еще нет номеров и мы ожидаем получить первый набор номеров, когда он будет готов.

А теперь давайте перейдем к пространству имен. Кристиан сделает доклад о том, как обстоят дела в этой области.

КРИСТИАН ХАЙТЕМА (CHRISTIAN HUITEMA): Добрый день. Я Кристиан Хайтема. Я работал над измерением данных и [состояния] DNS около полутора лет после исследования, призванного выяснить, что можно сделать с DNS. Я работал с Аланом и создавал фактические показатели.

Как сказал Пол, при разработке этих показателей должны использоваться определенные принципы. Эти утвержденные нами принципы в целом показаны на этом слайде. Впервые мы очень захотели сделать работу над этими показателями техническим процессом. Мы не хотим касаться политик. Целью этих показателей является описание состояния системы. В основном это не касается никаких выводов тем или иным образом.

Мы рассмотрели определение зон, которые мы хотим отслеживать, и в которых имеются потенциальные проблемы, определили показатели этих зон и способы их измерения.

Еще один принцип состоит том, что мы не хотим делать «моментальные снимки». Мы хотим иметь систему, стабильно работающую на протяжении длительного времени, представляющую, в основном, показатели, которые нам необходимы, для ежемесячной публикации новых значений, а также публикации значений за прошлые месяцы, чтобы мы могли оценить тенденции. Мы считаем, что эти тенденции практически также важны, как и фактическое текущее значение.

По этой причине мы очень много вкладываем в автоматизацию. В основном мы настраиваем измерение в различных местах и получаем непрерывную обратную связь и автоматизацию. Поэтому сайт, публикующий данные, автоматизирован, чтобы показатели предоставлялись автоматически каждый месяц и т. д.

Мы представим эти измерения во время презентации. И по той же причине, по которой мы не хотим касаться политик, мы хотим предоставлять измерения, как есть. Каждый раз, когда вы видите цифру, вы говорите «О, виджет X теперь показывает 29%. Почему?» Обычно мы отвечаем: «Мы не знаем, почему». Это значит, что у нас есть догадки, но ваши

догадки в равной степени правильны, как и наши. Поэтому мы не хотим включать эти догадки в публикуемые показатели. Эти показатели представляют собой непосредственные измерения.

Еще один принцип заключается в том, что мы очень осторожны, чтобы не допустить проблем с конфиденциальностью. Все данные, которые мы публикуем, являются статистическими по природе. Все наши инструменты имеют открытый код и все наши результаты публикуются, чтобы их можно было проанализировать.

На предыдущих заседаниях мы сделали пару докладов. В Абу-Даби мы сделали доклад по показателям ITHI, например. Мы разбили эти показатели на семь категорий. Первое, мы смотрим на точность данных WHOIS. Далее, мы смотрим на поведение корневых серверов и уровень злоупотреблений, который имеется там в определенной степени. Прошу прощения, злоупотреблений доменными именами, злоупотребление системой доменных имен. Мы смотрим на корневой трафик DNS.

Для всех показателей, перечисленных здесь, имеются источники данных. Например, в области WHOIS мы сотрудничаем с отделом соблюдения договорных обязательств ICANN. В области злоупотребления доменными именами мы работаем с платформой отчетности о случаях

злоупотребления доменами. Для измерения корневого трафика, в отношении рекурсивных серверов, регистратур IANA для параметров DNS и в области развертывания DNSSEC мы работаем со снимками корневого трафика и снимками рекурсивных сопоставителей. Также мы сотрудничаем с рекурсивными сопоставителями, чтобы [выполнять измерения], что дает нам эту статистику.

График работы, по которому мы работали в течение последнего года, по определению показателей. Что мы имеем сейчас — это доклад о первых данных. За последние два месяца мы настроили первоначальные измерения и получили данные для M1, M2, M3 и M7. Джефф Хьюстон сделает доклад о данных для M5 после этого доклада. Также за время предыдущей совместной работы мы смогли получить первоначальный набор данных для метрик M4 и M6, которые говорят об использовании клиентами DNS.

Показатель M5 мы интегрируем по мере его разработки. Мы собираемся создать канал и получить больше измерений, чтобы у нас были более обширные данные для показателей M4 и M6. Мы собираемся расширить эти данные и опубликовать их на сайте ITHI ICANN.

