№3
апрель 2016

MSK IX

Интернет изнутри

Связность

В Интернете каждый соединен с каждым - верно?

Является ли Интернет единым согласованным доменом связности, где каждый может соединиться с каждым?

Содержание:

Иформационный сборник «Интернет изнутри»

По всем вопросам пишите на info@internetinside.ru

Порядковый номер выпуска и дата его выхода в свет:
Выпуск №3, дата выхода: апрель 2016 г.

Публикуется при поддержке АНО «ЦВКС «МСК-IX»

Главный редактор:
Андрей Робачевский

Зам. главного редактора:
Новикова Татьяна

Дизайн:
Чернега Наталья

Связность и взаимосвязанность

главный редактор,
Андрей Робачевский

Дорогой читатель!

Перед вами очередной номер, в котором мы решили взглянуть на вопрос связности в Интернете.

Связность мы воспринимаем как данное. За исключением временных неполадок, в принципе каждый может связаться с каждым. Но каким образом эта связность обеспечивается в системе независимых сетей с посегментной маршрутизацией и коммутацией пакетов? Как отметил Джефф Хьюстон в своей статье «В Интернете каждый соединен с каждым – верно?», «удивительно, каким образом связность Интернета остается такой стабильной и всеобъемлющей, учитывая тот факт, что этот результат зависит от потребностей рынка без каких-либо конкретных гарантий правильного исхода».

Связность обеспечивается на всех уровнях, начиная с физических каналов уровня коммутации IP-пакетов и заканчивая уровнем приложений. Однако говоря о качестве связи, мы имеем в виду не только собственно связность, но и производительность – эффективную пропускную способность виртуального канала между отправителем и получателем. Фундаментальную роль здесь играют транспортные протоколы, и в особенности протокол TCP. Об эволюции этого протокола, позволившей сохранить актуальность на протяжении более четырех декад развития Интернета, рассказывает статья «Транспортные протоколы».

Наконец, связность также означает взаимосвязанность и взаимозависимость. И в такой системе как Интернет по сути независимые сети – коммерческие, научно-образовательные и государственные предприятия – вынуждены сотрудничать и координировать определенную деятельность для решения насущных глобальных проблем Сети. Одной из таких проблем является проблема безопасности глобальной системы маршрутизации, когда ошибка или намеренные мошеннические действия одной из сетей может иметь серьезные последствия для многих других.

Компания Dyn располагает разветвленной сетью «наблюдательных точек», позволяющих отслеживать и анализировать подозрительные изменения в системе маршрутизации. Даг Мадори, сотрудник Dyn, в своей статье «Украина превращается в источник ложной маршрутизации» расследует новые способы маскировки своей деятельности, к которым прибегают мошенники.

Ну и, конечно, нельзя не упомянуть ставшие традиционными разделы «Путевые заметки» Ольги Александровой-Мясиной и «Ученые шутят», в котором Леонид Тодоров расскажет сказку о мудром Кавусе и троллях.

Итак, перед вами третий выпуск. Надеемся, что он не разочарует. Расскажите нам, что вам понравилось, а что – нет, о чем бы вы хотели прочитать в следующих номерах. Как всегда, ждем ваших отзывов и предложений по адресу info@internetinside.ru.

В Интернете каждый соединен с каждым - верно?

Джефф Хьюстон (Geoff Huston)

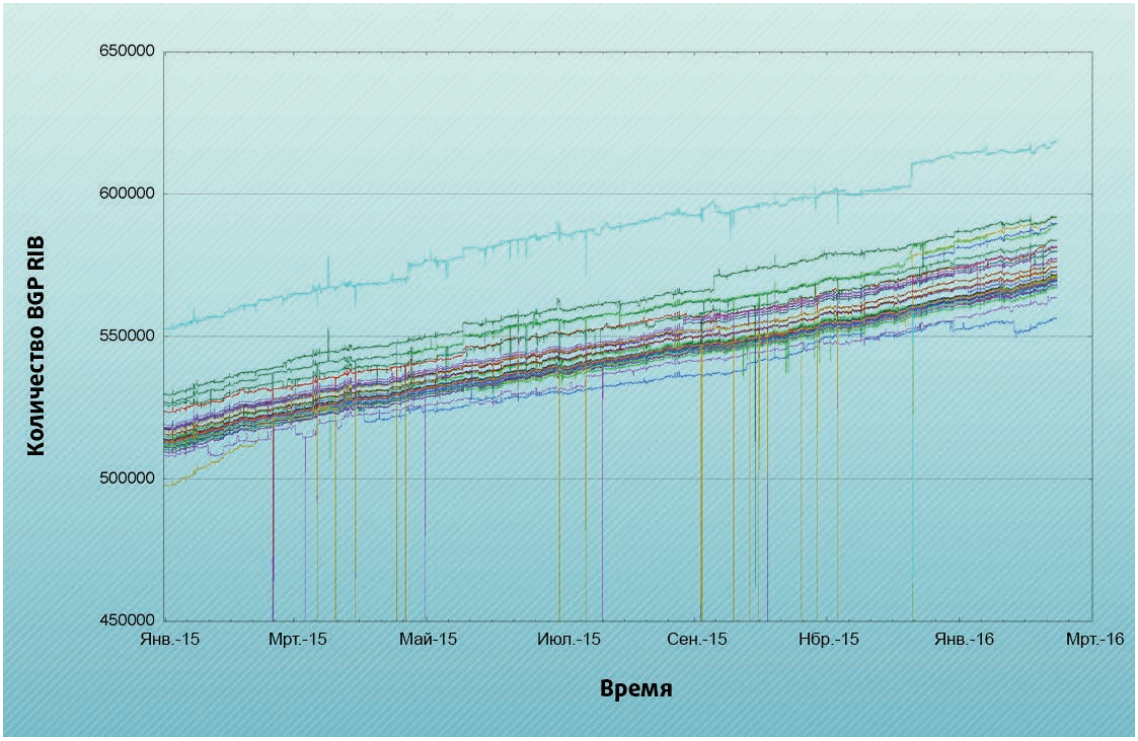
Является ли Интернет единым согласованным доменом связности, где каждый может соединиться с каждым? И каким удивительным образом связность Интернета остается такой стабильной и всеобъемлющей, учитывая тот факт, что этот результат зависит от потребностей рынка без каких-либо конкретных гарантий правильного исхода. В этой статье Джефф Хьюстон исследует глобальную систему маршрутизации Интернета с различных наблюдательных точек и обнаруживает адресное пространство, недостижимое для многих. Джефф предлагает несколько объяснений этому явлению, раскрывая внутреннюю механику отношений между операторами.

Ниже приведен график (рис. 1), над которым я уже некоторое время размышляю, иллюстрирующий тему, которую я бы хотел рассмотреть в данной статье.

Существует целый ряд «маршрутизационных коллекторов», которые собирают проекции междоменной маршрутизации с нескольких различных точек наблюдения, расположенных по всему Интернету. Рисунок 1 показывает график количества записей маршрутов, видимых в рамках ежедневного моментального снимка таблицы маршрутизации с каждого пира одного из серверов проекта Route Views. Каждая отдельная линия на этом графике представляет собой число маршрутов, анонсированных серверу Route Views каждым из его пиров.

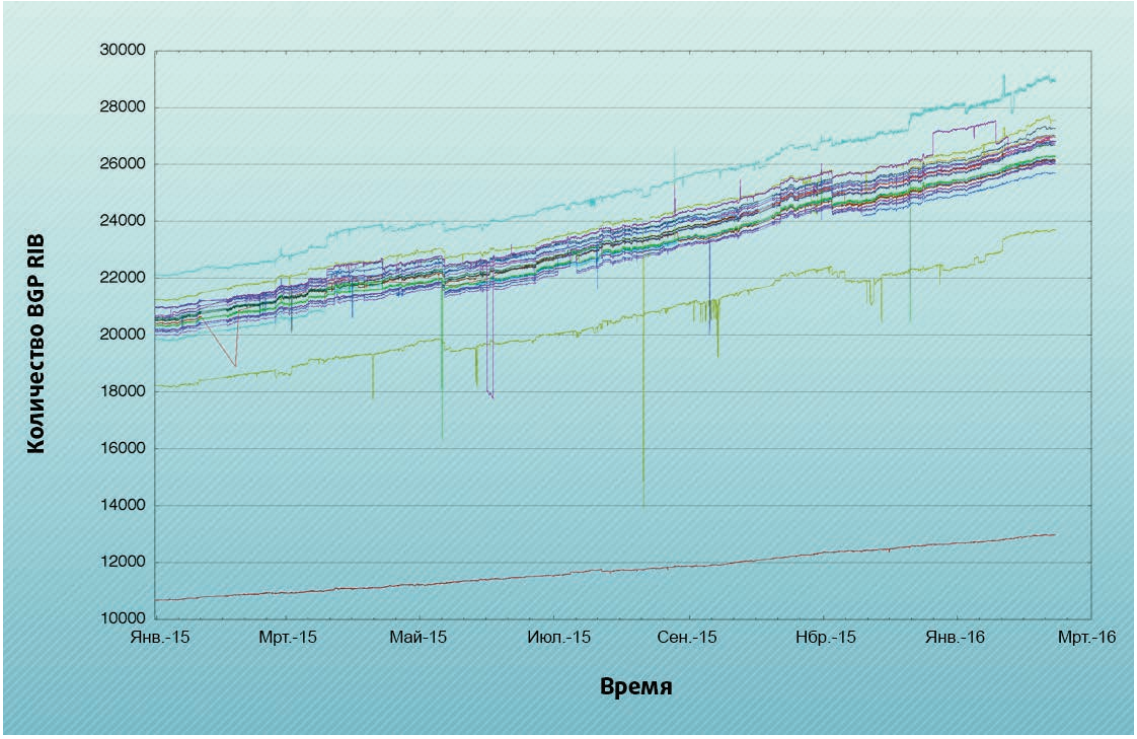
Существует ряд серверов, которые собирают анонсы BGP в различных точках наблюдения и публикуют все множество этих отдельных потоков информации о маршрутизации. В данной статье я использую данные, полученные в рамках проекта Route Views (http://www.routeviews.org), поскольку собранные данные имеют удобный размер и легко перемещаются скриптами. RIPE NCC также обслуживает набор коллекторов маршрутов в рамках услуги Routing Information Service (RIS), которая содержит большое собрание моментальных снимков таблиц маршрутизациии обновлений BGP (https://www.ripe.net/analyse/internet-measurements/routing-information-service-ris).

Рис.1. Количество маршрутов, объявленных каждым пиром (одноранговым узлом) BGP согласно проекту Route Views за 2015 г.



Этот график показывает, что вид на систему маршрутизации Интернета немного различается для каждого пира Route Views. Несомненно, если бы Интернет представлял собой единый, согласованный домен связности и если бы каждый мог единообразно «дотянуться» до любой другой точки, то не ожидали ли бы мы в таком случае, что проекция системы маршрутизации, системы, которая в конечном счете гарантирует, что каждый соединен с каждым, была везде одинаковой? И если это было бы правдой, и проекция системы маршрутизации обеспечивала бы одинаковый вид Интернета,

Рис.2. Количество маршрутов IPv6, объявленных каждым пиром BGP согласно проекту Route Views за 2015 г.



независимо от конкретной точки наблюдения, то вместо нарисованных на рис. 1 примерно 40 разных линий за определенный год – каждая для отдельной точки наблюдения за маршрутизацией – мы должны были бы увидеть одну и ту же линию, повторенную более 40 раз.

Однако данный график демонстрирует, что каждая из этих точек наблюдения видит немного разный Интернет. И эти различия не являются временными. На протяжении всего 2015 года эти разные точки наблюдения устойчиво видят разное число маршрутов в своей локальной системе маршрутизации. Поэтому данный феномен не является чем-то, что произошло в конкретный момент времени и что будет откорректировано в ходе обычной работы протокола маршрутизации. Эти различия устойчиво наблюдаются в течение всего года. Некоторые объявления маршрутов просто не видны для части точек наблюдения. Одна из причин, почему мы используем эти «коллекторы» маршрутизации, заключается в сборе разных ракурсов и в понимании полученных различий. Поэтому я бы хотел более подробно взглянуть на эти различия и увидеть, можно ли сделать какие-либо выводы о связности Интернета.

Возможно, что эти различия возникли как некий результат размера и возраста, и что «почтенный» Интернет, работающий по протоколу IPv4, получил эти расходящиеся ракурсы маршрутизации исключительно вследствие своего размера и возраста. Поэтому более «молодая» и меньшая по размерам система маршрутизации может по этой причине предоставить единый ракурс. Второй рассматриваемый график (рис. 2) относится к Интернету IPv6 за 2015 год. Вместо примерно 600 тысяч маршрутов он содержит всего 20 тысяч. Однако мы по-прежнему наблюдаем, что некоторые сети видят больше маршрутов, чем другие, и эта картина остается устойчивой и стабильной в течение всего года.

Что означают эти различия? Возможно, мы видим доказательство фрагментированности Интернета, когда из некоторых точек Интернета нельзя дотянуться до других мест? Являются ли эти различия в ракурсах для разных пунктов наблюдения

Рис.3. Доля анонсов с более специфическими маршрутами с точки зрения AS131072.

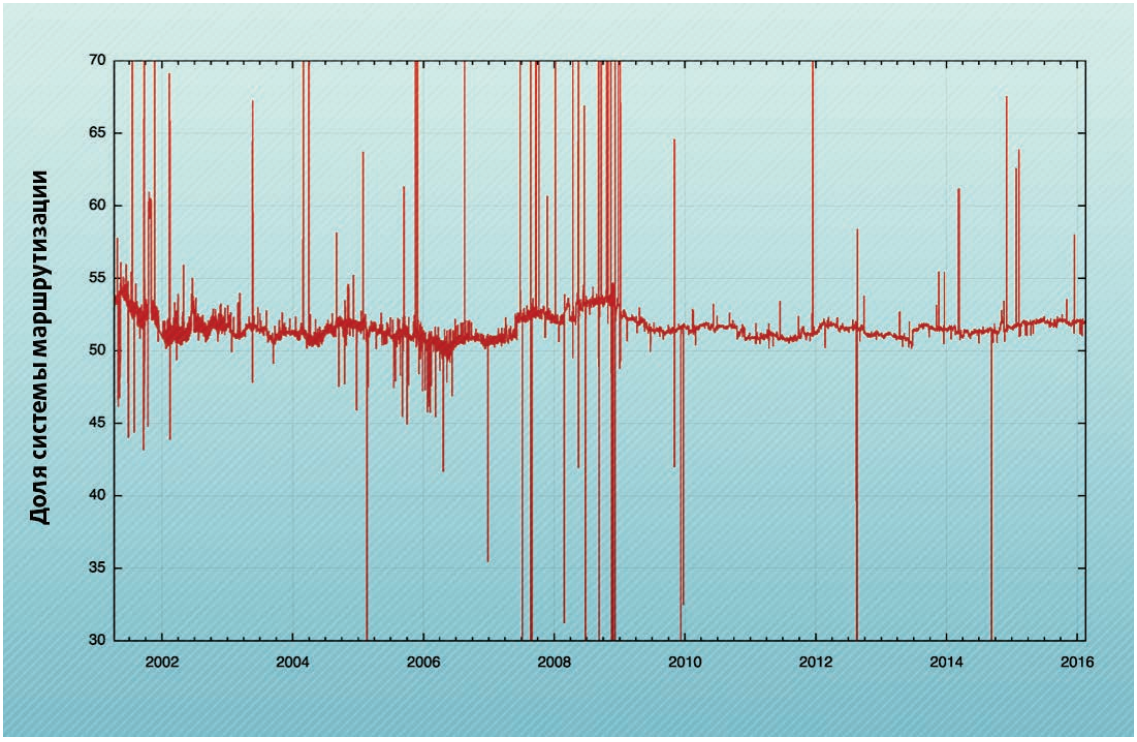
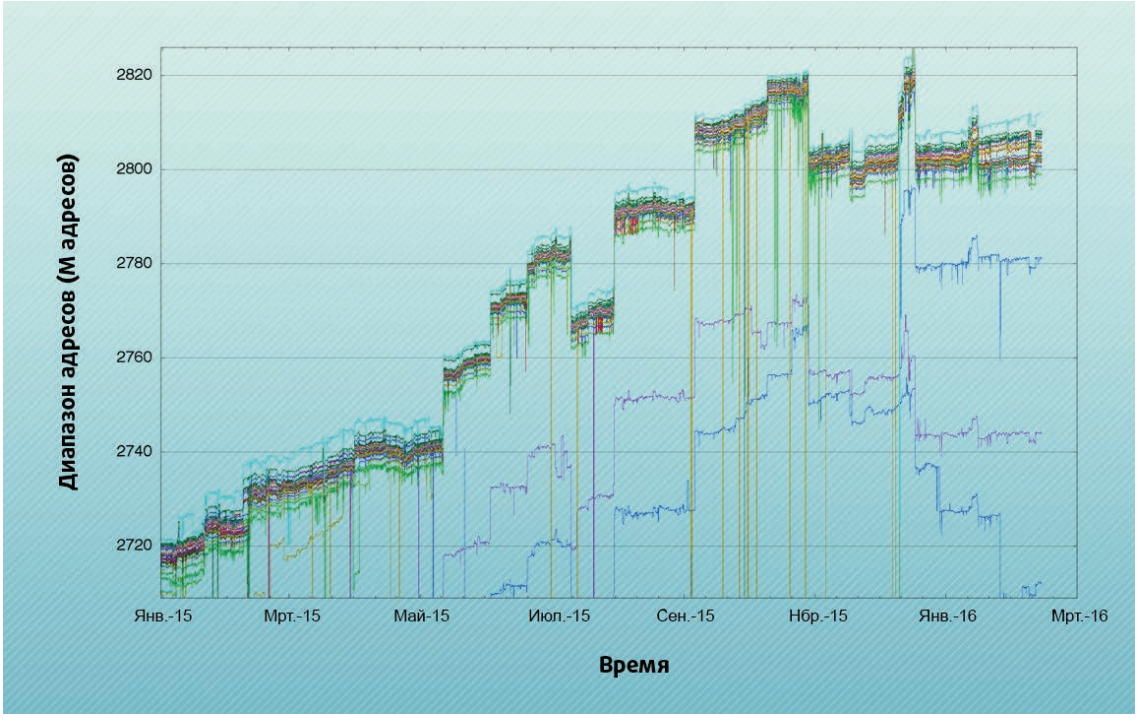


Рис.4. Суммарный диапазон достижимых адресов IPv4, анонсированных каждым пиром BGP проекта Route Views за 2015 г.



за маршрутизацией признаками разрывов в ткани связности Интернета?

Перед тем как переходить к выводам, полезно собрать вместе дополнительные данные. Одно из возможных объяснений различия в количестве объявленных маршрутов заключается в том, что система маршрутизации содержит два компонента информации: базовая достижимость и информация, относящаяся к политике, или правилам, определяющим, каким образом можно добраться до места назначения.

В общих чертах последнюю часть можно назвать набором анонсов для «инжиниринга трафика». Они не меняют общий набор адресов, которые анонсированы как достижимые в системе маршрутизации, однако предоставляют уточнения относительно того, каким образом следует «добираться» до конкретного адреса. Одна половина из 600 тысяч входных данных в системе маршрутизации IPv4 анонсирует «достижимость» как агрегированные объявления или «корневые» префиксы маршрутизации. Вторая половина добавляет к некоторым из этих базовых данных достижимости оговорки, уточняющие достижимость за счет предложения немного отличающихся путей (или нет!) к месту назначения. Поэтому возможно, что различия, видимые из каждой точки наблюдения, фактически не являются различиями в базовой связности, а представляют собой различия в более специфических анонсах, и они демонстрируют, что в разных частях сети могут существовать разные предпочтительные пути, по которым добираются до определенных мест назначения. Смотреть Рис. 3.

Различие в количестве маршрутов, видимое с каждой точки наблюдения, может быть объяснено тем фактом, что усилия по установлению скрупулезного контроля над определенными путями, которые использует междоменный трафик за счет объявления более специфических маршрутов, до некоторой степени намеренно локализованы. Распространение этих префиксов управления трафиком аналогичным образом умышленно ограничено конкретной областью или регионом. Поэтому в качестве вывода можно было бы заключить, что разброс в количестве объектов маршрутизации, видимых в каждой точке наблюдения обоих сетей – IPv4 и IPv6, – не

Рис.5. Суммарный диапазон достижимых адресов IPv6, анонсированных каждым пиром BGP проекта Route Views за 2015 г.