Первый показатель, M1. M1 — это отслеживание точности данных WHOIS. Это реализуется путем измерения точности, т.е. количество жалоб. Мы не регистрируем фактическое

количество жалоб. Мы берем количество жалоб, которые были проверены отделом соблюдения договорных обязательств ICANN. На данный момент это количество составляет немного менее 6 на миллион. Это наши первые данные, поэтому исторические данные отсутствуют. Но мы собираемся отслеживать тенденции со временем.

Все эти данные показывают, что усредненные значения не дают понимания реального положения. Если я скажу вам, что подается в среднем 6 жалоб на миллион зарегистрированных доменов [номеров], то это означает лишь то, что это «в среднем». Мы построили график, который [неразборчиво] частоту жалоб. Общее количество жалоб по оси Y, по оси X количество регистраторов, расставленных от регистратора с наибольшим количеством жалоб до регистратора с наименьшим количеством.

Из этого видно, что распределение не [неразборчиво]. Если бы у каждого регистратора было столько жалоб, то это была бы прямая линия по диагонали. Однако мы видим иное. Здесь изображена достаточно кривая линия, очень наклоненная к оси Y. Фактически, на шесть регистраторов приходится минимум 50% жалоб. Это не целое число, это на шесть регистраторов приходится немного больше 50% жалоб. На 44 регистратора приходится 90% жалоб. Всего это почти 2000

регистраторов. Поэтому мы наблюдаем очень [смещенное] распределение.

Как я сказал, это [неразборчиво] число. И это не вывод и не обоснование причины. Однако это то, что мы наблюдаем.

КАТИ ПЕТЕРСЕН: Прошу прощения, Кристиан. У нас есть вопрос онлайн.

КРИСТИАН ХАЙТЕМА: Да?

КАТИ ПЕТЕРСЕН: От Кати Кляйман, «Как вы понимаете, что жалобы WHOIS подтверждены? Как мы знаем, некоторые жалобы подаются злоумышленниками».

АЛАН ДЮРАН: Я отвечу на этот вопрос. Мы тесно сотрудничали с отделом соблюдения договорных обязательств ICANN. Мы рассматриваем не все жалобы. Мы рассматриваем только жалобы, относящиеся к точности данных. Существует множество других типов жалоб, которые мы не учитываем.

В отделе соблюдения договорных обязательств ICANN имеется процесс, где они рассматривают такие жалобы и оценивают их. Если они считают, что для жалобы имеются

достаточные основания, то они отправляют, как они это называют, первое уведомление. Если ответа нет, то они отправляют второе уведомление, затем третье и после этого, возможно вплоть до [отмены регистрации]. Это достаточно хорошо отработанный процесс, хорошо задокументированный в отделе соблюдение договорных обязательств ICANN.

Поэтому, чтобы ответить на этот вопрос еще раз, мы учитываем только те жалобы, которые относятся к точности базы данных WHOIS регистрации, и жалобы, которые были утверждены, которые прошли через этап первого уведомления.

КРИСТИАН ХАЙТЕМА: Спасибо, Ален. Это показатель M1, касающийся точности WHOIS. Набор показателей M2 относится к злоупотреблению системой доменных имен. Мы в этом отношении работаем с платформой отчетности о случаях злоупотребления доменами (DAAR). Они отслеживают четыре типа злоупотреблений: количество сайтов, используемых для фишинг-доменов, количество сайтов, использованных доменами вредоносного ПО, количество управляющих серверов ботнета и количество спам-доменов. Этот показатель определяется, как количество «взломанных» доменов на 10000 доменных имен.

Здесь мы видим глобальные усредненные значения, которые, в основном, представляют 4 или 3 порядок для первых трех типов злоупотреблений, и намного более высокое значение для спам-доменов, потому что спам достаточно распространен.

Теперь, так же, как и с показателем M1, мы видим, что эти усредненные значения не говорят о реальном положении вещей. Если посмотреть на распределение по TLD, мы видим, что, например, когда речь заходит о фишинге, то на один gTLD приходится более 50% всех фишинг-доменов. И все фишинг-домены относятся всего к 11 gTLD. Мы видим тот же тип смещения распределения для других доменов. Поэтому это определенно является показателем структуры этой проблемы.

Мы пытались сделать те же измерения для регистраторов, но мы не хотим тратить слишком много времени на данные регистраторов, потому что наши данные регистраторов необходимо оценить с помощью процесса WHOIS. Еще на эти данные налагаются все ограничения на использование данных WHOIS, такие как ограничение во времени и все такое. Мы публикуем только те данные, которые можем проверить, которым мы можем доверять, и сегодня это немного преждевременно.

Однако мы намерены представить этот перекос данных в некоего рода таблице, как эта, в которой видно, на сколько gTLD приходится минимум 50% фишинг-доменов, доменов вредоносного ПО и пр. Затем сколько доменов приходится минимум на 90% этих вариаций. Как я сказал, мы измеряем показания. Мы не интерпретируем данные и не определяем причины происходящего.

Данные M1 и M2 предоставляются отделом соблюдение договорных обязательств ICANN и платформой DAAR. Эти показатели характеризуют качество данных. Данные M3 и M4, которые мы рассмотрим позже, относятся к фактическому трафику DNS, который мы наблюдаем. M3 относится к корневому трафику. Корневой трафик мы измеряем с помощью измерений в корне L. В основном мы снимаем один показатель в день для каждого корневого сервера L. Эти показания снимаются в случайное время, поэтому они относятся к различным периодам времени, когда мы извлекаем статистические данные. Затем мы извлекаем все эти значения и обобщаем их раз в месяц. Так мы получаем эти показатели.

Здесь показан первый показатель: сколько корневых запросов получает ответ «нет такого домена»? Это достаточно большое число. В целом это почти две трети корневого трафика — запросы, не имеющие конкретного значения. В отношении

оставшихся запросов мы смотрим, сколько этих запросов может быть кэшировано сопоставителями. Опять же, мы видим, что это достаточно большая часть, почти 30%. Запросов, о которых у нас нет информации в отношении кэширования, возможно, они не были кэшированы, где-то около 6–6,5%. Этот показатель мы отслеживаем каждый месяц. Здесь показано текущее и усредненное значения и круговая диаграмма, показывающая разбивку по доменам.

Для запросов с ответом «нет такого домена», представляющих большую часть, мы попытались разбить диаграмму на составляющие. Что приводит к чему? Мы поняли, что имеем четыре компонента: зарезервированные имена, имена, которые были зарезервированы IETF, такие как .local, и имеется пять или шесть таких имен, на которые приходится около 3,4% трафика; строки с частой утечкой, такие как .home, составляющие 9,3% трафика; и частые шаблоны, которые мы видим в данных. Это не часто встречающиеся строки. Каждое имя появляется только в течение очень небольшого отрезка времени. Их много, множество различных имен, но везде шаблоны, и мы пытаемся понять, что это за шаблоны. И потом все остальное. Имеется также около 10% шаблонов, которые мы не можем непосредственно пояснить ни одним из процессов.

Для специальных имен, определенных в RFC 6761, мы видим, что большая часть использования приходится на домен .local. Это около 2,77% корневого трафика на сегодняшний день. Присутствуют и другие зарезервированные домены, но их гораздо меньше: .localhost представлен достаточно незначительно, .invalid представлен достаточно незначительно, других доменов ничтожно мало.

Что касается строк с частой утечкой, мы получаем наиболее часто встречающиеся в корне строки и в текущей вариации в текущей реализации рассматриваем только те, которые встречаются минимум 0,01% времени.

На этом слайде представлены только те строки, которые встречаются 0,02% времени, потому что чем ниже значение, тем менее мы уверены в результатах. Ну еще по причине удобочитаемости в PowerPoint.

Опять же, мы видим, что имеется одно имя .home, на которое приходится 3,5% запросов, наблюдающихся в корне. Также имеется группа других имен. Дополнительная задача заключается в том, что мы можем абсолютно измерить утечку этих имен в корне и можем отслеживать этот показатель от месяца к месяцу, и мы знаем, какие имена используются и какие имена повторяются. Мы видим, что изменения незначительны от месяца к месяцу. Некоторые имена

исчезнут, но мы видим, что имеется некое ядро часто используемых имен, встречающихся постоянно.

Я говорил, что ряд имен, наблюдаемых в этом корневом трафике, не является специальным доменом и не соответствует часто используемым строкам. Это просто случайные имена. Как видно на этой диаграмме, мы распределили эти имена по длине. Видно, что часть этих имен имеет длину от 7 до 15 символов. Более длинные имена мы здесь не представили, потому что их очень-очень мало.

Многие из этих имен длиной от 7 до 15 символов при случайном замере показателей, выглядят как нечто, случайно сгенерированное компьютерами. Но они такие не все. Очень сложно различить, что случайно генерируется компьютером, и что представляет некую нумерацию в какой-нибудь сети Wi-Fi, например. Однако это нечто, что мы хотели бы отслеживать и с чем мы хотим работать и анализировать в дальнейшем.