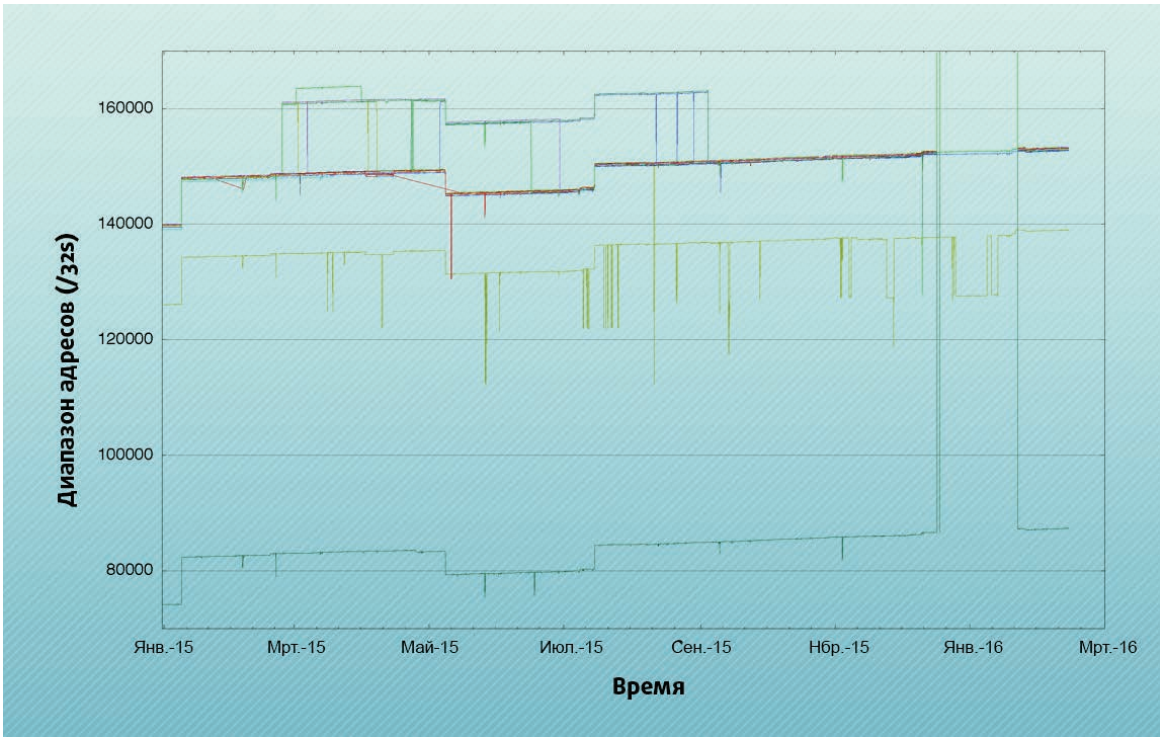
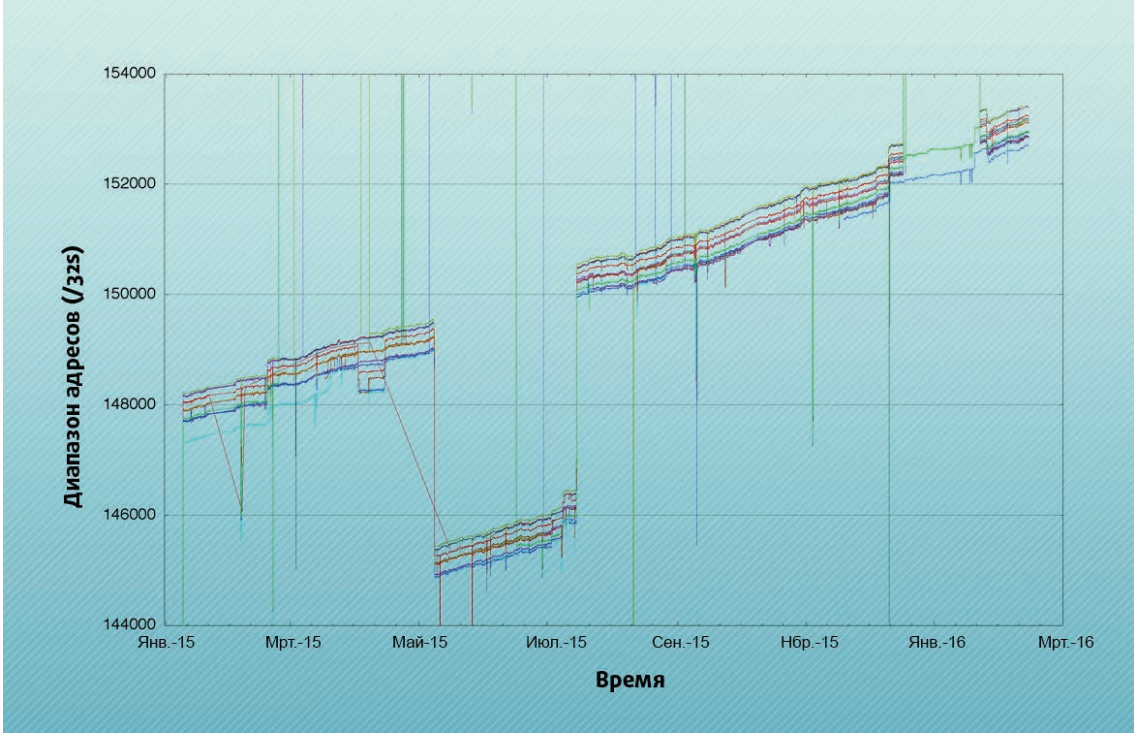


Рис.6. Детальный взгляд на суммарный интервал достижимых адресов IPv6, объявленных каждым пиром BGP в рамках Route Views за 2015 г.



в рамках всего года: те пиры, которые «видят» более крупный интервал адресов по сравнению с другими пирами, последовательно испытывают это на протяжении всего года. Если пренебречь тремя исключениями, то те пиры, которые объявляют меньший интервал адресов, делают это устойчиво в течение всего периода. Одно из исключений указывает на тенденцию увеличения диапазона достижимости, приближая его к диапазону достижимых адресов группы, тогда как два других пира анонсируют диапазон, который сокращается в течение года.

происходит вследствие какой-либо фундаментальной разницы в достижимости любых адресов, а просто объясняется желанием установить дополнительный контроль над путями перемещения трафика.

Один из способов протестировать обоснованность такого вывода заключается в том, чтобы еще раз взглянуть на данные коллекторов маршрутов, но в этот раз вместо того, чтобы обращать внимание на количество маршрутов, которые видны в каждой точке наблюдения, следует посмотреть на общий диапазон адресов, которые анонсируются как достижимые. Если эта теория является верной, то общий диапазон адресов должен быть одинаковым для каждой точки наблюдения, и мы можем вполне оправданно поверить, что Интернет действительно представляет собой однородную область связности.

И опять, данные не согласуются с этой теорией. При рассмотрении данных (рис. 4) можно заметить, что для IPv4 существует диапазон из примерно пяти миллионов адресов, при этом некоторые точки наблюдения «видят» более значительный набор адресов, анонсированных как достижимые, по сравнению с другими точками наблюдения. И опять прослеживается четкое постоянство

На графике для IPv6 можно заметить ту же ситуацию, при которой разные точки наблюдения «видят» разный диапазон адресов. Для IPv6 существуют крупномасштабные различия, например, один из пиров маршрутизации «видит» только половину суммарного интервала адресов пира, который анонсировал самый большой диапазон (рис. 5). Различие заключается в том, что этот пир не анонсирует маршрут /16, который анонсируют все другие пиры. Этому случаю легко найти объяснение: существует только один анонс /16 в «мире» IPv6, а именно – 2002::/16, анонсирование туннельного входа для технологии туннельного перехода 6to4. Возможно, что данная сеть не использует шлюз 6to4, или, что более вероятно, в сети расположен локальный

Рис.7. «Некворумные» анонсы адресов IPv4 для каждого пира за 2015 г.

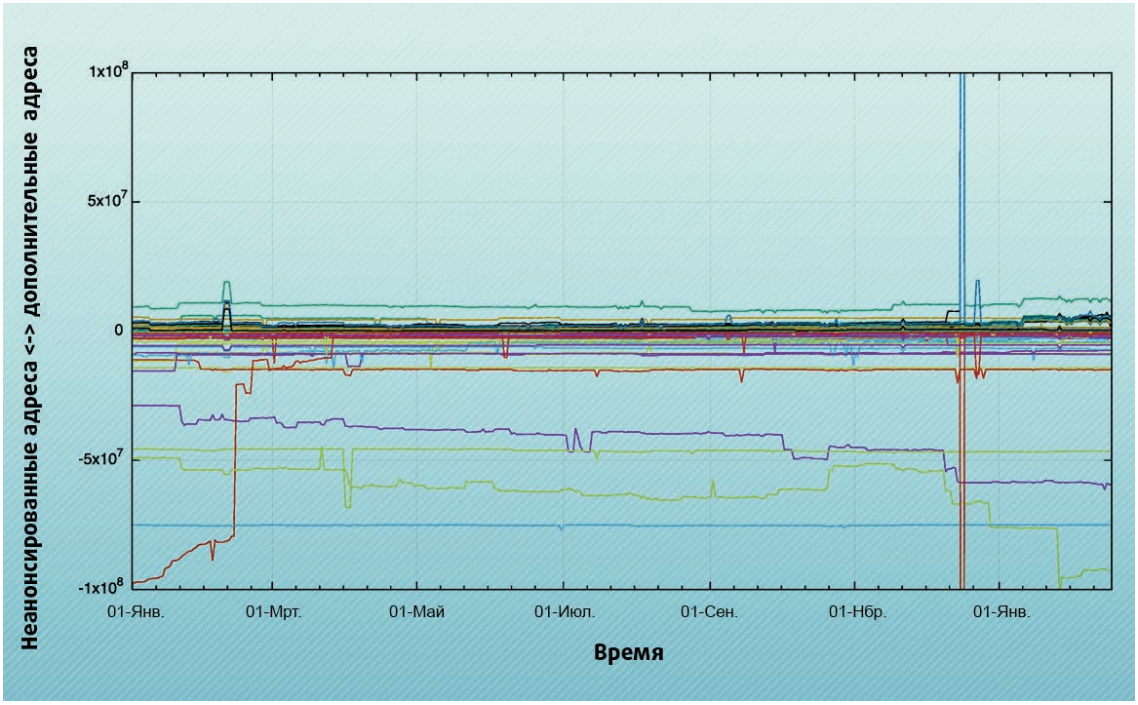


Таблица 1. Вид от Route Views на анонсированные таблицы маршрутизации IPv4 каждого пира, 17 февраля 2016 г.

AS	# Маршруты	Кворум (/8)	Пропущенные /32	Дополнительные /32	Имя AS
AS286	570,258	166.89	279,808	416,656	KPN KPN B.V., NL
AS293	581,940	166.79	2,002,432	4,471,122	ESNET - ESnet, US
AS852	571,209	166.59	5,376,256	325,632	TELUS Comms, CA
AS1221	571,660	166.06	14,228,736	199,552	Telstra, AU
AS1239	568,202	166.02	14,801,192	0	SPRINTLINK - Sprint, US
AS1299	563,988	163.38	59,226,624	66,048	TELIANET TeliaSonera AB, SE
AS1668	569,862	166.61	4,953,856	39,168	AOL Transit Data Network, US
AS2152	571,882	166.74	2,842,112	1,123,584	CSUNET-NW, US
AS2497	577,689	166.39	8,639,232	4,958,976	Internet Initiative Japan, JP
AS2914	569,261	166.46	7,479,608	4,992	NTT America. Inc., US
AS3130	570,348	166.89	290,560	9,600	RG-BIWA - RGnet. LLC, US
AS3257	569,221	166.83	1,264,608	54,016	TINET-BACKBONE, DE
AS3277	581,618	166.83	1,194,240	5,018,144	RUSNET-AS NPO RUSnet, RU
AS3303	571,302	166.66	4,088,064	310,784	SWISSCOM, CH
AS3356	566,679	164.14	46,459,904	1,792	Level 3 Communications, US
AS3549	569,244	166.77	2,337,216	72,448	Level 3 Communications, US
AS3561	569,285	166.37	8,959,232	1,536	Savvis, US
AS3741	571,241	166.62	4,758,528	304,640	IS, ZA
AS5413	570,940	166.85	852,480	182,816	Daisy Communications Ltd, GB
AS6539	574,633	166.88	438,784	100,864	GT-BELL - Bell Canada, CA
AS6762	571,435	166.88	478,464	12,032	SEABONE-NET, IT
AS6939	580,094	166.77	2,348,800	3,721,728	Hurricane Electric, US
AS7018	567,633	162.44	74,979,328	45,056	AT&T Services, US
AS7660	556,519	161.39	92,543,856	1,421,960	AsiaPac Advanced Network, JP
AS8492	581,296	166.84	1,122,048	4,744,960	OBIT-AS OBIT Ltd., RU
AS11686	573,309	166.76	2,376,952	1,081,856	Education Networks, US
AS13030	569,627	166.84	1,078,016	2,768,397	INIT7, CH
AS20771	589,700	166.79	1,871,360	4,280,576	Caucasus Online, GE
AS20912	576,857	166.77	2,221,822	2,743,621	Panservice, IT
AS22652	574,852	166.89	241,664	342,272	Fibrenoire, CA
AS23673	584,638	166.48	7,168,256	11,248,384	Cogetel Online, KH
AS37100	574,864	166.66	4,088,576	4,175,104	SEACOM-AS, MU
AS40191	571,528	166.90	40,448	163,200	AS-PRE2POST-1, CA
AS47872	592,253	166.82	1,391,360	4,277,664	SOFIA CONNECT EOOD, BG
AS53364	570,002	166.87	587,264	54,016	AS-PRE2POST-2, US
AS58511	589,579	166.82	1,500,822	6,139,973	Connectivity IT, AU
AS62567	570,671	166.90	104,448	111,104	ASN-NY2 - Digital Ocean, US
AS200130	571,269	166.86	772,096	2,549,760	ASN-1 Digital Ocean, EU
AS202018	570,636	166.90	72,704	103,936	ASN-3 Digital Ocean, NL
AS393406	570,588	166.89	187,648	116,736	NY3 - Digital Ocean, US

Полный анализ всех пиров из моментального снимка BGP проекта Route Views, зафиксированного 17 февраля 2016 года, показан в таблице 1. На основе этой таблицы очевидно, что никто не видит тот же самый набор адресов, что и любой другой участник, при этом разброс может быть достаточно велик.

Пиры IPv4 в рамках Route Views

Пиры: 40

Кворум: 25

Анонсы: 618,124

Кворум-анонсы: 570,372

Диапазон: 167.48 (/8s)

Диапазон кворума: 166.90 (/8s)

Таблица 2. Вид от Route Views на анонсированные таблицы маршрутизации IPv6 каждого пира, 18 февраля 2016 г.

AS	# Маршруты	Кворум (/8)	Пропущенные /32	Дополнительные /32	Имя AS
AS33437	26,013	87,194	275	120	HOTNIC - Hotnic LLC, US
AS2914	26,144	87,296	174	2	NTT America, Inc., US
AS47872	27,037	87,466	4	123	SOFIA-CONNECT, BG
AS3277	27,699	87,467	3	170	RUSNET-AS NPO RUSnet, RU
AS1239	25,694	87,155	314	1	SPRINTLINK - Sprint, US
AS30071	23,696	73,520	13,949	4	OCCAID, US
AS20912	26,666	87,466	4	126	Panservice, IT
AS37100	26,199	87,412	58	6	SEACOM-AS, MU
AS31019	26,730	87,459	11	160	MEANIE, NL
AS701	25,989	87,025	445	1	Verizon Business, US
AS200130	26,609	87,467	3	129	Digital Ocean, EU
AS7018	26,127	87,019	451	3	AT&T Services, US
AS202018	26,608	87,467	3	121	Digital Ocean,NL
AS393406	26,608	87,467	3	121	Digital Ocean, US
AS3257	26,006	87,299	171	1	Tinet, DE
AS62567	26,608	87,467	3	121	Digital Ocean,US
AS53364	26,428	87,306	164	1	AS-PRE2POST-2, US
AS22652	27,377	87,412	58	2	FIBRENOIRE-INTERNET, CA
AS40191	27,562	87,466	4	122	AS-PRE2POST-1, CA
AS6939	25,979	86,932	537	120	Hurricane Electric, US
AS3741	26,885	87,434	36	123	IS, ZA
AS2497	26,251	87,463	7	121	Internet initiative Japan, JP
AS13030	26,876	87,394	76	27	Init7, CH
AS209	26,653	86,949	520	120	CENTURYLINK, US

Пиры IPv6 в рамках Route Views

(Интервалы адресов показаны в единицах /32)

Пиры: 24
Кворум: 15
Анонсы: 28,935
Кворум-анонсы: 26,232
Диапазон: 87,688
Диапазон кворума: 87,468
Диапазон кворума: 166.90 (/8s)

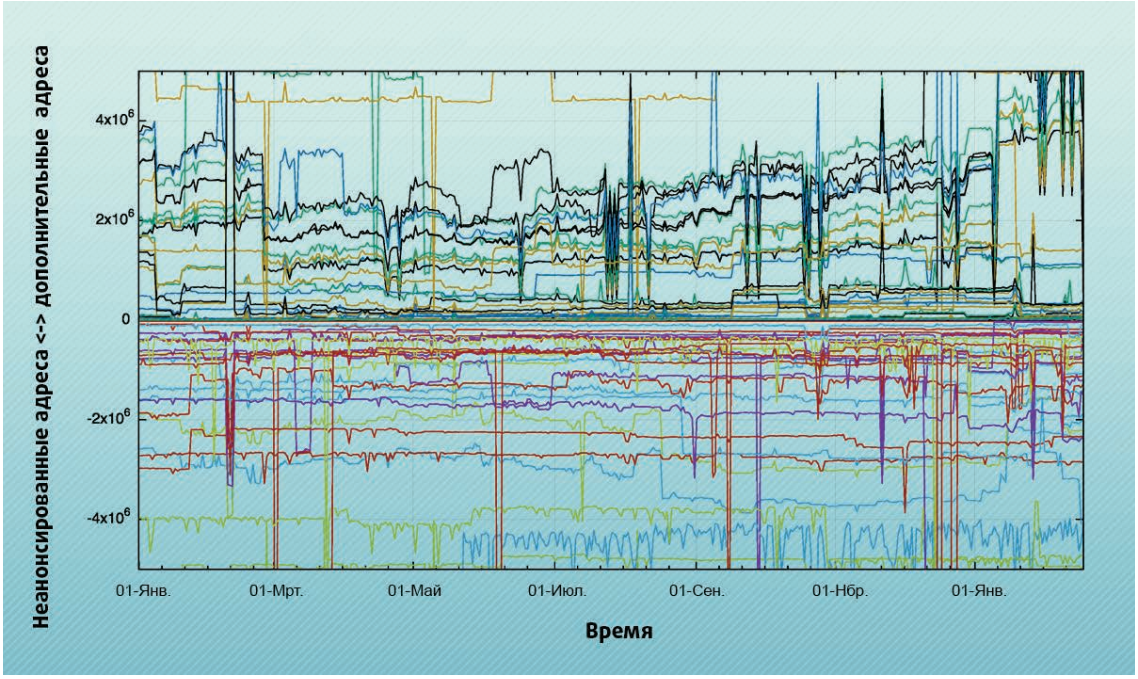
шлюз, анонс которого ограничен локальной областью, которая невидима в Route Views.

Рис. 5 предполагает, что IPv6 показывает тесный «кластер» достижимости адресов, и в этом отношении это, возможно, лучше, чем эквивалентная картинка для Интернета на базе IPv4. Однако это не является хорошим отображением Интернета на базе IPv6. Когда мы рассматриваем только данный центральный кластер (рис. 6), становится понятно, что здесь подобно IPv4 существует разброс, и некоторые сети «видят» намного больше IPv6-префиксов /32 (в некоторых случаях на 600 больше), чем другие, что эквивалентно вариации в диапазоне достижимости адресов в размере /23.

Давайте продолжим детализацию в этой области и перейдем на еще один более подробный уровень. Одним из способов изучения этого является определение «ядра» анонсированных префиксов. В данном случае мы будем использовать пороговое значение в 5/8 от набора пиров или 62,5%. Если 5/8 (или больше) пиров объявляют какой-либо префикс адреса, то мы можем включить его в «кворумный» набор объявленных маршрутов. После этого мы можем взять набор «корневых» префиксов кворума и определить кворумный диапазон адресов, который затем можно использовать для сравнения с каждым пиром, анонсирующим таблицу маршрутов в Route Views. Что нас интересует, так это область, в которой отдельные пиры анонсируют больший диапазон адресов, видимый в кворумном наборе, и, подобным же образом, суммарный диапазон адресов, существующий в кворумном наборе, но явно не анонсируемый пиром. Например, оператор Telstra из Австралии (AS1221) анонсировал диапазон адресов, эквивалентный 166.06 /8 из кворумного набора. Примерно 14 миллионов адресов IPv4, которые образуют часть кворумного набора, не были анонсированы компанией Telstra, при этом Telstra включила 199 552 анонсированных адреса, которые не входили в кворумный набор.