Это корневой трафик. Когда мы завершили в прошлом году первое исследование, мы провели несколько экспериментов и очень быстро пришли к заключению, что в корне представлен не весь трафик пользователей. Если вы понимаете архитектуру DNS, вы знаете, что то, что вы наблюдаете в корне, уже было отфильтровано DNS-сопоставителями. Если бы DNS-сопоставители использовались во всех современных технологиях, определенных IETF, в корне наблюдалось бы

чрезвычайно мало трафика. Хорошие результаты они будут кэшировать. Также будут кэшироваться [неудовлетворительные] результаты. Потому мы не увидим этого. Таким образом, большой объем трафика в корне соответствует аномальному поведению.

Если мы хотим смотреть на то, что пользователи делают фактически, нам необходимо быть как можно ближе к клиентам. Поэтому мы и расположили точки измерения на рекурсивных сопоставителях и попытались посмотреть, что там происходит. Сколько запросов, выпущенных клиентами, направляются в зарегистрированные TLD, а не во все эти строки, которые мы видим в корне? Сколько направляется на эти зарезервированные имена IETF? Сколько приходится на часто используемые строки, которые мы там видим и что еще?

Вы помните, что когда мы смотрим на корневой трафик, мы видим, что эти запросы к несуществующим TLD представляют почти 2/3 трафика. Здесь в одной точке измерений — должен сказать, что у нас только одна точка измерения на сегодняшний день. Их количество будет расти. В этой одной точке эти несуществующие TLD представляют лишь 1% трафика, т.е. намного меньше.

Тенденции также отличаются. В зарезервированных именах мы видим небольшой трафик с .localhost, .local и почти отсутствующий трафик для других имен.

В часто используемых строках это стало для нас неожиданностью. Здесь доминируют локальные имена, такие как имя хоста, которые люди пытаются найти, при этом отправляют запросы неправильно и в, конце концов, отправляют [запрос], указывая имя хоста, как единое [токен]-имя, которое может быть воспринято, как домен верхнего уровня. Мы не публикуем значения этих имен, потому что здесь имеются проблемы с конфиденциальностью. Это обычно имена в локальной инфраструктуре людей, предоставляющих нам точки для измерений, поэтому мы их просто сводим в глобальную категорию «имена локальных хостов».

Если пойти дальше, то мы увидим очень мало трафика для таких имен, наблюдаемых в корне. Мы наблюдаем трафик для больших имен, таких как .home, но мы также видим трафик для имен типа .dns, .internal, или .unifi, в этом случае представляющие сети Wi-Fi, в которых они используются. Это был для нас один из уроков. В этот момент нам хотелось бы иметь больше точек измерений перед тем, как мы сможем сделать ясные заявления, однако мы наблюдаем разницу в трафике на уровне клиентов и на уровне корня.

Вы хотели что-то сказать?

АЛАН ДЮРАН: Я хотел бы немного добавить к тому, что сказал Кристиан. На данный момент мы работаем с рядом небольших организаций и уже имеем две организации, которые согласились принять участие в этом и уже предоставляют данные. Мне хотелось бы поблагодарить их здесь.

Одна из них — это Университет Кейп-Кост в Гане, и еще Университет Ла-Плата в Аргентине. Также мы сейчас работаем с третьей организацией Nawala, являющейся, более-менее, поставщиком услуг в Индонезии. Вчера вечером мы допоздна пытались помочь им установить инструментарий для всех этих измерений.

Также мы обращаемся к другим потенциальным партнерам. Наша цель состоит в получении как можно большего количества участников проекта. Если бы у нас было пять, шесть, может, до десяти точек к концу года, было бы очень хорошо. Нам хотелось бы видеть среди партнеров различные организации из сектора науки и образования, промышленные организации, поставщиков услуг, небольшие, крупные или очень крупные организации.

Однако мы начинаем с [неразборчиво] подхода. Мы начали с малого. Это помогло нам понять, как на самом деле все работает, и мы смогли настроить инструменты, находящиеся в нашем распоряжении. Сейчас мы разрабатываем процесс, чтобы автоматизировать все это. Мы можем сотрудничать с

более крупными организациями и надеемся даже, через некоторое время, с намного более крупными.

Я хотел бы поблагодарить Кристиана за [создание] инструментария и помощь, которую он предоставляет каждому при развертывании.

КРИСТИАН ХАЙТЕМА: Спасибо. Ну, фактически, Ален тоже много времени провел, выполняя работу по развертыванию. Особенность мировой инфраструктуры состоит том, что вы проводите много времени на телефоне или в чате за компьютером посреди ночи. Однако это часть обязанностей, я бы сказал.

Итак, показатели M3 и M4 представляют анализ двух частей трафика. Какой тип DNS-трафика мы видим в корне и на стороне клиента? С помощью данных M4 нам хотелось бы также увидеть, насколько хорошими или полезными являются эти регистратуры IANA, которые мы ведем для IETF? Мы не можем отслеживать все регистратуры IANA, потому что располагаем только данными [касающимися] DNS. Мы составили таблицы [для] DNS, посмотрите на параметры, являющиеся частью регистратур. Например, типы [r] или классы [r-кодов], но также параметры, используемые DNSSEC, или параметры, используемые DANE.

В отношении этих параметров мы хотели бы ответить на два вопроса. Первый: используют ли люди зарегистрированные данные? В основном мы говорили: «Хорошо, если таблица определяет 10 значений, сколько этих значений мы фактически видим в наборе данных минимум один раз?» Для некоторых таблиц ответ — ноль. Таких таблиц очень немного.

Для классических таблиц, таких как классы DNS или номера алгоритмов, ответ приблизительно 20% и 70%. Некоторые значения используются редко. Например, некоторые алгоритмы безопасности устарели и люди их больше не используют. Но мы их видим. Это дает нам уверенность, что дело IANA не напрасно.

Мы также еще хотели увидеть, пропускают ли люди регистрацию IANA и создают собственные значения непосредственно. В этом наборе мы это наблюдаем только для кодов параметров DNS, EDNS0 DNS, если точнее, тогда как имеется небольшой уровень использования экспериментальных значений, которые мы встречаем в природе. Глобально мы видим вот такую картину.

Теперь я хотел бы сказать пару слов об использовании сертификата TLSA и сгенерированном сертификате DANE. В моих данных я их не вижу. Поэтому у меня был длинный разговор с Виктором Духовным (Victor Dukhovny) на эту тему. Он мне сказал, что это нормально, потому что по большей

части DANE используется между почтовыми и авторитативными серверами. Почтовый сервер отправляет запросы авторитативному серверу напрямую. Потому этот трафик через наши точки измерения не проходит. Я работаю с ним, чтобы получить непосредственные данные трафика, имеющиеся у него в измерениях DANE, чтобы мы могли фактически оценить правильно использование таблицы DANE.

Это в основном представляет путь, как мы можем использовать эти измерения для отслеживания IANA. Мы отслеживаем не одну таблицу. На предыдущем слайде я представил данные четырех или пяти таблиц. Здесь представлен весь список, который мы ведем, и могли бы добавить со временем больше таблиц в список, когда мы определим, как анализировать данные и извлекать их.

Финальный показатель M7 характеризует развертывание DNSSEC. Мы начали оценку развертывания DNSSEC с анализа корневой зоны, чтобы увидеть, сколько TLD предоставляли ключ DNS. Это количество достаточно стабильно и держится на уровне 90%. Но мы надеемся, что оно изменится и со временем и достигнет 100%, однако изменения протекают очень медленно.

При анализе данных M4 мы поняли, что видели большую часть трафика, которая фактически представляет трафик

безопасности DNS. Мы видим это потому, что можем заметить, что когда клиент использует безопасность DNS, он размещают бит DO в запросах, которые [неразборчиво] в ответе. Поэтому, если мы можем измерить часть запросов, содержащих этот бит, мы можем сказать, «Эй, если мы видим, что клиенты это делают, то знаем, что вот столько клиентов используют DNSSEC».

Поэтому мы можем добавить к этим данным то, что мы хотим сделать в M7.2, представляющим процент запросов DNSSEC от клиентов, использующих DNSSEC. Если бы мы были честолюбивей, мы определенно также увидели процент запросов от рекурсивных сопоставителей, использующих и, что интересно, процент ответов от авторитативных серверов, предоставляющих ответы DNSSEC. Я думаю, что таким образом мы получим показатель фактического использования DNSSEC и сможем ответить на вопрос, в каком объеме используется DNSSEC на сегодняшний день? Мне кажется, что это может представлять интерес для сообщества.