Полный анализ всех пиров из моментального снимка BGP проекта Route Views, зафиксированного 17 февраля 2016 года, показан в таблице 1. На основе этой таблицы очевидно, что никто не видит тот же самый набор адресов, что и любой другой участник,

Рис.8. «Некворумные» объявления адресов IPv4 для каждого пира за 2015 г.



при этом разброс может быть достаточно велик.

Таблица 2 показывает аналогичный отчет для IPv6. И опять очевидно, что здесь не существует единого ракурса. Некоторые пиры видят очень мало адресов сверх кворумного набора, тогда как другие видят 120-130 дополнительных адресов /32. Картина для пропущенных адресов также является смешанной. Некоторые пиры имеют очень мало пропущенных данных, тогда как другие не видят более крупный набор адресов.

Хотя единый моментальный снимок системы маршрутизации способен проиллюстрировать, что в терминах явно анонсированного адресного пространства в разных частях Интернета существуют значительные различия, он не может показать, являются ли такие различия следствием обычной работы динамического протокола маршрутизации, а это означает, что моментальный снимок, сделанный одним днем или даже одним часом позже, покажет существенно отличающуюся картину, либо же эти различия являются структурными, в каковом случае картина, показывающая различия в анонсированных наборах адресов для каждого пира, будет относительно постоянной с течением времени.

Анализ ежедневных моментальных снимков маршрутизации, сделанных в рамках проекта Route Views, показан на рис. 7. Масштаб этого графика равен ±100 миллионов адресов или примерно шесть диапазонов адресов /8. На этом графике каждый пир вычерчен в виде двух линий. Линия в положительной области представляет собой количество дополнительных адресов, которые пир анонсирует в дополнение к адресам, образующим кворумный набор. Линия в отрицательной области – это суммарный объем адресного пространства, которое входит в кворумный набор, но не было анонсировано данным пиром. Некоторые пиры «видят» значительно меньший диапазон адресов по сравнению с другими, и эта разница остается стабильной с течением времени.

Если мы посмотрим на кворум более близко, изменив масштаб графика до ±5 миллионов адресов (эквивалентно диапазону адресов /10), то можно увидеть аналогичную картину стабильности этих наборов адресов. На этом уровне детализации можно заметить некоторые суточные колебания. Что интересно, в данных имеется скачок в начале 2016 года, когда кластер пиров стал объявлять примерно четыре миллиона адресов (один /10), которые не включаются в состав кворумного набора. В середине 2015 года этот пул дополнительных адресов составлял приблизительно два миллиона и увеличился в два раза за прошедшие семь месяцев.

Мы можем также взглянуть на это на уровне детализации /16 или 65 536 адресов.

Рис.9. «Некворумные» анонсы адресов IPv4 для каждого пира за 2015 г.

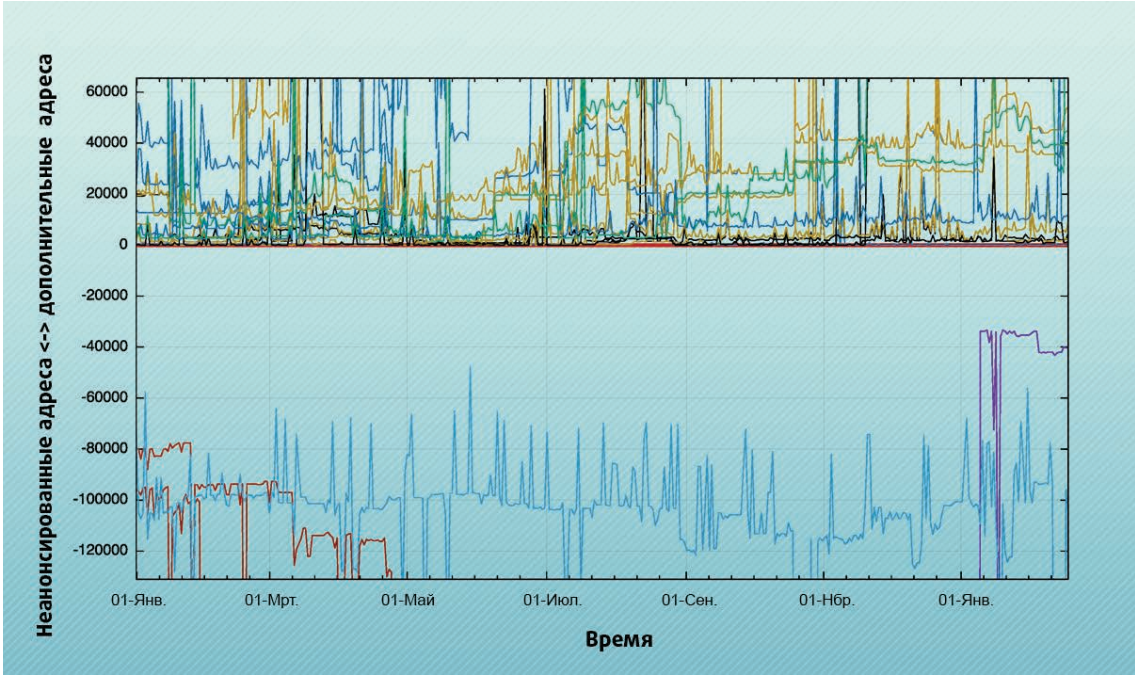
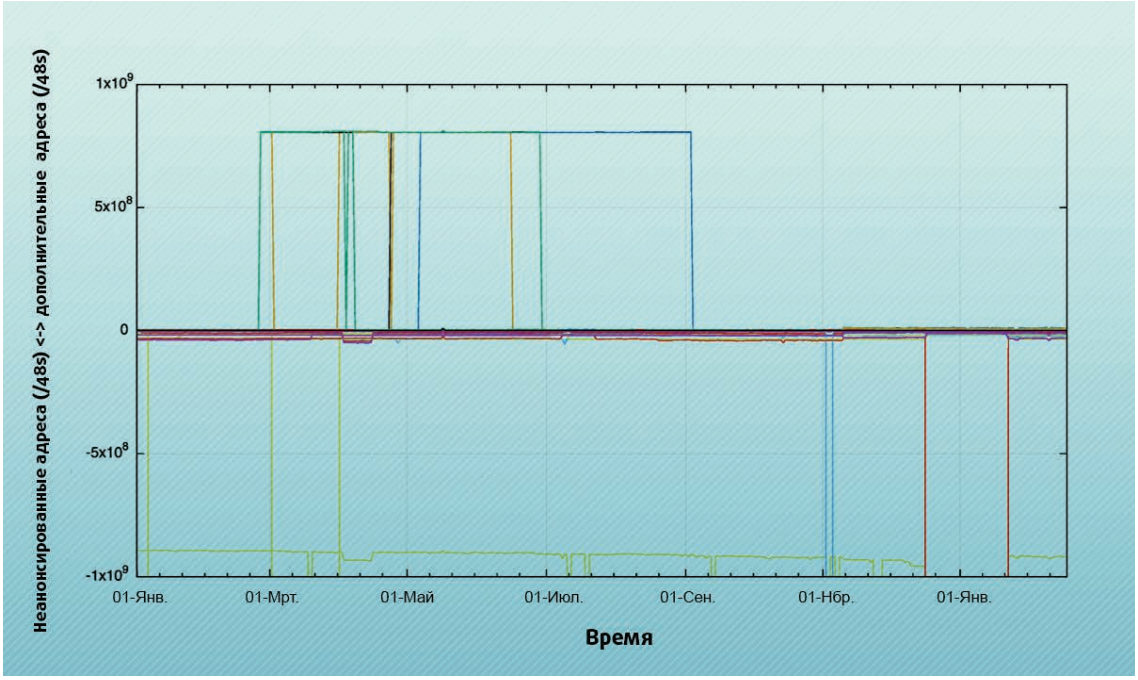


Рис.10. «Некворумные» анонсы адресов IPv4 для каждого пира за 2015 г.



Здесь ряд пиров «видят» дополнительные адреса, которые не входят в кворумный набор, а степень, в которой они отличаются от кворума, остается нестабильной в течение года. Несколько пиров отличаются от кворумного набора на (вплоть до) 130 тысяч адресов (рис. 9).

Аналогичная картина наблюдается для сети IPv6, где отдельные пиры могут «видеть» отличия диапазона адресов от кворумного набора, насчитывающие примерно 1000 миллионов /48 (рис. 10). Эти аномалии являются относительно долговременными, охватывая несколько месяцев.

На более высоком уровне детализации, охватывающем ±50 миллионов /48 (примерно один /22), можно заметить более значительную степень расхождения от кворумного набора, когда отдельные пиры не делают анонсы маршрутов, входящих в состав кворумного набора.

И наконец, при взгляде на те же данные в масштабе ±5 миллионов /48 (примерно один /27) можно заметить перемещения отдельных анонсов /32 внутрь и из расходящегося множества для каждого пира. При этом существует базовая стабильность этих различий в течение наблюдаемого периода времени.

Насколько серьезна эта проблема?

Другими словами, до какой степени попытки одной конечной точки Интернета напрямую связаться с другой конечной точкой затрудняются структурными разрывами связности Интернета? Возможно, что на этот базовый вопрос нельзя получить сколько-нибудь точный ответ.

Несмотря на то, что на этот вопрос, скорее всего, нельзя ответить, существуют некоторые представления, которые содержат косвенные указатели, помогающие количественно оценить серьезность этой проблемы в современном Интернете.

Первое наблюдение: то, что мы видим в системе маршрутизации как уровень управления данными, и то, что происходит на уровне пересылки данных, не всегда согласовано. Согласно тому, как это было показано в научной работе в IMC в 2009 году («Internet Optometry: Assessing the Broken Glasses in Internet Reachability», R. Bush, O. Maennel, M. Roughan, S. Uhlig, ACM SIGCOMM IMC, 2009), целый ряд сетевых операторов использовали маршрут «по умолчанию» для того, чтобы дополнить восходящую связность. Это подразумевает, что по мере того, как пакет проходит через последовательность восходящих соединений, он не обязательно должен следовать по явно объявленным маршрутам, а может перейти на маршрут по умолчанию. Что крайне важно для широко распространенной симметричной связности, так это то, что эти маршруты присутствуют как явные маршруты провайдеров первого уровня (tier 1), поскольку на этом уровне иерархии маршрутизации Интернета

Рис.11. «Некворумные» анонсы адресов IPv4 для каждого пира за 2015 г

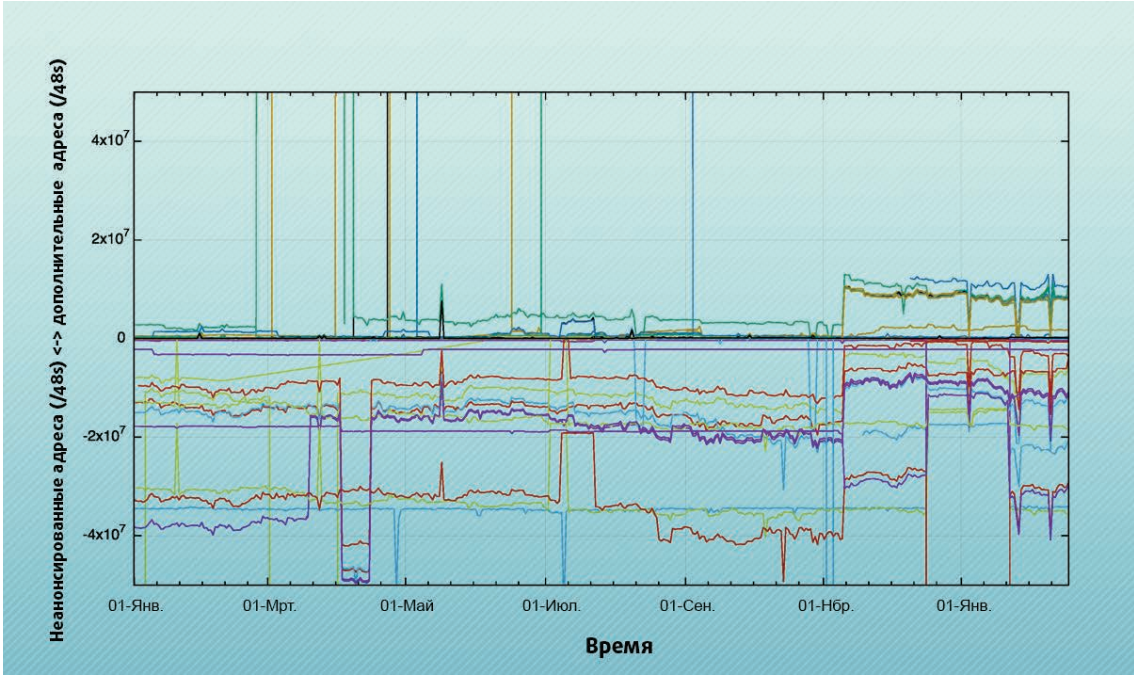
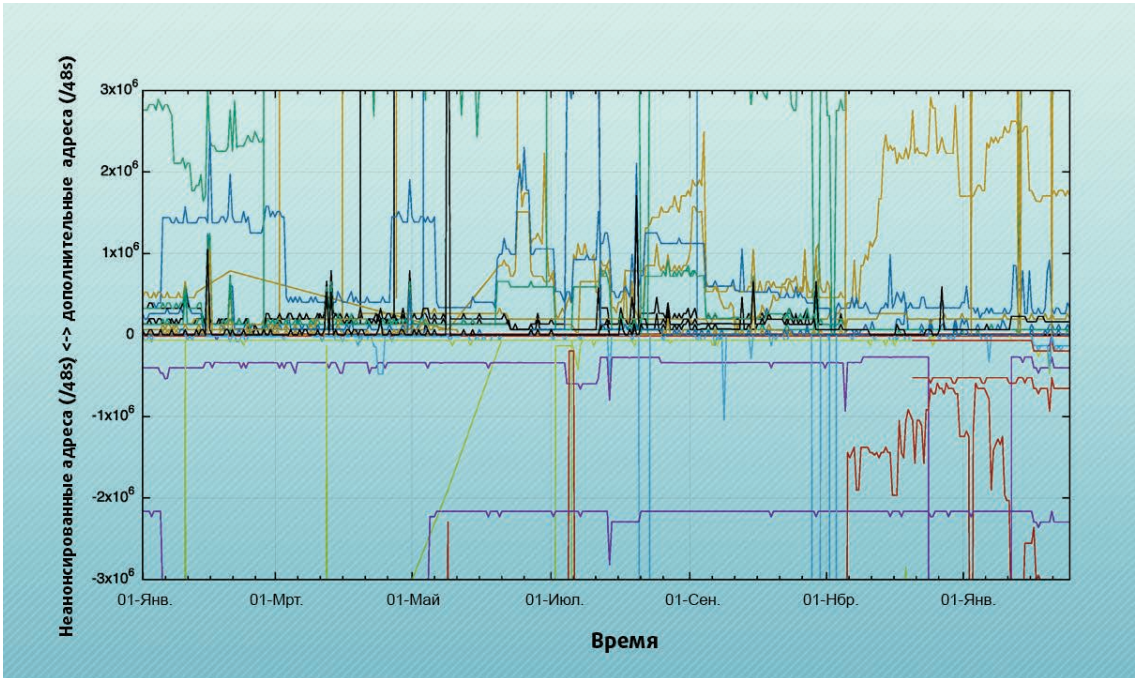


Рис.12. «Некворумные» анонсы адресов IPv6 для каждого пира за 2015 г.



маршруты по умолчанию не используются. Эта научная работа содержит сведения о том, что в то время примерно три четверти AS вели себя так, что это согласуется с восходящими маршрутами по умолчанию. Однако это открытие применимо к путям прохождения данных, которые ведут вдоль последовательностей «восходящих» отношений между AS. Довольно необычно видеть сетевого провайдера, направляющего свои маршруты по умолчанию через соединение пиров, и еще более необычным является маршрут по умолчанию, указывающий на сеть клиента. Поэтому использование маршрутов по умолчанию до некоторой степени помогает, но с помощью этого способа можно исправить только определенное подмножество аномалий маршрутизации, что ни в коем случае не является панацеей.

Вторая проблема – это асимметрия связности. Очень часто мы думаем, что нарушения связности являются симметричными, поэтому если конечная точка A не может отправить пакеты в конечную точку B, то предполагается, что и обратное тоже верно. В сетях с коммутацией пакетов, особенно в таких, как Интернет, которые используют однонаправленный протокол маршрутизации для поддержания своей внутренней топологии и достижимости, такие предположения о симметрии связности не являются верными. В рамках этих сценариев не являются редкими случаи, когда A не может передать пакеты в B, но B по-прежнему может передавать пакеты в A. Можем ли мы увидеть доказательство этому?

В APNIC мы используем систему измерения, в рамках которой каждый день примерно 4-10 миллионов конечных точек Интернета рекрутируются для выполнения небольшого набора базовых тестов на установление связи. Эти псевдослучайно выбранные браузеры, привлеченные практически со всех просторов Интернета, пытаются установить контакт с небольшим набором серверов, которые оборудованы и настроены для записи попыток соединения. Как это ни странно, не каждая попытка соединения получается успешной. Периодически возникает картина асимметричных неудач, при которых конечная точка может отправить пакет на сервер эксперимента, но попытка сервера ответить заканчивается безуспешно.

В течение 2015 года мы наблюдали за попытками соединения примерно 446 миллионов конечных точек IPv4 и зафиксировали 1,1 миллиона асимметричных неудач. Необработанные данные предполагают, что приблизительно одна из 400 попыток соединения заканчивается неудачей в подобной асимметричной манере. Является ли это базовым доказательством того, что не все соединено со всем в любой момент времени? Ежедневная частота неудачных соединений показана на рис. 13.

Рис.13. Частота отказов для асимметричных соединений IPv4.

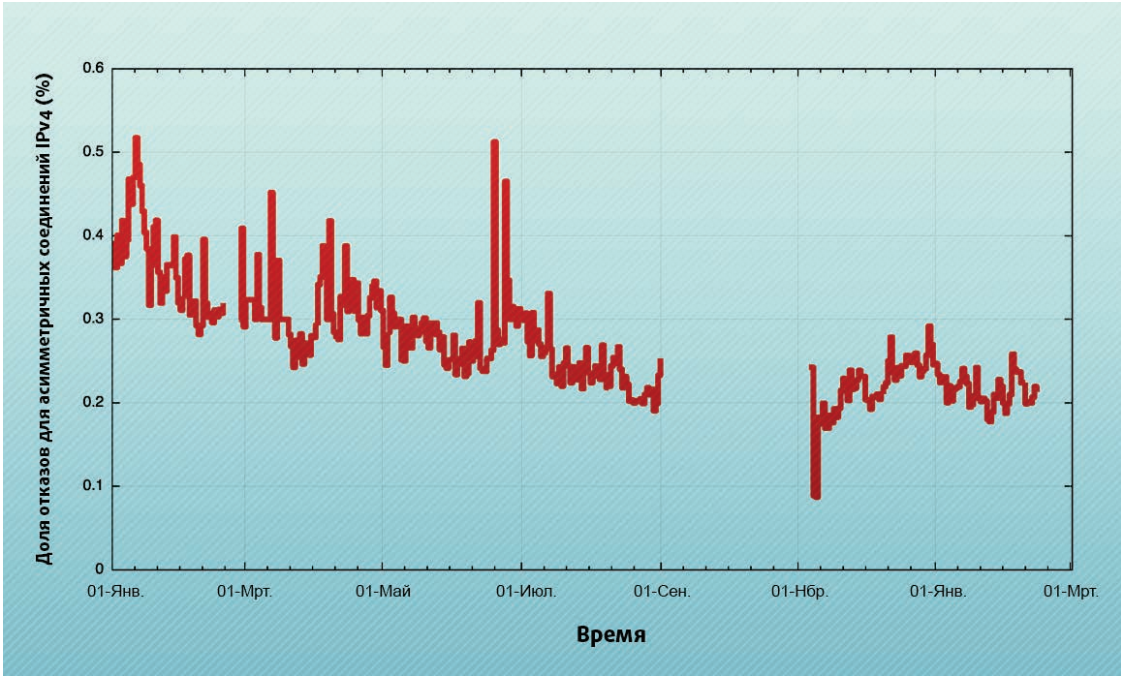
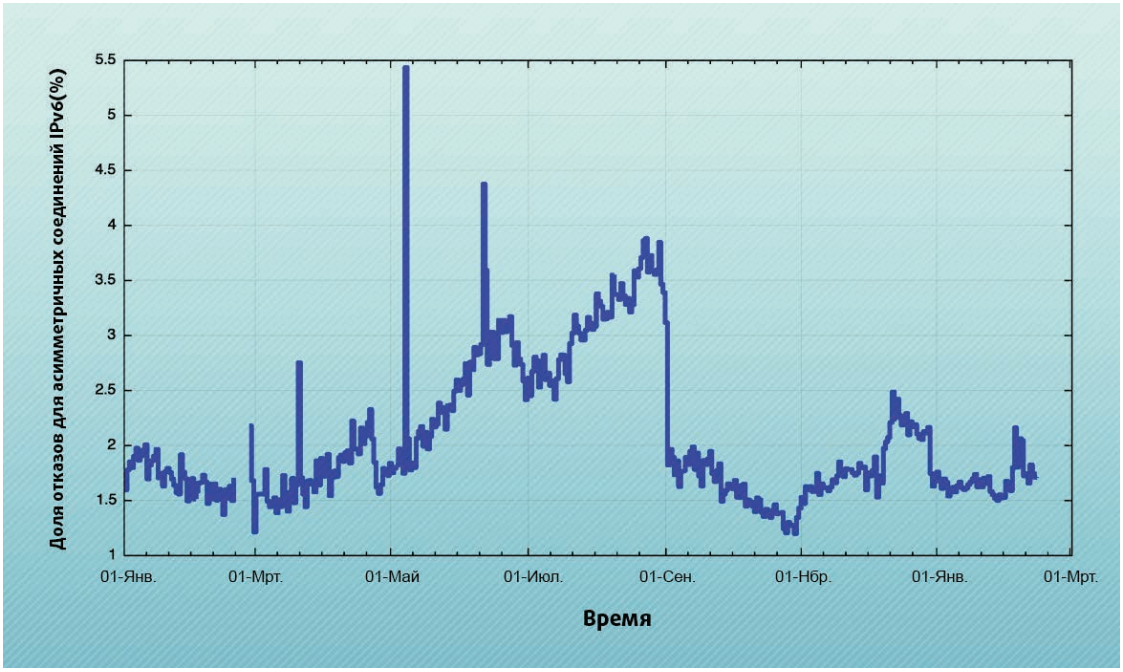


Рис.14. Частота отказов для асимметричных соединений IPv6.