Поэтому я просмотрел 6 из наших 7 показателей. Джефф Хьюстон сделает доклад о показателе 5 после меня. M7, как я говорил, очень стабилен, поэтому этот график говорит нам совсем немного.

Хочу поблагодарить вас за внимание и ответить на любые вопросы, если, конечно, они у вас есть.

РУБЕНС КУЛ (RUBENS KUHL): Рубенс Кул, домен .br. Я хотел бы добавить комментарий о том, что рекурсивный сервер, рекурсивная DNS, рекурсивные показатели основываются на трех рекурсивных серверах. И на данный момент у нас есть 50000 [неразборчиво] систем в интернете, поэтому, возможно, нам не стоит публиковать эти результаты, пока мы не получим 5000 рекурсивных серверов DNS, так как это не имеет статистической релевантности сейчас. Это как если один [неразборчиво] под микроскоп и делать выводы о всех тканях в мире на основании этого.

[Поэтому меня удивляет], что ICANN будет публиковать такой показатель, в особенности в области, где ICANN не располагает непосредственными данными о различных авторитативных корневых данных, когда она управляет одним из наиболее обширных корневых систем корневых серверов, потому что это [неразборчиво], например. Корень имеет очень хорошую статистическую значимость для корневых запросов. Однако, что касается рекурсивных запросов, то нам вовсе не стоит их публиковать, пока они не достигнут действительно хорошего порога статистической релевантности.

КРИСТИАН ХАЙТЕМА: Это очень хорошее замечание. Мы прибегали к различным аргументам в этом докладе и пояснили, что у нас есть только одна точка измерений на сегодняшний день, и что [неразборчиво] мы хотим сделать. Совершенно ясно, что мы хотим больше, чем одну точку. Не знаю, нужно ли нам 5000. Я был бы очень рад 5000, но не знаю, нужно ли нам это на самом деле.

Я планирую сравнить данные от различных сайтов после их подписки, и посмотреть, как они отличаются и что между ними общего. Идея заключается в том, что нам известно о наличии различий. Есть различия по времени, например, утром и вечером. В рабочие дни и в выходные ситуации также различаются. Мы знаем, что есть разница в отношении географии. Люди в Китае и в Америке не запрашивают один и тот же трафик. Мы знаем, что есть различия в виде занятости. Люди не делают один и тот же запрос в образовательной, правительственной, промышленной, частной или мобильной сети. Поэтому совершенно ясно, что мы хотели бы иметь данные каждой из этих зон.

Статистическая значимость тоже является проблемой, которую мы будем решать. Это совершенно определенно одна из наших задач. Но нам нужно начать с чего-то, поэтому мы эффективно собираем данные. И мы будем сравнивать источники, чтобы суметь ответить на ваш вопрос.

РУБЕНС КУЛ: Да, но я хотел бы на это ответить. Несмотря на то, что эти аргументы озвучены, они обычно указаны мелким шрифтом. Поэтому любой, читающий отчет, повторит это и опубликует данные в прессе, в соцсетях и не упомянет об этих оговорках. Такие аргументы практически бесполезны. Потому публикация — это медвежья услуга для сообщества. Это моя точка зрения.

Один комментарий о другом моменте, который был упомянут, это некоторые запросы регистраторов повлияли на ограничения во WHOIS и т.д. Существуют данные, которые я могу собрать из всех [расширенных] регистратур, т.е. BRDA — общие данные сокращенного варианта записи данных регистратуры. Эти данные регистратуры сокращенного объема уже содержат информацию о том, какой регистратор связан с определенным доменом. Поэтому делать запрос WHOIS не нужно. Данные уже имеются внутри ICANN и представляются со стопроцентной точностью. Может, захотите заглянуть в то, к чему есть доверие.

КРИСТИАН ХАЙТЕМА: Это интересное замечание. Точнее, хорошее замечание. Сначала хотелось бы сказать, что проект ITHI — это клиент проекта DAAR. Мы получаем данные от DAAR, чтобы при любом их решении мы об этом знали. Потому, во-первых,

хотел бы предложить переадресовать ваш вопрос людям из DAAR.

Второе, [и я, наверно, попытаюсь] их как-то переадресовать. Как я понимаю, нужно было исследование, обеспечивающее воспроизводимость. Это означает, что кто-то за пределами ICANN может повторить в точности то же самое исследование, открыть методологию и данные, находящиеся в доступе. Эти данные, которые вы упоминали, могут или могут не быть доступными за пределами, и это поставит ICANN в уникальное положение единственной организации, могущей провести это исследование. И там решили не идти этим путем. Они могут изменить подход в определенный момент и, возможно, Джон Крейн (John Crain) может ответить на это, но сейчас мы движемся именно в этом направлении. И как их клиент, мы принимаем то же самое решение.

Поэтому вопрос состоит в том, почему мы должны полагаться на WHOIS для определения регистраторов в противоположность использованию внутренних данных ICANN?

ДЖОН КРЕЙН (JOHN CRAIN): Если у нас будут все данные, доступные внутри организации, то я бы их не нашел. Это было бы просто замечательно. Но одна из вещей, которые мы пытаемся сделать, — это обеспечивать повторяемость для других

людей, что означает использование внешних источников. Единственное, что нам нужно от WHOIS — это ID регистратора. Мы фактически говорили раньше о том, где могут иметься источники внутри организации, поэтому можем переключиться на них, потому что, как я считаю, WHOIS может быть неудобен.

РУБЕНС КУЛ: Источник называется BRDA, поэтому вы, может, захотите взглянуть на них, хакнуть эти серверы и получить из них данные. Но даже если вы и будете использовать эти данные, это все равно обеспечит возможность повторить, просто немного усложнит для других людей использование запросов WHOIS. Однако не могут воспроизвести все с помощью запросов WHOIS, потому что это та же самая информация. Это полностью доступная информация.

ДЖОН КРЕЙН: Да понятно. Когда мы начинали проект, мы хотели все сделать точно так, как это сделали бы посторонние люди. Вот и вы сейчас говорите о воспроизводимости, поэтому мы думаем об этом.

РУБЕНС КУЛ: Хорошо.

АЛАН ДЮРАН: Хорошо. Есть вопросы в чате, Кати?

КАТИ ПЕТЕРСЕН: Нет.

АЛАН ДЮРАН: Нет вопросов? Хорошо. Хорошо, спасибо, Кристиан.

КРИСТИАН ХАЙТЕМА: Спасибо.

АЛАН ДЮРАН: Спасибо, что показали эти цифры впервые. Теперь я хотел бы пригласить Джеффа Хьюстона. Джефф в зале?

КАТИ ПЕТЕРСЕН: Тут есть вопрос.

АЛАН ДЮРАН: У нас есть вопрос.

СЕБАСТЬЕН БАШОЛЕ (SEBASTIEN BACHOLLET): Так как Джеффа в зале нет, я хотел бы задать один вопрос от человека, который не обладает техническими познаниями. Имеется ли какая-нибудь связь

между тем, что вы делаете, и какими-нибудь вопросами об обновлении ключей и данных, необходимых для понимания того, что происходит? Прошу прощения, это [неразборчиво] вопрос от [неразборчиво]. Спасибо.

КРИСТИАН ХАЙТЕМА: По состоянию на сегодня ответ «нет». Связи никакой нет. Это не показатель, который мы рассматривали изначально. На данный момент, как и в будущем, может быть больше обновлений, они могут быть более-менее регулярными и это может быть нечто, что мы хотели бы отслеживать. Поэтому сегодня мы говорили о семи показателях. Мы считаем, что понимаем их достаточно, чтобы быть способными их отслеживать. Мы обдумываем сейчас второй этап, на котором добавим больше показателей, чтобы рассмотреть другие типы проблем. Это может быть одна из областей, в которые мы хотим заглянуть и добавить к тому, что мы делаем сейчас. [отвечает на французском]

СЕБАСТЬЕН БАШОЛЕ: Я понимаю, о чем вы говорите. Просто хочу подробно пояснить свой вопрос, прошу прощения. Складывается впечатление, что сегодня нам не хватает данных для уверенности в том, что это правильное время для обновления ключей. Но не факт, что, когда мы сделаем обновление ключей, то каждый год вы сможете собирать данные [как

информацию для размышлений], которые можно использовать людям, которым необходимо решить, когда делать обновление.

КРИСТИАН ХАЙТЕМА: На сегодняшний день у нас нет данных, которые были бы им полезны.

СЕБАСТЬЕН БАШОЛЕ: Спасибо.