Существует целый ряд потенциальных причин для неудачи соединения в этом конкретном эксперименте. По данным невозможно провести различие между соединениями, инициированными в контексте эксперимента, и различными сканами адресов, которые сканируют адресное пространство Интернета с помощью TCP SYNS в портах 80 и 443. Поэтому маловероятно, чтобы все асимметричные неудачные попытки соединения, показанные на рис. 13, объяснялись асимметричной связностью. Однако весьма вероятно, что значительная часть неудач произошла вследствие данной конкретной формы фрагментации асимметричной связности в Интернете.

Сравнимая картина для асимметричных неудач IPv6 демонстрирует еще более плохую ситуацию (рис. 14).

Вполне вероятно, что в эту относительно высокую частоту асимметричных отказов, составляющую 1 из 50 соединений, вносит свой вклад целый ряд дополнительных факторов, включая подозреваемые проблемы с абонентским оборудованием (CPE, Consumer Premises Equipment), и фрагментарная поддержка IPv6, однако в эту разочаровывающую цифру вносит свой вклад и асимметричная сетевая связность.

Почему это происходит?

Или, если задать этот вопрос в противоположном смысле, почему весь Интернет полностью не соединен между собой? Несомненно, это тот случай, когда индивидуальные мотивы совпадают с общественным благом. Каждая подключенная сеть лучше всего обслуживается, когда она достижима из всех других подключенных сетей, и потенциально на сети пагубно отражается ситуация, когда существуют сети, которые не могут с ней связаться. Кроме того, это симметричное пожелание – то же самое применимо к набору сетей, с которыми может связаться данная подключенная сеть. Теоретическое значение соединения максимизируется, когда сеть способна достичь всех других подключенных сетей (и быть достигнутой ими).

Однако на практике невозможно приобрести услугу, которая гарантирует такую всеобъемлющую достижимость. Провайдеры услуг стремятся осуществить такие ожидания от имени своих заказчиков, однако всеобъемлющая связность попадает в категорию «лучшее из возможного», что отличается от «гарантии услуги».

Почему это происходит?

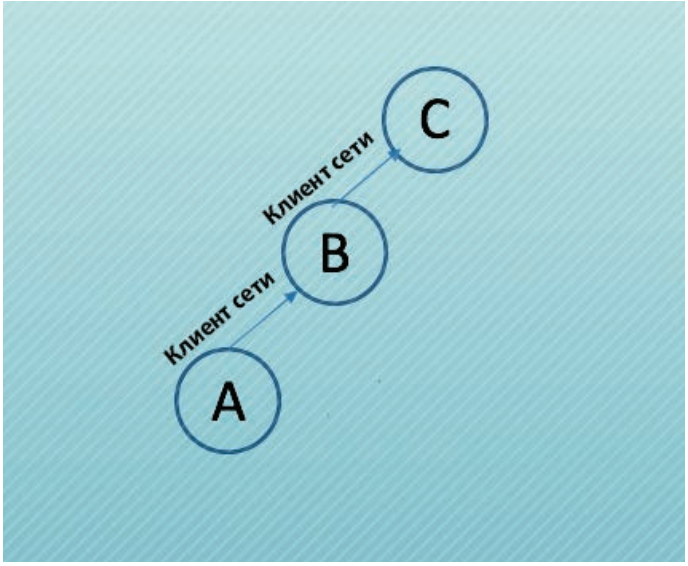
Всеобъемлющая связность не является требованием, налагаемым каким-либо регулирующим указанием, или предметом заранее подготовленного соглашения между сетевыми операторами. Межсетевые соединения имеют свой собственный рынок, а результаты можно рассматривать как рыночные.

Сеть каждого отдельного оператора услуг работает в рамках домена или «пиринга» либо «ярусности». «Ярусность» обозначает подразумеваемую структурированность сетей с учетом набора отношений клиента и провайдера. Такая структурированность обычно бывает иерархической, т.е. если A – это клиент B, а B – клиент C, то крайне необычной является ситуация, когда C является клиентом A (рис. 15).

В данном примере A – это сеть, работающая на уровне 3, B – на уровне 2 и C – на уровне 1, при этом деньги, связанные с предоставлением услуг связи и транзитных услуг, обычно текут по тому же пути. В данном примере отсутствуют любые другие отношения между AS, A ожидает, что B сможет предоставить полный набор маршрутов ко всем другим подключенным сетям, а B аналогичным образом полагается на C.

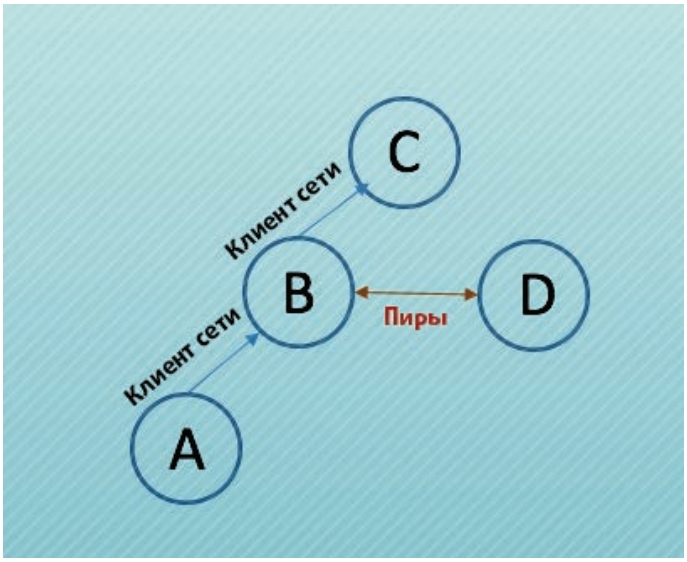
«Пиринг» обозначает немного другие отношения между сетями, в рамках которых ни одна из сетей не является клиентом другой. Ти-

Рис.15. Отношения клиент/провайдер.



повой шаблон пиринговых отношений заключается в том, что две пиринговые сети обмениваются информацией о достижимости своих собственных маршрутов и маршрутов своих клиентов, но не обмениваются маршрутами, о которых они узнали, являясь клиентом провайдера более высокого уровня, а кроме того, они не обмениваются маршрутами, о которых они узнали в результате других пиринговых отношений. Поэтому в рамках нашей простой схемы, состоящей из трех сетей A -> B -> C, если мы введем четвертую сеть – D, которая является одноранговой с сетью B, то D сможет узнать о том, как достичь адресов конечных точек, расположенных в сетях B и A, но не C (рис. 16).

Рис.16. Отношения клиент/провайдер и отношения пиринга.



Для D пиринговые отношения с A не приведут к расширению достижимости, и если бы даже это произошло, B может прийти к выводу, что «поскольку D имеет пиринговые отношения с одним из моих клиентов, D также должен быть моим клиентом». Пиринговые отношения обычно подразумевают, что обе сети расположены на одном и том же уровне в иерархии клиент/провайдер. Очевидно, что реальность мира коммерции гораздо изощреннее, чем приведенный пример, и сетевые операторы сумели изобрести многочисленные вариации на эту базовую тему, однако лежащие в основе принципы межсетевых соединений остаются относительно неизменными.

Если бы все эти отношения были статичными, то ситуация оставалась бы легко управляемой, однако курс системы постоянно меняется. Провайдеры меняют свои отношения в экосистеме связности. Появляются новые провайдеры, например, сети распределения контента, и конечно, происходят приобретения, слияния и разделения (компаний), после которых возникшие организации должны заново откалибровать свои позиции в мире связности.

На ситуацию можно посмотреть таким образом, что удивительно, каким образом связность Интернета остается такой стабильной и всеобъемлющей, учитывая тот факт, что этот результат зависит от потребностей рынка без каких-либо конкретных гарантий правильного исхода.

Возможно, результат не является столь уж удивительным. Еще один взгляд, нельзя не признать, что циничный, заключается в том, что все это – результат действия того, что можно приблизительно назвать «информационным картелем» провайдеров первого уровня, которые расположены в самом ядре связности. До тех пор, пока каждая подключившаяся сеть предпринимает усилия для того, чтобы ее маршруты были объявлены в по крайней мере одном маршрутизаторе уровня 1 через одно или несколько отношений клиент/провайдер, результатом будет некоторый уровень базовой связности. После достижения этой базовой связности вступает в действие пиринг, обеспечивающий минимизацию стоимости и/или улучшение обслуживания для выбранных маршрутов. В рамках этой точки зрения связность на уровне всего Интернета почти полностью определяется способностью ввести свои маршруты в картель провайдеров первого уровня. Эта группа взаимосвязанных пиринговых сетей по существу определяет, что подлежит подключению в Интернете. Поэтому другой взгляд на связность Интернета утверждает, что это не распределенный открытый рынок для связности, а вместо этого в его «ядре» мы имеем самовозобновляющуюся монополию маршрутизации!

Является ли Интернет полностью взаимосвязанным?

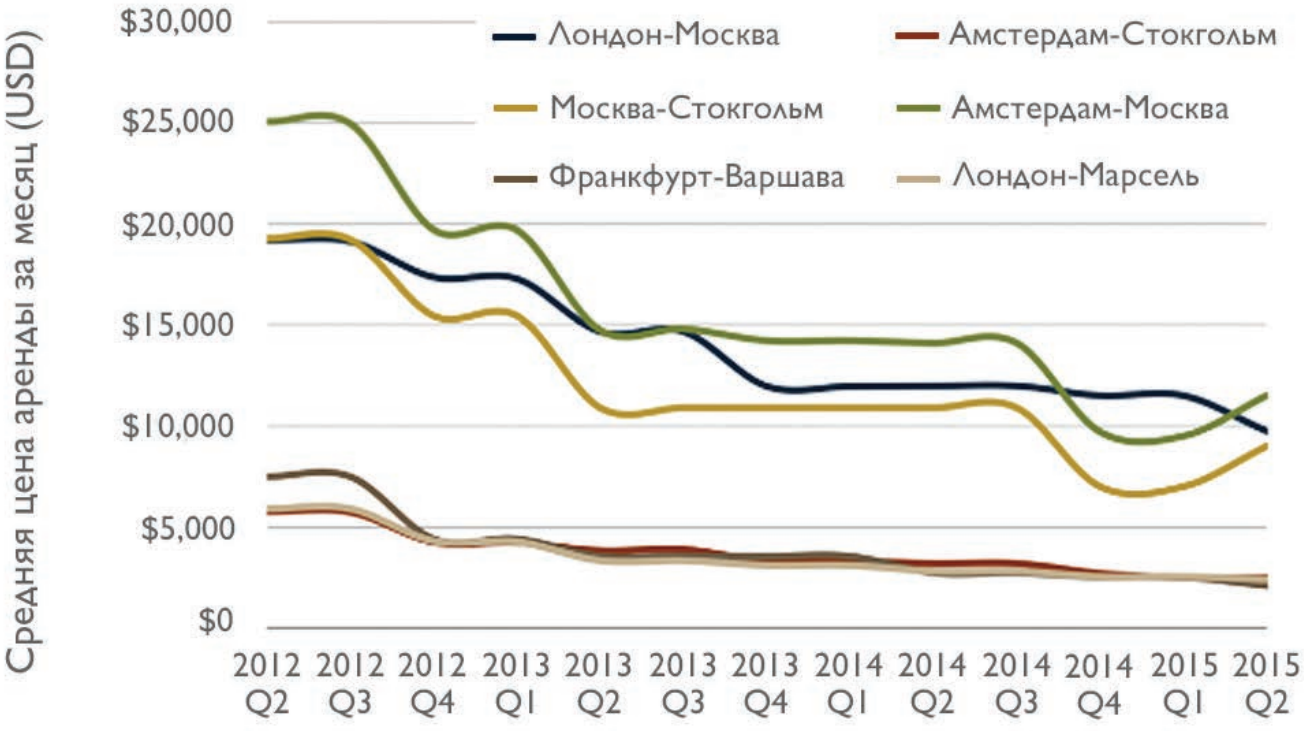
Нет.

По его границам существует серая зона асимметричной связности, внутри которой вы, возможно, сможете отправить мне пакет данных, но это отнюдь не означает, что вы получите мой ответ!

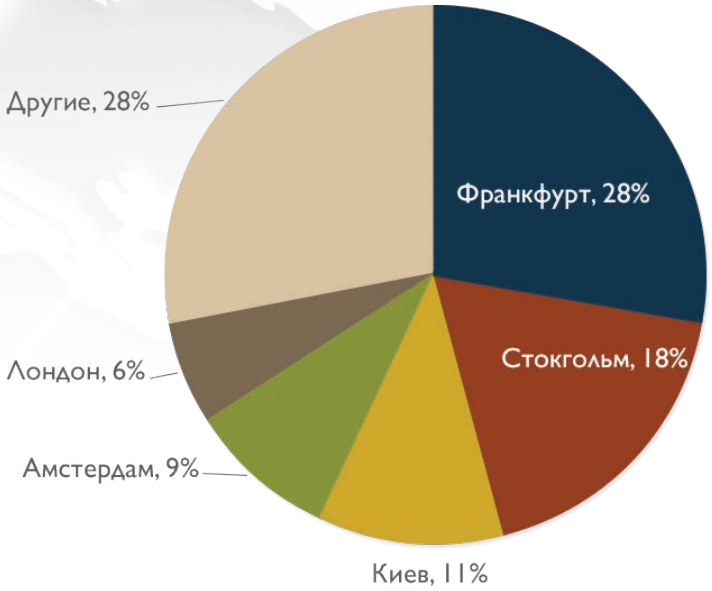
Источник: On the Internet Everyone is Connected to Everyone Else - Right?, http://www.potaroo.net/ispcol/2016-02/connected.html

МЕЖДУНАРОДНАЯ СВЯЗНОСТЬ

Тенденции ценообразования на некоторых машрутах

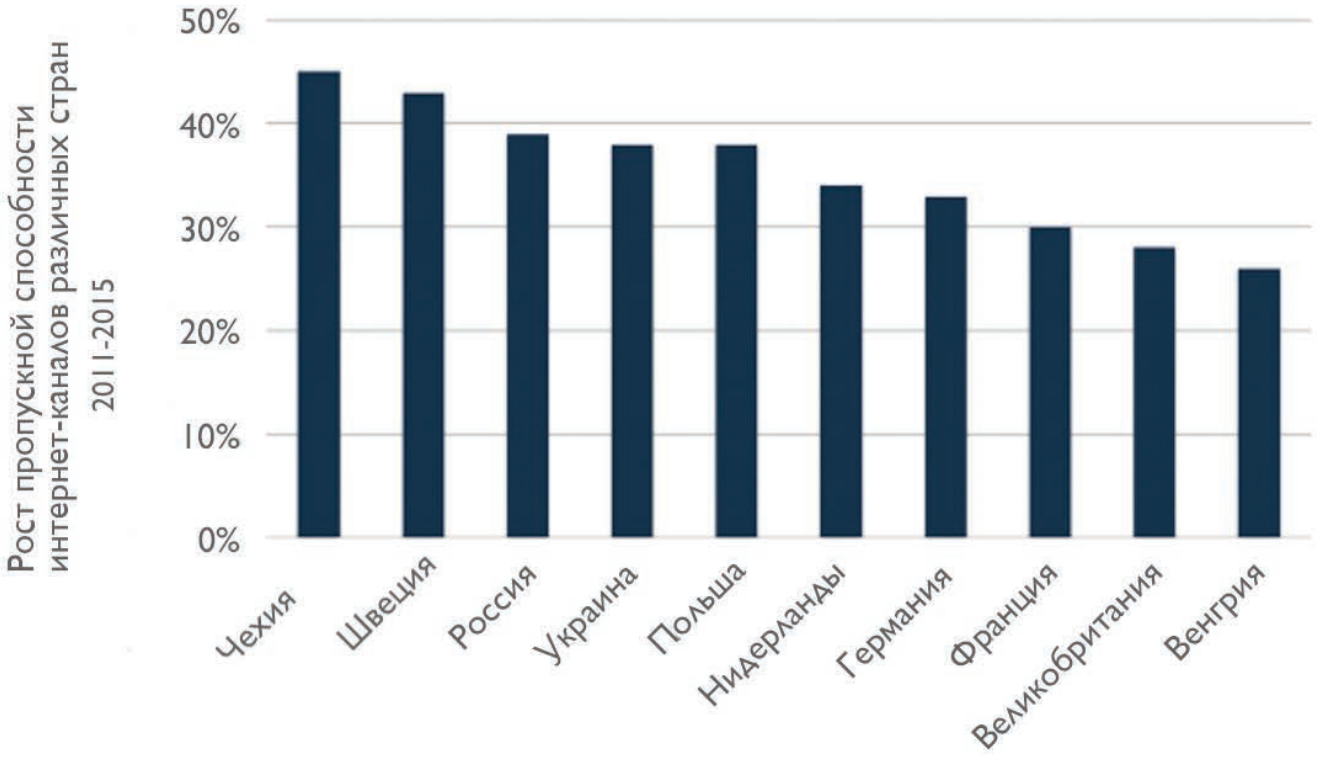


Источник: TeleGeography, www.telegeography.com



Пропускная способность международных каналов на Москву, 2015

Рост пропускной способности международных каналов, CAGR, 2011-2015



Источник: TeleGeography, www.telegeography.com

Транспортные протоколы

Джефф Хьюстон (Geoff Huston)

Хотя работа Интернета на низком уровне во многом происходит посегментно, так сказать hop-by-hop, работу основных приложений осуществляют два сквозных транспортных протокола: User Datagram Protocol (UDP) и Transmission Control Protocol (TCP). Джефф Хьюстон раскрывает в этой статье принципы работы и эволюцию этих протоколов. Особое внимание он уделяет протоколу TCP. Удивительно, что этот протокол способен обеспечить эффективное обслуживание вне зависимости от задействованной скорости – десятков бит в секунду или миллиардов бит в секунду. Кроме того, удивительно то, что TCP остается эффективным вне зависимости от того, сколько «разговоров» (сеансов) проходит по одному каналу – является ли он единственным или одним из миллионов параллельных «разговоров» TCP. Но что на самом деле удивительно, так это сама технология, сегодня используемая в миллиардах устройств, которая остается гибкой и адаптируемой. И мы все еще можем сделать ее лучше.

Одним из ранних усовершенствований модели интернет-протокола стало отделение изначального протокола Интернета от единой монолитной спецификации и разделение на Internet Protocol (IP) и на пару транспортных протоколов. Уровень IP предназначен для использования внутренними коммутаторами внутри сети для передачи пакетов в указанное для них место назначения, тогда как уровень транспортного протокола используется между системой отправителя и получателя. В этой статье я бы хотел взглянуть на то, что происходит с этими транспортными протоколами.

В сегодняшнем Интернете обычно используются два сквозных транспортных протокола: User Datagram Protocol (UDP) и Transmission Control Protocol (TCP). Почему только два? Несомненно, мы обдумывали много способов управления потоками данных через сеть с коммутацией пакетов, и в рамках такой открытой архитектуры, как Интернет, кажется разумным спросить, почему не видно множества сквозных транспортных протоколов. В таких попытках не было недостатка, и просмотр /etc/protocols в любой Юникс-подобной системе предоставит список из 130 таких протоколов, которые находятся на уровне поверх IP. Некоторые из них представляют собой специализированные протоколы, например, протокол 89, используемый протоколами маршрутизации OSPF, а другие являются протоколами инкапсуляции, например, протокол 41 для инкапсуляции IPv6-in-IPv4. Однако в этом списке также присутствует достаточное количество сквозных протоколов, таких как протокол 27 – Reliable Datagram Protocol (RDP), протокол 132 – Stream Control Protocol (SCTP). Однако большинству этих сквозных транспортных протоколов оказалось трудно выжить в общедоступном Интернете. Проблема сегодняшнего дня заключается в том, что промежуточные устройства взяли верх. Значительная доля Интернета укрывается за всякого рода типами промежуточного программного и аппаратного обеспечения, например, в виде брандмауэров (межсетевых экранов) и фильтров, работающих по принципу явного разрешения, а не исключения. Эти средства обычно разрешают протоколы 6 и 17 (TCP и UDP), однако в отношении разрешенных сквозных протоколов на этом все и заканчивается. Это не просто паранойя брандмауэров. У нас имеются выравниватели нагрузки, формирователи пакетов и многие другие виды перехватывающих промежуточных устройств, которые способны распознавать и управлять сеансами TCP, но обычно они больше ничего не делают. Поэтому для всех практических целей Интернет разрешает только сквозные транспортные протоколы TCP и UDP. Давайте взглянем на эти два протокола более подробно.

UDP

UDP – это очень простая абстракция базовой IP-датаграммы. И подобно самому IP, UDP является ненадежным средством. Зеркально отображая сервисные характеристики IP, пакеты, отправленные с помощью UDP, возможно, будут или не будут получены по их целевому месту назначения. Пакеты UDP могут перегруппировываться, дублироваться или теряться. В UDP не существует управления потоком данных или его регулирования. Дискретизация пакетов в UDP является явной: если данные разделены отправителем на два пакета UDP, то получатель должен принять данные с помощью двух отдельных операций чтения.

UDP предназначен для использования в чрезвычайно простых транзакциях, которые не требуют сеансового контекста. Система доменных имен (DNS, Domain Name System) и Сетевой протокол синхронизации (NTP, Network Time Protocol) являются хорошими примерами приложений, которые используют UDP для поддержки очень эффективной модели транзакций запрос/ответ. Обычно отправитель генерирует пакет запроса вместе с пространством для ответа, а ответ представляет собой тот же пакет, но с заполненным полем для ответа.

В наши дни к UDP относятся гораздо более настороженно. Отсутствие сеансового контекста означает, что большинство транзакций являются незашифрованными и не только легко «прослушиваются» третьими сторонами, но и чрезвычайно уязвимы к атакам, связанным с введением в заблуждение, при которых некто другой (а не целевой получатель) генерирует ответ с возможной целью введения в заблуждение или обмана исходного отправителя запроса. Например, в современном Интернете средства перехвата DNS стали весьма обычным способом фильтрации контента. Более зловещим выглядит то, что протокол UDP стал общей платформой для формирования различных типов DOS-атак. Многие серверы UDP, например, официальные DNS-серверы, должны отвечать на все UDP-запросы, поэтому они не могут отвечать только на определенные «подлинные» запросы. В UDP все запросы имеют обыкновение выглядеть подлинными. Эта особенность используется злоумышленниками, которые помещают IP-адрес намеченной жертвы в исходный адрес UDP-пакета. При отправлении достаточно большого количества запросов, содержащих один и тот же исходный адрес, возможно «рекрутировать» UDP-серверы, превратив их в невольных злоумышленников, атакующих жертву. Поэтому сегодня протокол UDP «впал в немилость».

Если не UDP, тогда, вероятно, TCP.

TCP

TCP – это надежный сквозной протокол управления потоком данных. Поток данных, входящий в TCP-сокет на одном конце, будет считан как поток данных на другом конце. Поскольку это потоковый протокол, дискретизация пакетов скрыта от приложений, как и механика управления потоком, обнаружение потерь и повторная передача, создание и завершение сеансов. TCP не обеспечивает никакой синхронизации потока, но поддерживает его целостность.

Внутри любой сети с коммутацией пакетов при возникновении соперничества за общий выходной порт либо при длительном превышении объема передаваемых данных доступной емкости канала коммутатор пакетов будет использовать очередь для удержания лишних пакетов. После заполнения такой очереди коммутатор пакетов должен сбрасывать пакеты. Любой надежный протокол передачи данных, который работает в рамках сети, должен распознавать такую возможность и предпринимать меры по исправлению ситуации. TCP не является исключением. Один из подходов заключается в том, что каждая пара коммутаторов использует надежный протокол для их общего соединения, осуществляя обнаружение и корректировку потери пакетов посегментно (hop-by-hop). TCP использует другой подход и не делает никаких допущений о надежности каждого сегмента, или хопа. Вместо этого TCP использует сквозную последовательную нумерацию данных между двумя коммуникационными системами, которая позволяет идентифицировать контекст в рамках потока данных. После получения хостом последовательно включенного пакета он отправляет обратно отправителю порядковый номер последнего байта в пакете в качестве положительного подтверждения (ACK). Это подтверждение ACK представляет собой «кумулятивный ACK» в том смысле, что оно «говорит» о получении всех данных в потоке вплоть до порядкового номера, который содержится в ACK. В TCP не существует негативных подтверждений (NACK). После прибытия пакета вне заданного порядка получатель также отправляет ACK, но содержащий порядковый номер полученного байта с самым высоким номером, который соответствует заданному порядку. Такая форма конструкции надежного протокола именуется «сквозным» управлением, в отличие от «похопового» управления, поскольку TCP не допускает, что внутренние коммутаторы попытаются исправить потерю пакетов на уровне каждого сегмента.

TCP использует эти подтверждения ACK как некую форму обратной связи от получателя к отправителю для синхронизации потока данных. Подтверждения ACK поступают обратно отправителю с частотой, примерно равной интервалам, с которыми пакеты данных прибывали к отправителю. Если TCP использует эти подтверждения ACK для инициации дальнейшей отправки пакетов данных по сети, то он может отправлять данные в сеть с той же частотой, с которой они покидают сеть в пункте назначения. Такой режим работы именуется «ACK-синхронизацией». С точки зрения поддержки стабильности сети, такое конструкторское решение является очень дальновидным. В устойчивом состоянии этот механизм гарантирует, что TCP вводит данные в сеть с той же скоростью, с которой они выходят из сети на другом ее конце.

Немаловажно, что современный Интернет допускает, что большая часть сетевых ресурсов выделяется на прохождение TCP-трафика, а также предполагает, что алгоритмы управления потоком, используемые этими сеансами TCP, ведут себя примерно схожим образом. Если ресурсы коммутации и передачи сети рассматриваются как общий ресурс, то предположение о единообразном поведении сеансов TCP подразумевает, что сквозные транспортные сеансы будут вести себя аналогично и в условиях соперничества. В результате, с учетом разумного уровня упрощения, параллельные сеансы TCP самостоятельно найдут положение равновесия, чтобы дать каждому сеансу TCP равную долю общего ресурса. Другими словами, сама сеть не должна «справедливо» распределять проходящие через нее TCP-потоки – до тех пор, пока все потоки управляются единообразным алгоритмом управления, эти потоки будут взаимодействовать друг с другом таким образом, чтобы (вероятно) выделить равную долю сетевых ресурсов каждому активному TCP-потоку. По крайней мере, в теории.

Но совпадают ли между собой теория и практика? Происходит ли это в современном Интернете и что меняется в этих допущениях о поведении TCP?

Управление потоками данных TCP – TCP Tahoe и TCP Reno

Возможно, это удивительно, но TCP не имеет единого алгоритма управления потоками данных. Хотя спецификация обычного протокола TCP определяет, каким образом устанавливать и закрывать сеанс, а также определяет способ, которым полученные данные подтверждаются отправителю, спецификация базового протокола не устанавливает, каким образом обе конечные точки договариваются о скорости передачи данных между ними. Простой подход заключается в том, чтобы отправлять данные до тех пор, пока не наполнится буфер неподтвержденных данных отправителя. Полученные подтверждения ACK позволяют отправителю уменьшить этот буфер, после чего он опять может отправлять данные в сеть до его повторного наполнения.

Однако такой простой подход имеет свои проблемы при его использовании несколькими параллельными сеансами, приводящие к «коллапсу сети из-за заторов». Сеансы TCP взаимодействуют таким образом, что происходит крупномасштабная потеря пакетов, а потеря сигналов ACK заставляет всех отправителей повторно передавать данные и т.д. Проведенное в 1980-х исследование этого феномена привело к введению «управления потоками» в поведение TCP.

Одним из ранних алгоритмов управления потоками данных TCP стал TCP Tahoe, впервые использованный в операционной системе 4.3BSD. В рамках этой схемы управления потоками существуют две отличительные стадии поведения: этап «Медленный старт», когда скорость отправки удваивается каждый временной интервал передачи и подтверждения (RTT, Round Trip Time), и этап «Предотвращения заторов», когда скорость отправки увеличивается на фиксированную величину (один размер сегмента сообщения (MSS, Message Segment Size)) за каждый интервал RTT.

Этап медленного старта инициирует поток данных очень консервативным способом, отправляя в сеть всего один пакет и ожидая соответствующего подтверждения ACK. Однако каждый полученный ACK приводит к тому, что отправитель удваивает размер окна отправки. В результате отправитель будет последовательно отправлять в сеть 2, 4, 8 и т.д. пакетов по истечении каждого интервала RTT. Это приводит к экспоненциальному росту скорости передачи данных в сеть, и сеанс TCP очень скоро либо достигает максимальной скорости приема удаленного получателя, максимальной начальной скорости передачи отправителя («порог медленного старта» (ssthresh (slow start threshold)), либо он столкнет сеть в затор до точки потери пакетов. В первом случае скорость потока TCP стабилизируется на максимальной скорости получателя. Во втором случае, когда скорость отправки превышает локальный порог, отправитель затем переходит в режим предотвращения заторов и продолжает увеличивать скорость отправки до тех пор, пока либо не будет достигнута максимальная скорость получателя, либо произойдет потеря пакетов.

Если подтверждение ACK для пакета не поступает в течение интервала RTT, то TCP Tahoe предполагает, что отправленный пакет был потерян в сети. Это событие потери пакета приводит к тому, что локальное значение ssthresh устанавливается равным половине текущей скорости отправки, а сама скорость отправки возвращается к одному пакету при возобновлении процесса медленного старта. Идеализированная картина итогового поведения потока данных, управляемого протоколом TCP Tahoe, показана на рис. 1. Идея заключается в том, что при запуске процесса отправитель быстро зондирует скорость в направлении повышения с целью определения верхней границы устойчивой скорости передачи данных. К этому моменту отправитель знает примерную скорость. Логическая аргументация предполагает, что потеря из-за затора произошла в ходе последнего интервала RTT, и что максимальная устойчивая скорость отправки находится где-то между половиной скорости отправки последнего интервала RTT и полной скоростью отправки того же последнего интервала RTT. После этого Tahoe опять выполняет медленный старт, но останавливается на этой половинной скорости, которая, согласно результатам предыдущего этапа медленного старта, была устойчивой скоростью для сети. Отправитель входит в режим предотвращения заторов и осуществляет гораздо более осторожное зондирование с целью нахождения следующего диапазона скоростей отправки, увеличивая скорость отправки на один сегмент каждый интервал RTT. И опять при потере пакета Tahoe предполагает, что произошло пересечение границы, после чего задает ssthresh равным половине скорости отправки и перезапускает поток в режиме медленного старта.

Tahoe использует отсутствие ACK как сигнал о потере пакета в результате затора. Tahoe осуществляет внутреннюю оценку RTT и при неполучении ACK в течение интервала RTT (плюс некоторое время, выделяемое на джиттер (колебания задержки)) Tahoe осуществляет сброс в режим медленного старта. Такой ответ на потерю пакетов приводит к значительным задержкам в передаче данных, поскольку отправитель бездействует в ходе интервала ожидания, а при перезапуске процесс возобновляется с обмена одним пакетом, постепенно достигая скорости передачи, которая была до потери пакета.

Для решения этой проблемы протокол TCP Reno ввел механизм «быстрого восстановления». Этот механизм инициируется последовательностью из трех дублирующих ACK, полученных отправителем данных. Эти дублирующие подтверждения ACK генерируются пакетами, которые следуют за потерянным пакетом, где отправитель подтверждает каждый из этих пакетов

Рис.1. Идеализированное управление потоками данных TCP Tahoe.

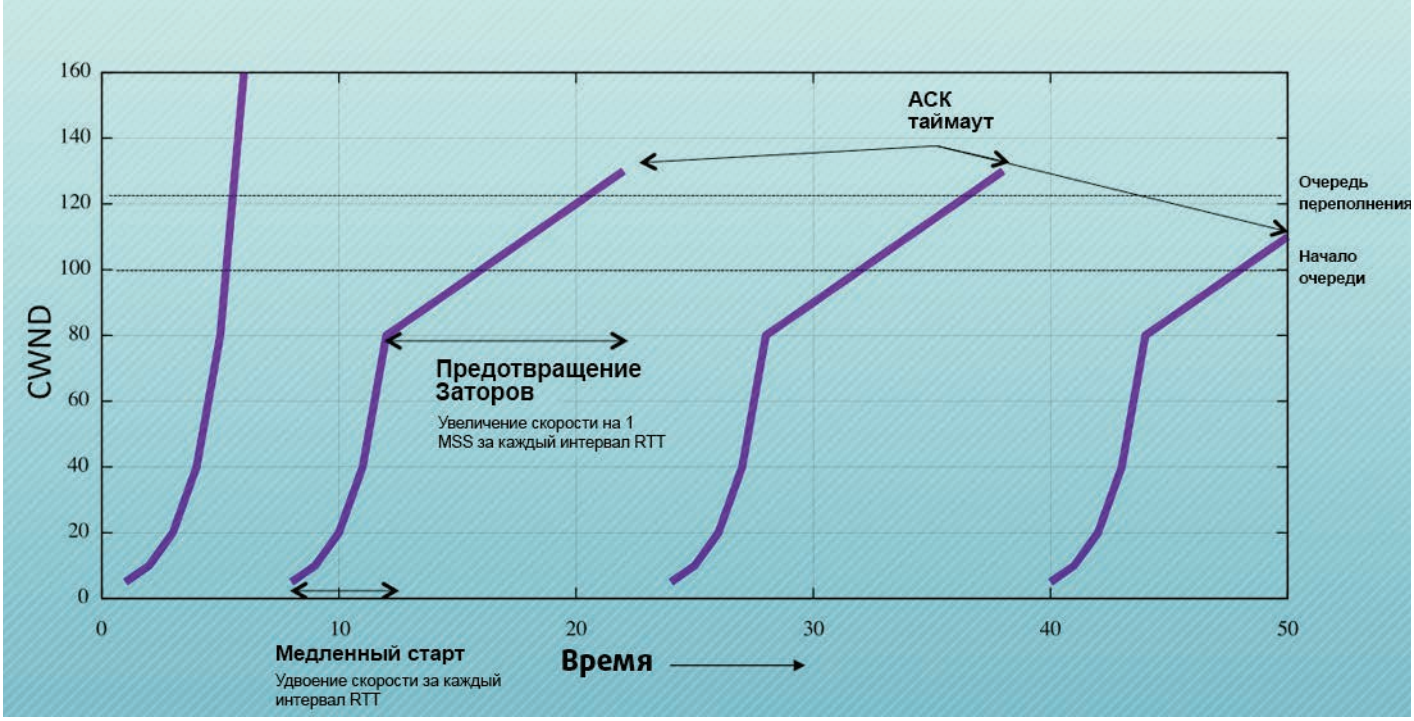
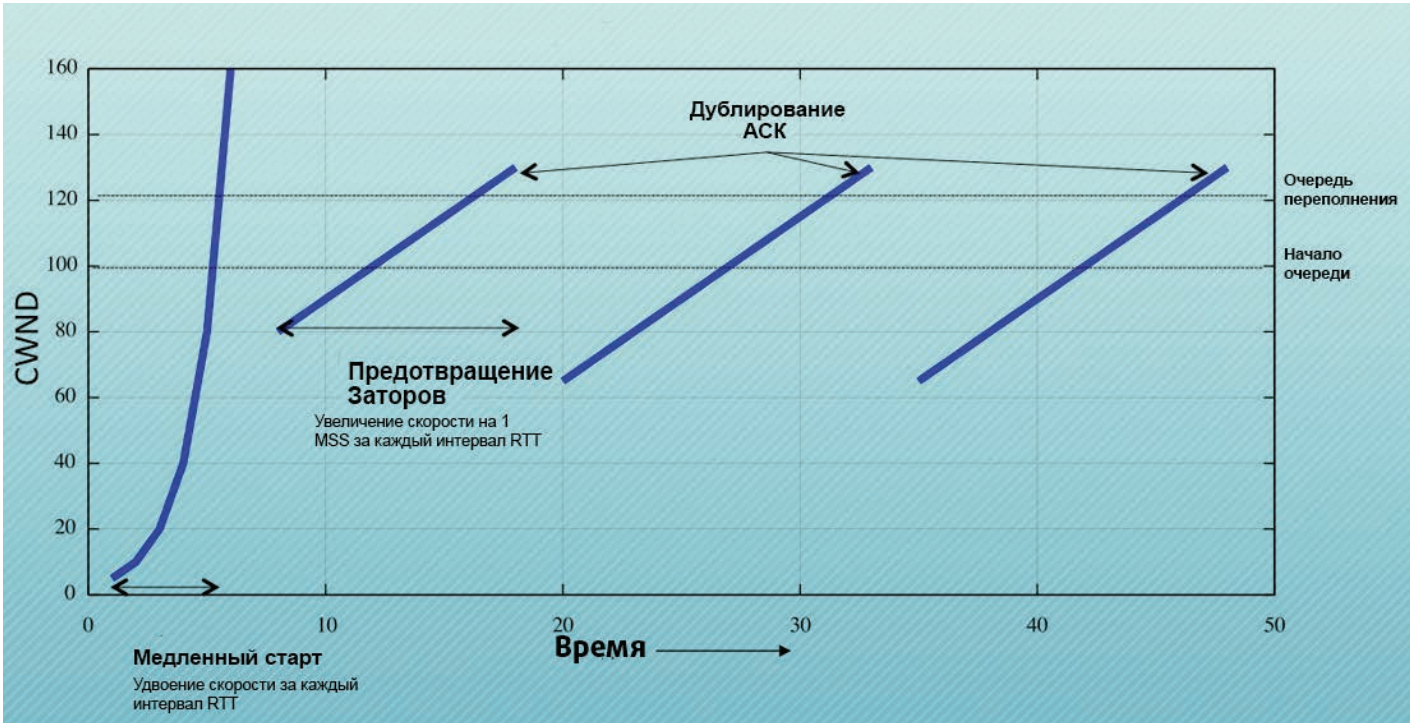


Рис.2. Идеализированное управление потоками данных TCP Reno



с помощью порядкового значения ACK последнего включенного в последовательность байта. В этом режиме отправитель немедленно повторно передает потерянный пакет и затем продолжает пошаговое продвижение ACK, пока продолжают прибывать дублирующие ACK. После того, как он получит свидетельство получения восстановительной передачи данных (согласно счетчику ACK, превысившему порядковый номер потерянных данных), отправитель возобновит режим предотвращения заторов при скорости, равной половине той скорости, которая использовалась во время получения дублирующих ACK.

Протокол резко реагирует на эти дублирующие сигналы ACK сетевого затора, но использует только половину предыдущей скорости отправки и затем возобновляет постепенное увеличение скорости отправки с целью достижения равновесия с параллельными сеансами TCP. Если ему не удается восстановить утерянные пакеты с помощью этого механизма быстрого восстановления, то он закрывает окно отправки и повторно входит в режим медленного старта с начальной скоростью 1. Рис. 2 показывает идеализированное поведение TCP Reno.

Целью перехода отправителя в режим предотвращения заторов является тщательное зондирование для нахождения точки образования затора в сети с постепенным увеличением давления потока данных на сеть. В режиме предотвращения заторов дублирующие ACK заставляют отправителя вдвое снизить скорость отправки, попытаться восстановить данные из потерянного пакета и, при успешном выполнении этих задач, продолжить работу в режиме предотвращения заторов, начав с новой скорости отправки.