ПАТ КЕЙН (PAT KANE): Здравствуйте. Пат Кейн, компания VeriSign. Просто дополнение к вопросу Себастьяна. Ранее сегодня представитель CTO ICANN говорил о снижении числа запросов DNSSEC на обновление KSK с прошлой осени до следующего октября. Поэтому, я думаю, важно, чтобы мы понимали из использования DNSSEC, какое это все имеет отношение к информированию об этом решении, потому что количество многих данных у людей с сопоставителями, у которых нет обоих ключей, ухудшается по сравнению с концом прошлого года. Было бы очень хорошо предоставить эту информацию Дэвиду как можно раньше. Спасибо.

АЛАН ДЮРАН: Спасибо. Это очень хорошее мнение. Как мы видели, Кристиан говорил о новом показателе M7.2, отслеживающем количество запросов, содержащих бит DO. Это может помочь в этом отношении с другими показателями, которые мы пытаемся создать в этом конкретном пространстве. Может, мы могли бы поговорить дополнительно, если у вас есть конкретные идеи, как это можно отслеживать.

Джефф вернулся? Хорошо, я прошу прощения. Один из наших докладчиков потерялся. Могу кратко рассказать о том, что мы планируем сделать.

Показатель M5 изначально был одним из показателей, отслеживающим сопоставители, но больше с точки зрения клиента. Мы просили Джеффа рассмотреть этот вопрос. У Джеффа есть система измерения, основанная на известном Google Ads. Мы ее уже использовали в других контекстах. Мы просили его использовать эту систему, чтобы исследовать, что можно сделать с точки зрения клиента, stub-сопоставителей.

Одна из вещей, которые мы хотели бы рассмотреть — это действительно ли сопоставители кэшируют данные? Иногда нам кажется, что да. Иногда кажется, что они не кэшируют или могут кэшировать в течение более короткого времени или могут кэшировать более продолжительное время. Мы

считаем, что можем получить некоторые измерения в этом отношении.

Мы также можем посмотреть на некоторые DNSSEC и распределения IPv6, чтобы понять, настроены ли сопоставители на DNSSEC или нет и могут ли они использовать IPv6 или нет. Потенциально мы можем найти наиболее часто используемые сопоставители. Эти сопоставители я могу определить визуально, потому что система основывается на данных Google Ads, используемой живыми пользователями, а не машинами. Потому мы получим данные не о связи «машина-машина», а «пользователь-машина».

Это может быть проект измерений [неразборчиво], также могущий предоставить нам информацию об обновлении ключей и сколько сопоставителей на самом деле необходимо, чтобы покрыть 95% или какой бы то ни было процент населения, который нам нужен.

Работа идет. Это новые показатели, которые хотел предложить Джефф. В том же духе, что и Кристиан описывал ранее, мы хотим настроить автоматизацию, чтобы мы могли собирать измерения и отслеживать все это во времени в течение нескольких лет.

Поэтому в целом, это проект, который мы хотели бы сделать с Джеффом.

Я прошу прощения за то, что он отсутствует. Возможно, были какие-то обстоятельства.

Если вопросов больше нет, то мы закроем заседание пораньше. Ага, вопрос.

РУБЕНС КУЛ: В принципе, это больше ответ на комментарий Пата Кейна. Почему ряд сопоставителей, предоставляющих информацию о DNSSEC, увеличил количество сопоставителей, предоставляющих информацию [неразборчиво] 2010 KSK? Мы еще не знаем, являются ли эти сопоставители проверяющими или нет. Потому это может быть кто-то, у кого действительно есть корневой ключ, но он не является проверяющим. Это не является вероятной проблемой при обновлении ключей. Если мы проведем какое-нибудь исследование, например, как с показателями, то, может, нам стоит взглянуть на проверяющие распознаватели со старыми ключами, а не только на распознаватели, предоставляющие информацию о старых ключах. Это не инструмент измерить все, что может спрогнозировать, что произойдет, когда мы обновим корневой ключ.

АЛАН ДЮРАН: Очень хорошее замечание, но я добавлю. Мы должны каким-то образом вместе взвесить это по количеству пользователей,

стоящих за этим сопоставителем. Если это нечто, используемый очень редко, который включается на 5 минут, то это может не иметь той же важности, как сопоставитель, обслуживающий миллионы клиентов.

РУБЕНС КУЛ: Принято.

АЛАН ДЮРАН: Итак, если других комментариев нет, то мы закрываем заседание. Следующая конференция ICANN будет посвящена политикам, поэтому никаких технических заседаний больше не будет и мы там не встретимся. Но мы увидимся со всеми в Барселоне.

[КОНЕЦ СТЕНОГРАММЫ]