В устойчивом состоянии алгоритму предотвращения заторов TCP Reno удается идеально избегать перезапуска сеансов с помощью медленного старта, вместо этого он пытается продолжить поток данных с использованием того же самого допущения о вероятной потере из-за затора. Это процесс Аддитивного увеличения (Additive Increase) в рамках окна отправки (на один сегмент каждый интервал RTT) и Мультипликативного уменьшения (Multiplicative Decrease) (посредством двойного сокращения размера) при наступлении затора или сокращенно процесс AIMD.

Алгоритм AIMD протокола TCP Reno имеет обыкновение подвергать буферы сети высокому уровню давления при наличии доступного буферного пространства и менее драматично реагировать в тех случаях, когда буферы в конце концов переполняются и достигают точки отбрасывания пакетов. Управление потоками данных TCP Reno является гораздо более эффективным по сравнению с его предшественником Tahoe, однако по-прежнему существует форма управления обратной связью «бум – спад», которая пытается перевести сеть в перегрузку затора и затем отступает, обеспечивая опорожнение буферов. Невзирая на эти недостатки, Reno является главной опорой Интернета на протяжении уже более чем двух десятилетий и остается эталонным ориентиром, с которым сравниваются другие алгоритмы управления потоками TCP.

Лучше чем Reno

Существует целый ряд причин для ухода от алгоритма управления потоками данных, используемого в Reno. Один из подходов заключается в обеспечении более ровного потока пакетов через сеть и в устранении некоторого «дерганья», присущего TCP Reno. Кроме того, имеется подозрение, что более «чувствительный» механизм управления потоками сможет добиться более высокого результата, чем TCP Reno. Иными словами, другой алгоритм управления потоками данных TCP может добиться лучшего, чем «справедливая доля» при соперничестве с набором параллельных потоков TCP Reno!

Первый из таких подходов, на который мы здесь обратим внимание, реализуется простым изменением. В попытке удвоить давление на другие параллельные сеансы TCP алгоритм AIMD может быть скорректирован при помощи увеличения скорости отправки на более крупную постоянную величину за каждый интервал RTT и посредством уменьшения скорости на менее значительную величину

при потере пакета. Такой подход используется протоколом MulTCP. Например, если скорость отправки увеличивается на 2 единицы MSS за каждый интервал RTT (вместо 1 единицы MSS в TCP Reno) и уменьшается на одну четверть (а не вдвое) при получении дублирующего ACK, то результирующее поведение будет, в приблизительном смысле, напоминать два параллельных сеанса TCP. В рамках сценария справедливого разделения такая форма управления потоком попытается обеспечить выделение этому сеансу двойного объема сетевых ресурсов по сравнению с эквивалентным сеансом TCP Reno.

Еще один вариант такого подхода – это «Высокоскоростной TCP», который увеличивает частоту зондирования в потенциально достижимую пропускную способность, наращивая свою скорость отправки на более крупную величину, сохраняя при этом скорость уменьшения на постоянном уровне. Поэтому вместо увеличения окна перегрузки на 1 за каждый интервал RTT (в случае Reno), высокоскоростной TCP использует расчет скорости увеличения за каждый интервал RTT, который превышает 1 по мере роста окна перегрузки. Этот протокол зондирует на предмет начала потери пакетов с гораздо более высокой частотой, чем любой из протоколов TCP Reno или MulTCP после того, как отправитель начинает работать с большим окном перегрузки и способен ускориться до гораздо более высоких скоростей потока в значительно более короткие интервалы времени.

BIC и его вариант CUBIC используют нелинейную функцию увеличения, а не функцию увеличения с постоянной скоростью. Вместо того, чтобы увеличивать скорость на фиксированную величину каждый интервал RTT, в режиме предотвращения заторов BIC запоминает частоту отправки на начало потери пакетов, и каждый интервал RTT увеличивает скорость на половину разницы между текущей частотой отправки и предполагаемой частотой «бутылочного горлышка». BIC быстро доводит сеанс до пропускной способности узкого места и затем осуществляет более осторожное зондирование после того, как скорость отправки приблизится к пропускной способности бутылочного горлышка (рис. 3). И опять, по сравнению с сеансом потока Reno, BIC должен выдавать превосходящий результат.

Другие алгоритмы управления потоками данных уходят от использования уровня потери пакетов в качестве управляющего фактора и имеют обыкновение чаще колебаться вокруг точки начала формирования очереди в маршрутизаторах вдоль сетевого пути. Такой вид управления обратной связью делает отправителя чувствительным к относительной временной разнице между отправкой пакетов и получением ACK. В качестве примера можно привести TCP, управляемый потоком «парами пакетов», в рамках которого частота отправки увеличивается до тех пор, пока временной интервал между отправляемыми двумя пакетами не станет равен интервалу времени полученных ACK. Если интервал ACK становится больше, то это интерпретируется как начало образования очереди на пути отправки, и происходит уменьшение частоты отправки до тех пор, пока временной интервал ACK опять не станет равным временному интервалу отправки.

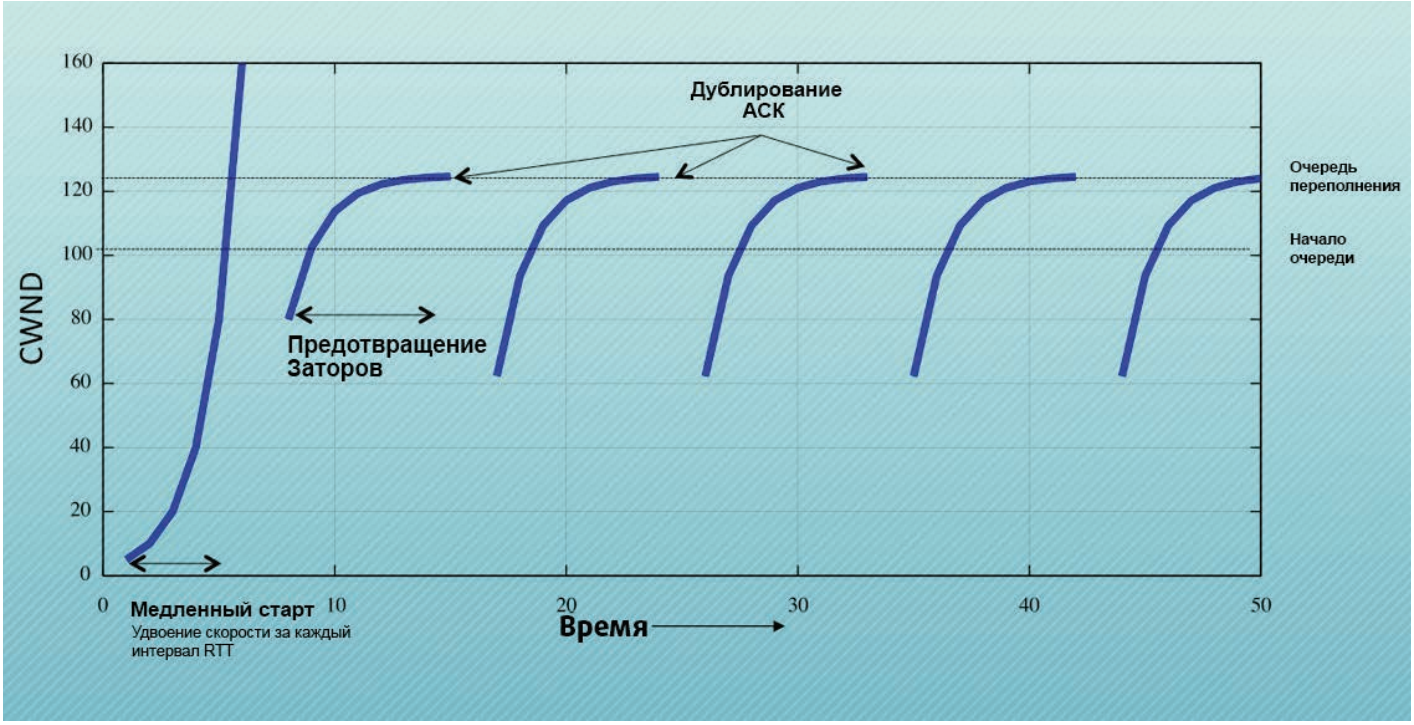
Недавно выпущенные системы Microsoft используют протокол Compound TCP, который объединяет TCP Reno и управление потоками данных на основе задержки. Этот алгоритм пытается измерить объем данных «в полете», удерживаемый в очередях (трафик с более высокой задержкой), и, после потери пакета, алгоритм уменьшает частоту отправки до уровня, находящегося ниже начала роста в очередях согласно оценке отправителя.

Системы Apple Macintosh используют New Reno – вариант управления потоками данных Reno, который усовершенствовал процедуру восстановления потерь Reno, но в остальном остается тем же самым алгоритмом управления AIMD.

Ядра Linux переключились на использование CUBIC – варианта алгоритма BIC, который использует кубическую функцию (а не экспоненциальную) для управления расширением окна.

Существует неустойчивое равновесие между сетью и поведением управления потоками TCP, сконцентрированное вокруг управления буферами очередей в коммутаторах сети. TCP имеет обыкновение интерпретировать потерю пакета как сигнал о том, что «бутылочное

Рис.3. Идеализированное управление потоками данных TCP BIC



горлышко» сетевого пути наполнило свои буферы очередей точки переполнения, и что поддержание той же скорости отправки только ухудшит ситуацию с потерей пакетов, приводя к неэффективности потока данных. Целью уменьшения скорости вдвое, как ответа на потерю в результате затора, являлось уменьшение частоты отправки потока до точки, расположенной ниже пропускной способности узкого места, позволяя, таким образом, очереди опустеть, а медленное увеличение в режиме предотвращения заторов гарантировало, что эта более низкая скорость потока продолжалась достаточно долго для полного «осушения» буферных накопителей очередей. Проблема здесь заключается в том, что очередь, которая никогда не пустеет ниже некоторого минимального уровня, ведет себя точно таким же образом, как линия задержки, объединенная с укороченной очередью. Поэтому уменьшение скорости вдвое и постепенное восстановление предназначены для использования полного размера очередей маршрутизатора и снижения уровня потери пакетов и транспортной неэффективности.

Алгоритмы управления потоками данных TCP, которые модифицируют такое поведение, имеют обыкновение лучше всего работать при использовании в среде, в которой все остальные потоки ведут себя более консервативно. В среде, в которой все параллельно работающие сеансы в рамках перегруженного канала используют AIMD-поведение подобное Reno, единственный сеанс, использующий более агрессивный ответ на потерю пакетов, например, CUBIC, обычно будет оказывать более высокое давление на параллельные Reno-подобные сеансы TCP и получит более крупную долю доступных сетевых ресурсов. Конечно, такая стратегия работает только в том случае, когда используются разнообразные алгоритмы управления потоками данных.

Скрещивая опоры: TCP в UDP

Другие подходы к эволюции сквозных транспортных протоколов еще дальше удалились от обычного TCP и изменили поведение как сервера, так и клиента. Один из способов, который позволяет приложению это сделать, заключается в отказе от использования реализации TCP, предоставленной операционной системой, и в реализации TCP-подобного надежного протокола управления потоками данных в потоке данных приложения и в использовании интерфейса UDP операционной системы для передачи пакетов в сеть и из сети.

Такой подход много лет использовался стриминговым приложением BitTorrent (LEDBAT) и совсем недавно – компанией Google в ее экспериментах с QUIC.

Google QUIC (Quick UDP Internet Connections) использует эмуляцию TCP в UDP. QUIC эмулирует надежный протокол со скользящим окном, используемый TCP внутри потока пакетов UDP. QUIC объединяет функциональность защищенных транспортных сеансов и слотов между обычным интерфейсом API HTTP/2, в качестве интерфейса приложения, и UDP (с использованием порта 443) в качестве сквозного транспортного протокола.

QUIC реализует управление потоками данных TCP CUBIC, а также добавляет к этому алгоритму целый ряд небольших изменений.

Это включает использование Выборочного подтверждения (SACK, Selective Acknowledgement) при его активации получателем, для обнаружения случаев отбрасывания нескольких пакетов в отдельном окне RTT. Кроме того, это включает подход, направленный на разделение механизмов управления заторами TCP от действий по восстановлению данных. Цель здесь заключается в том, чтобы разрешить отправку новых данных в ходе восстановления для поддержания синхронизации TCP ACK. Этот подход, получивший название Forward Acknowledgements with Rate Halving (FACK), предусматривает отправку одного пакета для каждых двух подтверждений ACK, полученных в то время, пока TCP восстанавливается от потери пакетов. Такой алгоритм эффективно уменьшает частоту отправки наполовину в течение одного интервала RTT, но не оставляет отправителя в замороженном состоянии, ожидая половинного опустения объема окна перегрузки перед тем, как продолжить отправку данных. Обычный алгоритм восстановления заставляет отправителя прекратить отправку в течение периода длительностью до одного интервала RTT, тем самым теряя точность неявной синхронизации ACK для сеанса. FACK позволяет отправителю продолжить отправку пакетов в сеть в течение этого периода, стремясь разрешить отправителю поддержание точного ракурса синхронизации ACK. Кроме того, FACK позволяют задать количество блоков SACK, которые определяют отсутствующий сегмент, перед повторной отправкой этого сегмента, обеспечивая для отправителя более высокий уровень контроля над чувствительностью к переупорядочиванию пакетов.

Реализация также включает Tail Loss Probe (TLP) – метод, который реагирует на тайм-аут ожидания подтверждения ACK при помощи повторной отправки «хвостового» сегмента, добиваясь отклика SACK отсутствующих сегментов, которые затем можно исправить с помощью FACK. Кроме того, он поддерживает Forward Retransmission Timeout (F-RTO) для смягчения мнимых случаев, когда отправитель возвращается к медленному старту после тайм-аута ожидания ACK, и Early Retransmit для поддержки случаев, когда дублирующие ACK получены при небольшом окне перегрузки отправителя с целью предотвращения переходов к медленному старту из-за мнимого состояния.

QUIC может сделать то, что невозможно даже после множества небольших изменений TCP, а именно, он способен «играть» с некоторыми фундаментальными механизмами TCP, поскольку не существует унаследованных проблем, когда модифицированный отправитель TCP обменивается данными с обычным, не модифицированным получателем TCP. Например, QUIC использует новый порядковый номер для повторно переданных сегментов, позволяя отправителю проводить различия между ACK для исходного сегмента и ACK для повторно переданного. QUIC также всегда использует шифрование TLS и планирует принять TLS 1.3, когда эта спецификация будет завершена. Он уже внедрил «рукопожатие» с нулевым интервалом RTT.

Google намеревается продолжить развитие, используя Упреждающую коррекцию ошибок (FEC, Forward Error Correction), что позволит получателю исправлять определенные виды потоков пакетов без какой-либо повторной передачи, а также добавляя Мульти-тракт (Multipath), что позволит платформам с несколькими сетевыми интерфейсами (например, мобильным устройствам) разделять нагрузку среди всех активных интерфейсов.

QUIC выполняет оценку пропускной способности как средство для быстрого достижения эффективной скорости отправки. SPDY дополнительно помогает QUIC, осуществляя мультиплексирование нескольких сеансов приложений в рамках единого сеанса сквозного транспортного протокола. Это позволяет избежать издержек старта каждого сеанса TCP, учитывая, что TCP требуется

некоторое время для определения пропускной способности узкого места коммуникационного тракта. Использование UDP также позволяет избежать промежуточных устройств, которые осуществляют перехват трафика с целью глубокого изучения пакетов в потоках TCP, и изменяет их объявленный размер окна с целью осуществления внешнего регулирования скорости потока TCP.

Определенно, это весьма интересный подход. Он избавляется от проблемы обратной совместимости, используя транспортный сеанс через UDP, поэтому в этом случае не существует подлежащих рассмотрению унаследованных ограничений TCP.

Однако существует одна проблема с использованием UDP как заменителя TCP. И хотя общедоступные отчеты Google по этой теме не были опубликованы, она остается источником беспокойства. Проблема касается использования UDP через Транслятор сетевых адресов (NAT, Network Address Translator) и времени привязки по адресу внутри NAT. В TCP Транслятор сетевых адресов получает указания от TCP. Когда NAT «видит» начальный пакет «рукопожатия» TCP «изнутри», он создает временную привязку по адресу и отправляет этот пакет в предназначенное для него место (конечно, вместе с оттранслированным адресом источника). Получение в NAT ответной части «рукопожатия» заставляет NAT подтвердить свою запись привязки и применить ее к последующим пакетам данного потока TCP. NAT удерживает это состояние до тех пор, пока не получит сигнал закрытия обмена или сигнал сброса, который закрывает сеанс TCP, либо пока не истечет таймер простоя. Для TCP Транслятор сетевых адресов пытается удерживать привязку, пока активен сеанс TCP. С точки зрения устройств NAT, протокол UDP другой. В отличие от TCP, в UDP нет информации о статусе потока данных. Поэтому когда NAT создает привязку UDP, он должен удерживать ее в течение определенного периода времени. В данном случае не существует четкого технического стандарта, поэтому реализации могут отличаться друг от друга. Некоторые NAT используют очень короткие таймеры и быстро «отпускают» привязку, что соответствует ожиданиям об использовании UDP в качестве простого протокола для запросов/ответов. Применение UDP в качестве эрзац-протокола фреймирования пакетов для реализации TCP на уровне пользователя требует, чтобы NAT удерживал привязку по адресу UDP в течение более длительных интервалов, соответствующих скрытому сеансу TCP. Некоторые NAT это делают, тогда как другие разрушают привязку даже в том случае, когда имеются активные пакеты UDP, нарушая таким образом скрытый сеанс TCP. QUIC предполагает, что NAT будет удерживать открытой неактивную привязку UDP в течение 30 секунд. Если NAT ведет себя более агрессивно, чем это предположение, то QUIC аварийно переключится на обычный протокол TCP.

Удивительно то, что протокол TCP способен обеспечить эффективное обслуживание вне зависимости от задействованной скорости – десятков бит в секунду или миллиардов бит в секунду. Кроме того, удивительно, что TCP остается эффективным вне зависимости от того, сколько «разговоров» (сеансов) проходит по одному каналу – является ли он единственным или одним из миллионов параллельных «разговоров» TCP.

Все это иллюстрирует уровень компромисса в современной среде между сквозными протоколами и промежуточным ПО сети. Сеансы TCP модифицируются активными промежуточными устройствами, которые пытаются управлять скоростью потока данных TCP при помощи активного изменения размеров окон внутри сеанса TCP, сводя на нет некоторые усилия этого сеанса по оптимизации своей скорости потока. TCP в UDP передает управление потоком данных TCP приложению и скрывает параметры потока TCP от сети. Однако сеансы UDP уязвимы к прерыванию из-за вмешательства NAT, поскольку некоторые NAT предполагают, что UDP используется только для микро-сеансов, а длительное время удерживаемые сеансы UDP представляют собой некоторый вид аномального поведения, которое должно быть отфильтровано при помощи удаления привязки порта UDP в NAT.

Конец сквозных протоколов?

И куда теперь?

То, что технология сквозных протоколов не является статичной и не костенеет, весьма бодрит. Наше понимание того, каким образом можно сделать сквозные протоколы работоспособными и эффективными, постоянно эволюционирует путем незначительных, но важных изменений.

Удивительно то, что протокол TCP способен обеспечить эффективное обслуживание вне зависимости от задействованной скорости – десятков бит в секунду или миллиардов бит в секунду. Кроме того, удивительно, что TCP остается эффективным вне зависимости от того, сколько «разговоров» (сеансов) проходит по одному каналу – является ли он единственным или одним из миллионов параллельных «разговоров» TCP. Но что на самом деле удивительно, так это сама технология, сегодня развернутая в миллиардах устройств и остающаяся гибкой и адаптируемой. И мы все еще можем сделать ее лучше.

Сквозная эра определенно еще не закончилась!

Источник: Transport Protocols, http://www.potaroo.net/ispcol/2015-10/e2eprotocols.html

YANG и NETCONF/RESTCONF получают широкое развитие в отрасли

Махеш Джетанандани (Mahesh Jethanandani) и Бенуа Клэз (Benoît Claise)

За последнюю пару лет протокол NETCONF и язык YANG получили широкое развитие в сетевой индустрии. Они перешли от этапа определения в этап реализации. В рамках IETF количество разрабатываемых моделей YANG растет невероятными темпами. Однако быстрый рост числа моделей YANG не обходится без проблем, главная из которых заключается в их координации. Хотя все модели великолепно определяют, каким образом можно конфигурировать или отслеживать конкретные функции, они также должны взаимодействовать с моделями, разрабатываемыми в IETF и в других SDO.

В 2003 году в рамках RFC 3535 «Overview of the 2002 IAB Network Management Workshop» («Обзор состоявшегося в 2002 году семинара IAB по сетевому управлению») были задокументированы результаты диалога между сетевыми операторами и разработчиками протоколов о концентрации усилий IETF на дальнейшей работе по управлению сетями. Семинар идентифицировал 14 требований операторов и определил «удобство использования» как ключевое требование для любой новой системы управления сетями. Такое удобство использования включает способность управлять сетью, а не просто устройством в сети, и устанавливает, что должно существовать четкое разграничение между конфигурационной, рабочей и статистической информацией устройства. Кроме того, эти требования включают способность развернуть конфигурацию, проверить ее правильность перед практическим задействованием и выполнить откат назад к предыдущей конфигурации в случае неудачи.

Эти 14 требований операторов привели к созданию в том году рабочей группы NETCONF, рабочей группы NETMOD в 2008 году и к развитию базовых моделей данных для управления сетями. Результатом этих усилий стали появившиеся в 2011 году RFC 6241, 6242, 6243 и 6244 (переработанные и созданные на основе RFC 4741, 4742, 4743 и 4744 соответственно) для основанного на XML протокола NETCONF и связанные с ними RFC 6020 и 6021 для языка моделирования данных YANG, которые появились в 2010 году.

За последнюю пару лет протокол NETCONF и язык YANG получили широкое развитие в сетевой индустрии. Они перешли от этапа определения на этап реализации. В рамках IETF количество разрабатываемых моделей YANG растет невероятными темпами. Новые модели YANG разрабатываются в области систем эксплуатации и управления (OPS, Operations and Management), а также в таких областях, как Маршрутизация (RTG), Интернет (INT), Транспорт (TSV) и Безопасность (SEC). Однако наиболее впечатляющие примеры принятия моделей YANG происходят в рамках проекта с отрытым исходным кодом Open Daylight, где выход Lithium сопровождался публикацией более чем 480 моделей YANG.

Другие организации-разработчики стандартов (SDO) также инициировали проекты по разработке моделей YANG. Например, организация Metro Ethernet Forum стала первопроходцем в разработке моделей YANG для управления производительностью (PM, Performance Management) и управления ошибками (FM, Fault Management) в рамках Service OAM (SOAM); в настоящее время эта организация

Одним из результатов популярности YANG является то, что операторы, желающие разработать свой собственный протокол управления, используют YANG в качестве языка моделирования данных. Сюда относится CoMI, который определяет интерфейс управления для ограниченных устройств. Даже среди существующих протоколов NETCONF и RESTCONF имеются разные кодировки (например, XML и JSON) для моделей YANG.

работает над моделями YANG на уровне сервисов. Кроме того, Институт инженеров по электротехнике и электронике (IEEE, Institute of Electrical and Electronics Engineers) утвердил проект для моделей 802.1x и 802.1q, проявляя интерес к разработке модели 802.3. Аналогичным образом сектор по стандартизации телекоммуникаций Международного совета по телекоммуникациям (ITU-T, International Telecommunication Union Telecommunication Standardization Sector) заинтересовался разработкой модели G.8032. Информацию обо всех моделях организаций-разработчиков стандартов можно найти в GitHub (https://github.com/YangModels/yang).

Быстрый рост числа моделей YANG не обходится без проблем, главная из которых заключается в их координации. Хотя все модели великолепно определяют, каким образом можно конфигурировать или отслеживать конкретные функции, они также должны взаимодействовать с моделями, разрабатываемыми в IETF и в других SDO. Первый практический пример координации происходит в области маршрутизации и проводится организацией Routing Area YANG Coordination Forum. Координация работ по разработке YANG в рамках IETF и в других SDO попадает в сферу ответственности директора Области эксплуатации и управления (OPS, Operations and Management) Бенуа Клэза (Benoît Claise), который работает в сотрудничестве с группой YANG Model Coordination Team.

Рабочие группы IETF, работающие над аспектами развития модели YANG, включают:

- LIME (модели YANG OAM)
- L3SM (модель YANG для сервиса L3VPN)
- SUPA (модели YANG последовательной политики)
- I2NSF (модели YANG, связанные с обеспечением безопасности)

Для того, чтобы помочь в разработке моделей YANG, с докторами YANG можно связаться как по электронной почте, так и в течение недели, когда происходят совещания IETF в рамках сеансов редактирования/рекомендаций YANG. Кроме того, доступны несколько инструментов для разработки и компиляции моделей YANG (см. http://trac.tools.ietf.org/area/ops/trac/wiki/YangModelCoordGroup для получения полного списка).

Вероятно, наиболее важным из таких инструментов является pyang – средство компиляции YANG на основе языка Python, которое осуществляет синтаксическую проверку и позволяет генерировать такие выходные форматы, как UML, модель на основе дерева, YIN и т.д. Эти инструменты должны запускаться с опцией IETF (--ietf) для того, чтобы проверять правила YANG, определенные в RFC 6087. До сих пор многие модели YANG не могут быть корректно скомпилированы. Онлайновый графический эквивалент инструмента pyang можно найти по адресу: http://yangvalidator.com; этот инструмент берет файл YANG или проект/RFC IETF, извлекает модель YANG и затем проверяет ее правильность.

Благодаря опыту, накопленному при разработке и внедрении некоторых моделей YANG, рабочая группа NETMOD получила отзывы и комментарии по YANG 1.0. На основе этой информации в настоящее время окончательно дорабатывается версия YANG v1.1. Эта новая версия является отладочной версией языка YANG; она устраняет двусмысленности и дефекты, которые содержались в исходной спецификации.

После завершения специфицирования NETCONF и YANG операторы могут начать их использование для целей конфигурирования и мониторинга. Однако некоторые операторы уже начали использовать закрытые (патентованные) интерфейсы API REST, предоставленные другими производителями, для управления своими сетями. RESTCONF представляет собой REST-подобный протокол, работающий через HTTP, который используется для доступа к данным, определенным в YANG. REST-подобный API не предназначен для замены NETCONF, он скорее предоставляет упрощенный интерфейс, тем самым удовлетворяя потребности разработчиков приложений. По этой причине рабочая группа NETCONF решила добавить в концепцию поддержку протокола RESTCONF. Протокол RESTCONF поддерживает два формата кодирования: XML и JSON.

Хотя эта возможность часто остается незамеченной, устройства могут также посылать уведомления, определенные в модели YANG. Недавно принятая концепция NETCONF включает обновление для Уведомлений о событиях (Event Notifications) NETCONF и развитие механизма «уведомления по подписке» (subscription-and-push), который позволяет клиентским приложениям запрашивать уведомления об изменениях в хранилище данных. Эти возможности откроют NETCONF для мира телеметрии, посылая данные приложениям систем сетевого управления (NMS).

Одним из результатов популярности YANG является то, что операторы, желающие разработать свой собственный протокол управления, используют YANG в качестве языка моделирования данных. Сюда относится CoMI, который определяет интерфейс управления для ограниченных устройств. Даже среди существующих протоколов NETCONF и RESTCONF имеются разные кодировки (например, XML и JSON) для моделей YANG.

В конечном счете, главное значение имеют модели данных. В рамках индустрии существует явная потребность в стандартных моделях данных, которые помогли бы облегчить управление и, что более точно, программируемость сетей, объединяющих аппаратные средства разных производителей. YANG четко позиционирует себя как тот самый язык моделирования данных для этих стандартных моделей. Если мы хотим, чтобы все модели YANG безболезненно работали друг с другом, то эта задача по координации моделей возложена на нас – участников IETF.

Источник: YANG and NETCONF/RESTCONF Gain Traction in the Industry, http://www.internetsociety.org/publications/ietf-journal-november-2015/yang-netconf

Протокол OSPF: состояние канала

Эси Линдем (Acee Lindem)

Рабочая группа OSPF – одна из старейших рабочих групп IETF. После более чем двух десятилетий деятельности можно было бы подумать, что она перешла в режим поддержки и обслуживания. Однако в действительности мы находимся на распутье, в поиске пути стандартизации гибких механизмов расширения на основе метода тип-длина-запись TLV. Благодаря этим механизмам протокол OSPF может использоваться для поддержки некоторых новых, захватывающих приложений, таких как сегментная маршрутизация.

Протокол Open Shortest Path First (OSPF) является одним из двух внутренних протоколов маршрутизации (IGP, Interior Gateway Protocol), которые стандартизируются в рамках IETF. Этот протокол широко используется как в корпоративных сетях, так и в сетях провайдеров услуг, он является предпочтительным протоколом управления в оптических сетях. Рабочая группа OSPF (OSPF Working Group) – это одна из старейших рабочих групп IETF. После более чем двух десятилетий деятельности можно было бы подумать, что она перешла в режим поддержки и обслуживания. Однако в действительности мы находимся на перекрестке, поскольку необходимо стандартизировать гибкие механизмы расширения на основе метода тип-длина-запись (type-length-value, TLV).

Для OSPFv2 мы решили оставить базовый протокол в неизменном виде и инициировать отдельные сообщения о состоянии канала (Link State Advertisements, LSA), чтобы объявлять TLV для таких приложений, как сегментная маршрутизация (segment routing) и максимально избыточные деревья (maximally redundant trees). Проект документа «OSPFv2 Prefix/Link Attributes» (атрибуты канала/префикса OSPFv2) прошел последний звонок рабочей группы, после чего была запрошена его публикация. В настоящее время существует не менее пяти известных реализаций. Однако у этого подхода есть один недостаток, который заключается в том, что атрибуты дополнительных приложений объявляются независимо от базовой топологии OSPF и IP-достижимости. Поэтому приложения должны увязывать базовые LSA с LSA атрибутов.

Для OSPFv3 предлагается более амбициозный подход, в рамках которого даже базовые LSA заменяются на LSA, полностью основанные на методе TLV. Благодаря такой кодировке вся информация для данного префикса или канала может быть объявлена в одном и том же сообщении LSA, что значительно упрощает реализацию и уменьшает накладные расходы. Эти механизмы превращают версию протокола OSPFv3 в идеального кандидата на роль внутреннего протокола маршрутизации (IGP) следующего поколения, поскольку OSPF имеет заметные преимущества по сравнению с другими протоколами IGP: с его помощью можно разделять информацию и объявлять ее с помощью нескольких LSA – в отличие от монолитных блоков протокольных данных (PDU). При использовании OSPFv3, если в сети происходит изменение топологии или достижимости, то требуется повторно объявлять только те LSA, которые затронуты этими изменениями. Эти механизмы определяются в проекте расширенных сообщений LSA протокола OSPFv3. Обсуждения, редактирование и анализ прошли хорошо, и в настоящее время мы ожидаем реализаций. В основном реализации мешают два препятствия. Первое препятствие заключается в том, что OSPFv3 отнюдь не так широко распространен, как OSPFv2, и в результате существует меньше стимулов для его расширения. Второе препятствие – это сложность, добавляемая механизмами обратной совместимости протокола.

Благодаря этим механизмам расширения базовых сообщений LSA протокол OSPF используется для поддержки некоторых новых, захватывающих приложений. Наиболее важным из них, вероятно, является Сегментная Маршрутизация (Segment Routing), поскольку она пользуется существующей технологией передачи данных Multi-Protocol Label Switching (MPLS) без задействования каких-либо специфических протоколов управления MPLS (т.е. без LDP или RSVP). Кроме того, Сегментная Маршрутизация упрощает управление трафиком и лучше поддерживает IPFRR (IP Fast Reroute), поскольку позволяет направлять пакеты через любые граничащие системы (adjacency).

В число других приложений, использующих кодировки OSPF на основе метода TLV, входят алгоритм IPFRR, известный как Maximally Redundant Trees (MRT), поддержка метода мультикастинга BIER (Bit-Indexed Egress Replication) и поддержка дополнительных метрик OSPF в спутниковых сетях.

Программируемость на основе моделей (MDP, Model-Driven Programmability) является обычным требованием для многих рабочих групп IETF. В рамках рабочей группы OSPF мы сформировали проектную группу с участием специалистов от нескольких производителей, которая еженедельно собиралась на протяжении почти целого года для определения общей модели YANG для OSPF. Мы достигли консенсуса, даже несмотря на значительные различия в конфигурации производителей. Одним из ключевых решений было согласие на принятие модели, ориентированной на виртуальную маршрутизацию и передачу (VRF, Virtual Routing and Forwarding), а не модели, ориентированной на протоколы. В рамках VRF-ориентированной модели конфигурации протокола для от-

дельных VRF-экземпляров содержатся внутри этих VRF, а не консолидируются внутри отдельного экземпляра протокола маршрутизации. Еще одно ключевое иерархическое решение, которое окажет воздействие на многочисленные модели IETF, заключается в том, следует ли принимать предложение Open Config о группировании (информации) рабочего состояния на том же уровне, что и конфигурация. В текущей версии модели OSPF существуют отдельные иерархии YANG конфигурации и рабочего состояния. Однако ко времени публикации этой статьи модель, по всей вероятности, будет пересмотрена. Поскольку это решение влияет на многие модели YANG, соответствующее обсуждение также происходит в рамках рабочей группы NETMOD.

И наконец, наша рабочая группа также рассматривает способы для масштабирования протокола OSPF за пределы своих практических ограничений. Одно из таких предложений – это возможность определения так называемой топологически-прозрачной зоны (TTZ, Topology Transparent Zone), позволяющее выделить произвольную часть сети OSPF в качестве полной «сетки» соединений между маршрутизаторами, ограждающими эту абстрагированную топологию. Еще одно предложение касается соседей-«заглушек» OSPF (stub neighbor), призванное оптимизировать лавинное распространение LSA в звездообразных (веерных) топологиях.

Источник: Open Shortest Path First: The State of The Link State, http://www.internetsociety.org/publications/ietf-journal-july-2015/open-shortest-path-first

«Слишком много неизвестных терминов?»

Type-length-value (TLV)

TLV (тип-длина-значение) — широко распространённый метод записи коротких данных в компьютерных файлах и телекоммуникационных протоколах.

Метод определяет простую двоичную структуру из трёх полей: тег, длина данных и собственно данные. Первые два поля имеют фиксированную длину (обычно один или два октета на поле), длина третьего поля определяется значением второго поля (значение указывается в байтах). Тег является идентификатором данных, определяя их назначение.

Источник: https://ru.wikipedia.org/wiki/Tag-length-value

Сегментная маршрутизация

Сегментная маршрутизация (СМ) основана на парадигме маршрутизации от источника. Узел направляет пакет через упорядоченный список инструкций, называемых сегментами. Сегментом может являться любая инструкция, топологическая или основанная на услугах. Семантика сегмента может быть локальной по отношению к конкретному узлу или глобальной, распространяющейся на определенную область.

СМ позволяет направить поток данных через любой топологический путь или цепь услуг, и в то же время состояние каждого потока хранится только в узле, входном в СМ-область.

СМ можно непосредственно применить к архитектуре MPLS. Сегмент в этом случае является меткой MPLS. Упорядоченный список сегментов представляет собой стэк меток. Обрабатываемый сегмент находится на верхушке стэка. После окончания обработки сегмента соответствующая метка удаляется из стэка.

Источник: http://www.segment-routing.net/

Bit Index Explicit Replication (BIER)

Архитектура BIER предназначена для передачи данных мультикаста. Она позволяет оптимизировать передачу мультикаст-пакетов через «область мультикаста». Однако она не требует специального протокола для построения распределительных деревьев мультикаста и не требует от промежуточных узлов сохранения состояния. Когда мультикаст-пакет передается в область, входной маршрутизатор определяет выходные маршрутизаторы, которым необходимо отправить пакет. Входной маршрутизатор инкапсулирует пакет в заголовок BIER. Этот заголовок содержит строку битов, в которой каждый бит соответствует определенному выходному маршрутизатору. Исключение требования хранения состояния для каждого потока и необходимости отдельных протоколов для построения дерева приводит к существенному упрощению системы.

Источник: https://tools.ietf.org/html/draft-ietf-bier-architecture-03

Virtual Routing and Forwarding (VRF)

В компьютерных IP-сетях под VRF понимают технологию, которая позволяет нескольким экземплярам таблицы маршрутизации сосуществовать на одном и том же маршрутизаторе. Поскольку эти экземпляры независимы, те же самые или пересекающиеся блоки IP-адресов могут использоваться без возникновения конфликтов.

Источник: https://en.wikipedia.org/wiki/Virtual_routing_and_forwarding

Эволюция IANA: на финишной кривой

Андрей Робачевский

Эта третья и последняя статья из серии, рассказывающей об истории и эволюции IANA, включая текущую стадию передачи надзорной роли правительства США в руки международного сообщества. В этой статье рассматривается заключительный этап работы международного сообщества над проектом предложения, завершившийся 10 марта 2016. В этот день Координирующая Группа ICG объявила, что она одобрила общий согласованный план передачи IANA и направила окончательное предложение на рассмотрение NTIA через Правление ICANN.

В предыдущей статье на тему эволюции IANA мы говорили о процессе и основных моментах предложений по передаче координирующей роли NTIA в осуществлении функций IANA в руки мирового сообщества. Предложений было подано три, от каждого «операционного сообщества» – сообщества, непосредственно пользующегося услугами IANA. От сообщества протоколов, сообщества номерных ресурсов и сообщества имен.

Каждое из сообществ независимо определило процесс разработки предложения. Так, сообщество протоколов в лице IETF создало рабочую группу IANAPLAN, которая провела работу, руководствуясь процессом разработки стандартов IETF. Сообщество номерных ресурсов в лице РИРов выбрало 15 представителей в специальную группу по разработке Консолидированного предложения РИР по координирующей роли – CRISP. Для разработки предложения для имен была создана группа CWG – Сквозная рабочая группа сообщества имен (также получившая название CWG-Stewardship, CWG-Координирующая роль), сформированная из представителей поддерживающих организаций и комитетов, имеющих отношение к доменным именам.

В изначально установленный срок – 15 января 2015 года – были поданы два предложения – от протоколов и номерных ресурсов. Предложение от имен поступило пять месяцев спустя – 25 июня 2015 года. Наконец, к концу октября того же года группа ICG (Координационная группа по передаче координирующей роли в исполнении функций IANA) завершила работу над консолидированным предложением. Однако уже к концу лета было очевидно, что изначальный план нереалистичен, и в сентябре 2015 года NTIA продлило контракт еще на один год, установив очередную целевую дату – сентябрь 2016 года.

Дело в том, что хотя консолидированное предложение было формально готово, оно не могло быть отправлено в NTIA через Правление ICANN из-за существенной неразрешенной зависимости. Часть предложения, относящегося к именам, зависела от реализации механизмов подотчетности ICANN. Перед отправкой заключительного предложения группа ICG запросила подтверждение от CWG в том, что ее требования подотчетности выполнены.

Как я писал в предыдущей статье, над этой проблемой работала другая группа – Сквозная рабочая группа сообщества по повышению подотчетности ICANN (Cross-Community WG Group – Accountability, CCWG). В ее состав вошли 28 членов, номинированных поддерживающими организациями и комитетами ICANN, а также 169 участников. Работать группа начала еще в 2014 году, но год спустя конца этому проекту еще не была видно.

Как же получилось, что задача передачи ограниченной и во многом формальной роли NTIA переросла в реформу ICANN?

Подотчетность

Дело здесь в основном связано с функцией IANA по координации внесения изменений в корневую зону DNS. Задача эта не такая уж сложная, но политически очень значимая. Можно сказать, что собственно эта функция и является основным мотивирующим фактором в процессе передачи роли NTIA, призванным во многом устранить асимметричную роль правительства США в управлении корнем DNS.

Задача выбора модели управления также усложнялась тем, что в отличие от сообщества номерных ресурсов и параметров протоколов, не существует единого сообщества имен. Вместо этого имеется сообщество администраторов национальных доменов и сообщество общих доменных имен. Поскольку управление корневой зоны имеет ключевое значение для работы всей глобальной системы DNS, необходимо учитывать интересы и интернет-пользователей, и правительств государств.

Наконец, и это, пожалуй, самое важное, все эти сообщества представлены организациями поддержки – ccNSO, gNSO – и различными комитетами – ALAC (At-Large Advisory Committee) и GAC (Government Advisory Committee). Все эти структуры являются частью самой ICANN, то есть для имен не существует внешней по отношению к ICANN организации, или организаций, которые представляли бы интересы различных сообществ имен.

В этом еще одно существенное отличие от номерных ресурсов и параметров протоколов, для которых эту представительскую роль выполняют РИРы и IETF соответственно. И если эти сообщества выбрали договорные отношения с оператором IANA как наиболее

простую форму установления подотчетности, в случае имен эта возможность отсутствовала: получалось, что ICANN заключала договор с ICANN и являлась подотчетной самой себе.

Как я писал в предыдущей статье, для разрешения этой загвоздки было предложено создание нового отдельного юридического лица «IANA после передачи» (Post Transition IANA, PTI) в форме некоммерческой корпорации (точнее — корпорации по обеспечению общественных интересов, зарегистрированной в штате Калифорния). В соответствии с планом, эта корпорация должна была являться филиалом или дочерней компанией ICANN, а ICANN — ее полным и единственным собственником.

Хотя заключение договора с отдельной организацией-провайдером услуг IANA вроде бы передавало контроль за этой функцией сообществу, неувязка все же оставалась. В конце концов, ICANN заключала договор со своей же дочерней организацией, что хотя безусловно обеспечивало лучшее разделение функций определения политик от администрирования реестров и корневой зоны в соответствии с этими политиками и прозрачность этих отношений, по-прежнему оставляло значительную часть контроля в руках ICANN и ее Правления. Например, ICANN могла не утвердить или недопустимо урезать бюджет PTI, также, ICANN могла вставлять палки в колеса, зайди речь о передаче функций IANA от PTI к какой-либо сторонней организации.

Поэтому вопрос общей подотчетности ICANN оказался неотделимым от вопроса передачи координирующей роли.

CCWG – основные вопросы

Окончательное предложение от сообщества имен, разработанное группой CWG и интегрированное в совместное предложение ICG, содержало семь требований подотчетности ICANN. Эти вопросы и стали основой работы и рекомендаций CCWG (https://www.icann.org/en/system/files/files/ccwg-accountability-supp-proposal-work-stream-1-recs-23feb16-en.pdf). Давайте кратко рассмотрим каждое из них.

1. Бюджет ICANN и IANA

Основное требование заключается в предоставлении возможности сообществу утверждать или отклонять бюджет ICANN после его утверждения Правлением, но до вступления бюджета в силу. Основной целью этого требования является прозрачность расходов на работу IANA (PTI), особенно в случае совместного использования ресурсов, а также предотвращение возможности неадекватного финансирования оператора услуг IANA и, как следствие, ухудшения качества этих услуг.

Рекомендация №4 предложения CCWG позволяет сообществу отклонить стратегический и операционные планы ICANN, а также годовой бюджет ICANN и IANA.

2. Механизмы наделения сообщества большими полномочиями

В основном, здесь имеются в виду полномочия по отношению к Правлению ICANN – возможность отзыва отдельных членов или роспуск всего Правления, возможность осуществления надзора над ключевыми решениями Правления, в частности, относительно рекомендаций по результатам проверки функций IANA (IFR) и бюджета ICANN. Также сюда была включена возможность утверждения изменений в основной устав ICANN.

Рекомендация №4 предложения CCWG наделяет сообщество так называемыми семью дополнительными полномочиями (см. рис 1). Например, сообщество отвечает за утверждение изменений в Основной устав, а также уполномочено как отозвать отдельного члена Правления, так и расформировать Правление целиком.

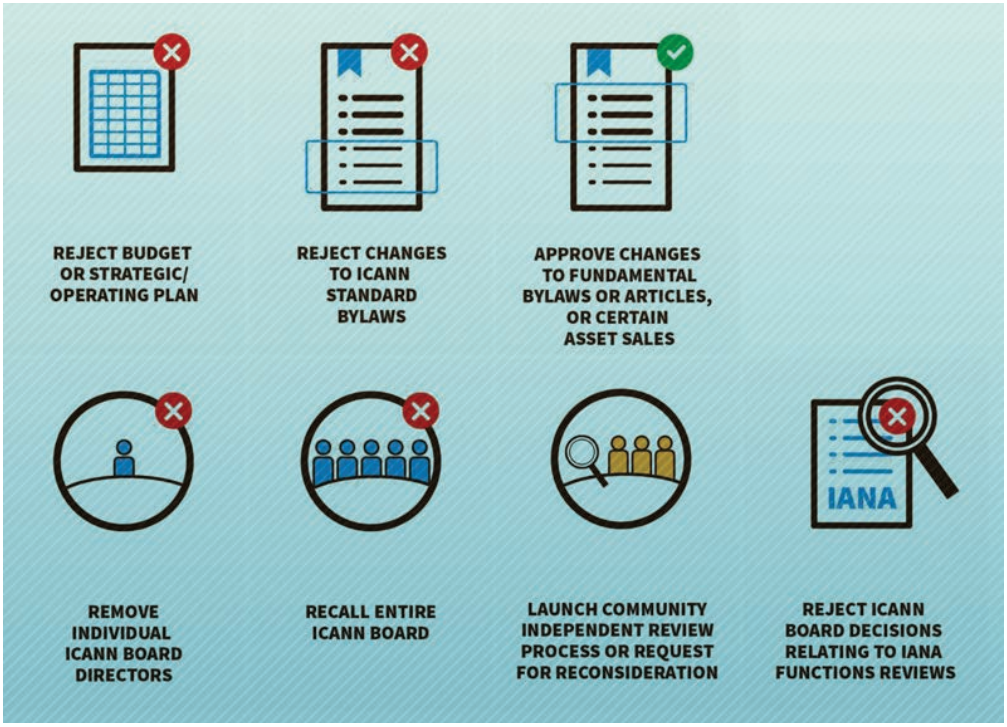
3. Проверка функционирования IANA (IFR) и процесс отделения

Процесс проверки функционирования IANA помимо рекомендаций по улучшению может указать на необходимость начала процесса отделения – поиска нового «дома» для IANA. Этот процесс не имеет предопределенного разрешения и включается только тогда, когда остальные механизмы разрешения конфликтов ни к чему не привели. Управлением этим процессом будет заниматься созданная для этой цели специальная группа, и ее рекомендации могут распространяться от «дальнейших действий не требуется» до поиска нового оператора функций IANA или реорганизации PTI. Для того, чтобы проверки имели силу, они должны быть отражены в Уставе ICANN. Это является сутью рекомендации №9 CCWG.

4. Постоянный комитет потребителей

Создание Комитета, уполномоченного эскалировать неразрешенные проблемы организациям поддержки ccNSO и gNSO. Эти организации, в свою очередь, должны получить

Рис.1. Семь новых полномочий. Источник — Проект CCWG-Accountability (https://www.icann.org/en/system/files/files/ccwg-accountability-supp-proposal-work-stream-1-recs-23feb16-en.pdf)



Что за реестры обслуживает IANA?

В процессе обсуждения различных предложений по передаче координирующей роли NTIA в руки мирового сообщества много говорилось о функциях IANA, о реестрах, которые она обслуживает. Но что собой представляют эти реестры?

Реестры различны по формату и содержанию. Объединяет их одно – они все призваны обеспечить уникальность присвоенных имен, адресов, номеров и других идентификаторов. Все реестры делятся на три основные категории: реестры параметров протоколов, реестры IP-адресов и реестры имен.

Реестр параметров протоколов определяет уникальные значения определенных полей протокола и их семантику. Например, для протокола маршрутизации BGP в IANA существует несколько реестров: BGP Message Types, BGP OPEN Optional Parameter Types, BGP Path Attributes и еще пара десятков.

Например, часть реестра BGP Path Attributes показана в таблице 1.

Ссылка определяет спецификацию RFC, в которой определено поле, и его значения, а также даны инструкции IANA по обслуживанию (и во многих случаях – созданию) реестра. Документ RFC, который создает реестр, также определяет «политику», или правила, по которым новые значения могут быть присвоены. Эта относящаяся к реестрам информация определена в секции RFC «IANA Considerations», или «Соображения для IANA».

Таблица 1. Реестр BGP Path Attributes

Значение	Код/имя	Ссылка
1	ORIGIN	RFC4271
2	AS_PATH	RFC4271
3	NEXT_HOP	RFC4271
4	MULTI_EXIT_DISC	RFC4271
5	LOCAL_PREF	RFC4271
6	ATOMIC_GGREGATE	RFC4271
7	AGGREGATOR	RFC4271
8	COMMUNITY	RFC1997
...	...	...

Для адресов реестров меньше, но принцип такой же. Основная структура определяется IETF. Так, например, IETF через RFC4291 определяет адресное пространство IPv62000::/3 как «Global Unicast», предназначенное для распределения между сетями, подключенными к глобальному Интернету через региональные регистратуры (РИРы). Большая часть оставшегося пространства IPv6 зарезервирована IETF для будущего использования.

Сам же реестр «Global Unicast» выглядит примерно, как показано в таблице 2.

Таблица 2. Реестр «Global Unicast»

Префикс	Десигнация	Дата	Whois	Статус
2001:0e00::/23	APNIC	2003-01-01	whois.apnic.net	ALLOCATED
2001:1200::/23	LACNIC	2002-11-01	whois.lacnic.net	ALLOCATED
2001:1400::/23	RIPE NCC	2003-02-01	whois.ripe.net	ALLOCATED
2001:1600::/23	RIPE NCC	2003-07-01	whois.ripe.net	ALLOCATED
2001:1800::/23	ARIN	2003-04-01	whois.arin.net	ALLOCATED
...	...	...	...	...

Наконец, когда речь идет о доменных именах, то наиболее часто упоминаемый реестр – это корневая база данных (http://www.iana.org/domains/root/db), содержащая информацию, относящуюся к делегированию того или иного домена: тип домена (национальный, общего назначения), организацию-спонсор, административный и технический контакты, а также серверы имен, обслуживающих домен.

Но IANA также обслуживает и несколько других, связанных с доменными именами реестров. Например, реестр IDN, представляющий коллекцию так называемых IDN-таблиц, которые представляют собой разрешенные кодовые точки (буквы), допустимые для регистрации многоязычных доменных имен в соответствующих реестрах.

полномочия, позволяющие им эффективно решать указанные проблемы.

Рекомендация №3 предложения CCWG требует включения этой структуры в Основной устав ICANN.

5. Процесс обжалования

Суть этого требования заключается в предоставлении возможности обжалования проблем, не получивших удовлетворительного разрешения, например, посредством эскалации через Постоянный комитет потребителей. Эта возможность может быть реализована путем создания Независимой группы рассмотрения (Independent Review Panel, IRP). Рекомендация №7 CCWG предлагает соответствующие изменения в Основной устав и содержит более детальную структуру процесса.

6. Структура управления PTI

В своем предложении группа CWG-Stewardship определила основные моменты структуры PTI. В частности, определено, что юридически PTI является дочерней организацией ICANN. Также были задокументированы основные требования к Правлению PTI. Для закрепления этих решений, они должны быть отражены в Уставе ICANN.

Хотя в предложении группы CCWG-Accountability отсутствуют конкретные статьи Устава, оно рекомендует (Рекомендация №3), что эти статьи должны войти в Основной устав.

7. Основной устав

Все перечисленные предложения должны быть отображены в Уставе ICANN. Соответствующие статьи Устава формируют так называемый Основной устав, внесение изменений в который требует утверждения сообществом и в результате имеет более высокий порог. В настоящий момент отсутствует требование проведения публичных консультаций с сообществом – и любые изменения могут быть внесены в Устав 2/3 голосов Правления.

Как мы уже видели, Основной устав предлагается в Рекомендации №3.

Интеллектуальные права

Другим важным моментом, который остался неразрешенным в консолидированном предложении ICG, был вопрос, относящийся к интеллектуальным правам, – а именно к торговому знаку IANA и домену iana.org.

В предыдущей статье я отметил,что предложение сообщества номерных ресурсов (группа CRISP) содержало четкие требования относительно независимости держателя этих ресурсов, который, очевидно, не мог являться оператором IANA. По мнению группы CRISP, если ICANN продолжала бы быть держателем этих ресурсов, это могло существенно затруднить возможность перезаключения контракта и перехода к другому оператору – ключевой элемент решения вопроса надзора.

План CRISP предлагал передать эти ресурсы в руки независимой организации, а IETF Trust (http://trustee.ietf.org/) рассматривался в качестве возможного кандидата.

IETF согласился с этим предложением и Правление IETF Trust выступило с заявлением о готовности Trust принять на себя эти обязательства. Сообщество имен пребывало в нерешительности относительно этого вопроса и формально не высказалось ни в пользу, ни против такого решения. Вместо этого юридической компании Sidley Austin LLP было поручено провести анализ нескольких сценариев решения вопроса с различными держателями интеллектуальной собственности.

Юристы Sidley рассмотрели три сценария: ICANN продолжает являться держателем этих ресурсов, PTI становится держателем интеллектуальной собственности и, наконец, ресурсы передаются независимой организации, такой как, например, IETF Trust.

Однако после непродолжительного обсуждения этих предложений стало очевидно, что первые два сценария неприемлемы, так как противоречат принципам управления интеллектуальной собственностью IANA, разработанным сообществом номерных ресурсов (CRISP).

Вопрос повис в воздухе.

Между тем, IETF Trust представил схему, как управление интеллектуальной собственностью может быть осуществлено путем отдельных договоров с операционными сообществами. Эта рамочная схема была призвана разрешить сомнения относительно возможности одного из сообществ - IETF – узурпировать права. И случилось чудо – было принято совместное (между представителями операционных сообществ) прагматичное решение взять за основу IETF Trust как держателя собственности – и не заниматься разработкой более сложных и теоретически более элегантных решений, как, например, создание новой организации.

Таким образом, этот элемент перешел в разряд вещей, относительно которых существует принципиальная договоренность, и реализация которых должна произойти до окончания контракта NTIA.

Зри в корень

Вопрос управления корневой зоной DNS, разумеется, является одним из ключевых в процессе передачи координирующей роли NTIA. Здесь можно выделить три элемента:

- Утверждение изменений в корневую зону, в настоящий момент выполняемое NTIA. В ходе обсуждений группой CWG было решено полностью исключить этот шаг. Проще решения трудно было придумать.

- Утверждение существенных архитектурных и операционных изменений в систему обслуживания корневой зоны, в настоящий момент также выполняемое NTIA. Примером такого изменения является внедрение DNSSEC в корне. Эта функция будет передана Правлению ICANN, действующему по рекомендации специального комитета с достаточно широким представительством.

- Договор на обслуживание корневой зоны в настоящее время существует в виде Кооперативного соглашения между NTIA и Verisign. Об этом договоре я писал в предыдущей статье. Будущее этого договора неизвестно, но ясно, что если он будет продолжать существовать, он должен быть модифицирован для исключения требования утверждения изменений. Также в любом случае должен быть заключен дополнительный договор между ICANN/PTI и Verisign. Этот вопрос относится к фазе реализации.

Последний рывок

10 Марта 2016. «Координирующая Группа ICG объявляет сегодня, что она одобрила общий согласованный план передачи IANA и направила окончательное предложение для передачи NTIA через Правление ICANN.

ICG единогласно поддерживает это предложение и рекомендует, чтобы все затронутые стороны реализовали его. ICG утверждает, что это предложение и все связанные с ним процессы соответствуют критериям, изложенным в нашем уставе и мандате, в том числе критериям NTIA».

Днем позже Правление ICANN постановило передать это предложение на рассмотрение NTIA.

В тот же день NTIA опубликовало на своем сайте заявление руководителя NTIA **Лоренса Стриклинга** (Lawrence E. Strickling) (https://www.ntia.doc.gov/blog/2016/reviewing-iana-transition-proposal). В частности, он отметил: *«Теперь NTIA начнет процесс рассмотрения этого предложения – мы надеемся, в течение 90 дней, – чтобы определить, соответствует ли он критериям, которые мы определили, когда объявили о передаче».*

Что ж, будем ждать.

Сказка о мудром Кавусе и троллях

«Сказывает Фади ибн-Шехаде аль-Бейрути, мир с ними обоими, что жил в Тегеране мудрец по имени Кавус Арастех и слыл он великим знатоком управления Интернетом. Всю жизнь положил Кавус на то, чтобы противостоять интернет-происками Великого сатаны. Он боролся с ним в Международном союзе электросвязи, громил на всемирном Форуме по управлению Интернетом, обличал на заседаниях Правительственного консультативного комитета (ПКК) интернет-корпорации ICANN. Завидев Кавуса, дрожали мелкой дрожью тысяча тысяч демонов и шайтанов мультистейкхолдеризма и пересыхала от страха ядовитая слюна у них во рту, и заслышав его гневные, обличительные речи, бежали они прочь, как песок, носимый самумом, и мудрые слова его гремели набатом в ушах правоверных.

Но чем дольше боролся Кавус, тем сильнее становился коварный враг, и однажды стремительным, как пущенная из лука стрела, имейлом долетел до Кавуса из родных краев наказ поддерживать мультстейкхолдерные процессы и передачу функции IANA в руки тех самых мультистейкхолдерных демонов и шайтанов. Слабый пал бы духом, глупец смутился, но истинная мудрость, как учит нас премудрый Фади ибн-Шехаде аль-Сингапури, мир с ними обоими, в том, чтобы заставить ветер конъюнктуры дуть в паруса дхоу твоего успеха. И вновь зазвучало в высоких собраниях пламенное слово Кавуса, и преисполнился благостью ПКК, и воспряли духом мультистейкхолдеры», — так гласит 1002-ая ночь Шехерезады.

Я любил смотреть и слушать выступления Кавуса — они добавляли соли и перца в пресную, усыпляющую атмосферу интернет-официоза. Особенно забавно было видеть его на заседаниях в МСЭ, где система звукопоглощения не позволяет услышать слов, не произнесенных в микрофон: снимаешь наушники и видишь, как у человека в шаге от тебя пульсируют от напряжения жилы на шее, как он отчаянно жестикулирует, и все беззвучно, как в космосе.

В какой-то момент меня познакомили с Кавусом. На фоне сомнительной свежести маек и мешковатых джинсов армии стейкхолдеров он смотрелся щеголем с лондонской Сэвил-роу: безупречный костюм, строгий галстук в горошек, седина и неожиданно тихий голос и печальный взгляд исполненных истинно восточной неги персидских очей. Я не мог не выразить своего восхищения его мастерством троллинга. Ответ Кавуса, искренний и наивный, был неожиданным:

— А что такое троллинг?

— Вы это серьезно? — Чашка с кофе чуть не выпала у меня из рук.

Пришлось рассказать, что есть троллинг и привести в качестве примера его же собственные речи. Кавус ушел в себя, а затем со вздохом промолвил:

— Мне все это надо серьезно обдумать.

Мы расстались, чтобы на следующий день встретиться на заседании ПКК. Войдя в зал, я увидел Кавуса. Как всегда в безупречном костюме-двойке он на всех парах двинулся в мою сторону. Приблизившись, Кавус поднял пальцы а-ля Уинстон Черчилль в знаке победы V и заговорщицки прошептал **«Let's keep trolling!»** («Давай будем троллить и дальше»). И танцующей походкой снявшего тяжкое бремя с души человека двинулся к собратьям по ПКК, которые, как ласточки на проводах, весело щебеча, уже занимали отведенные им за длинным столом места.

«Как учит прославленный Фади ибн-Шехаде аль-Лосанджеласади, мир с ними обоими, если человек не идет к мультистейкхолдеризму, то мультистейкхолдеризм приходит к человеку, и тот, кто властвует над своими троллями, обретет покой и благоволение, и будут его славить в подлунном мире как великого мудреца все те годы и десятилетия, что идет процесс передачи функции IANA в управление мультистейкхолдерного сообщества — мир с ними обоими», — так с наступлением рассвета заканчивает дозволенные речи 1002-ой ночи Шехерезада.

Украина превращается в источник ложной маршрутизации

Даг Мадори (Doug Madory)

С помощью мощной сети мониторинга системы маршрутизации Интернета, Даг Мадори анализирует отдельный класс случаев мошеннической маршрутизации, основанной на временной «аренде» чужого адресного пространства в преступных целях. Преступники прибегают ко все более ухищренным способам сокрытия своей деятельности, но пристальный анализ позволяет выявить эти случаи.

Прошлой осенью министр внутренних дел Украины объявил о создании национальной Киберполиции для защиты страны от любых киберугроз, начиная от мошенничества с кредитными картами и заканчивая борьбой с компьютерными вирусами. Неплохо было бы добавить к этому списку еще один пункт: мошенническую BGP-маршрутизацию из Украины. В прошлом году мы сообщали об инциденте, в ходе которого украинский интернет-провайдер «Вега» перехватил маршруты у компании British Telecom (включая маршруты Научно-исследовательского центра ядерного оружия Великобритании), событии, которое возможно относится к разряду безобидных ошибок. Однако та мошенническая маршрутизация, которую мы наблюдаем сейчас на Украине, явно разработана таким образом, чтобы быть незамеченной. В этом блоге мы рассмотрим некоторые аспекты этого нового явления.

Государственные органы отмечают некоторые странности

Масштаб проблемы стал более заметен в прошлом году, когда государственным органам некоторых стран пришлось столкнуться с проблемой мошеннического использования своего адресного пространства. В июле прошлого года министру иностранных дел Нидерландов пришлось отвечать на парламентские запросы, касающиеся инцидента, при котором «организаторы компьютерных атак» завладели IP-адресным пространством, принадлежавшим в прошлом году министерству иностранных дел. В ходе этого инцидента, который имел место 18 ноября 2014 года, компания Decision Marketing (AS62228), действующая из Софии (Болгария), начала глобально анонсировать одиннадцать BGP-маршрутов, которые ей не принадлежали.

Эти маршруты включали в себя следующие адреса (смотреть таблицу 1).

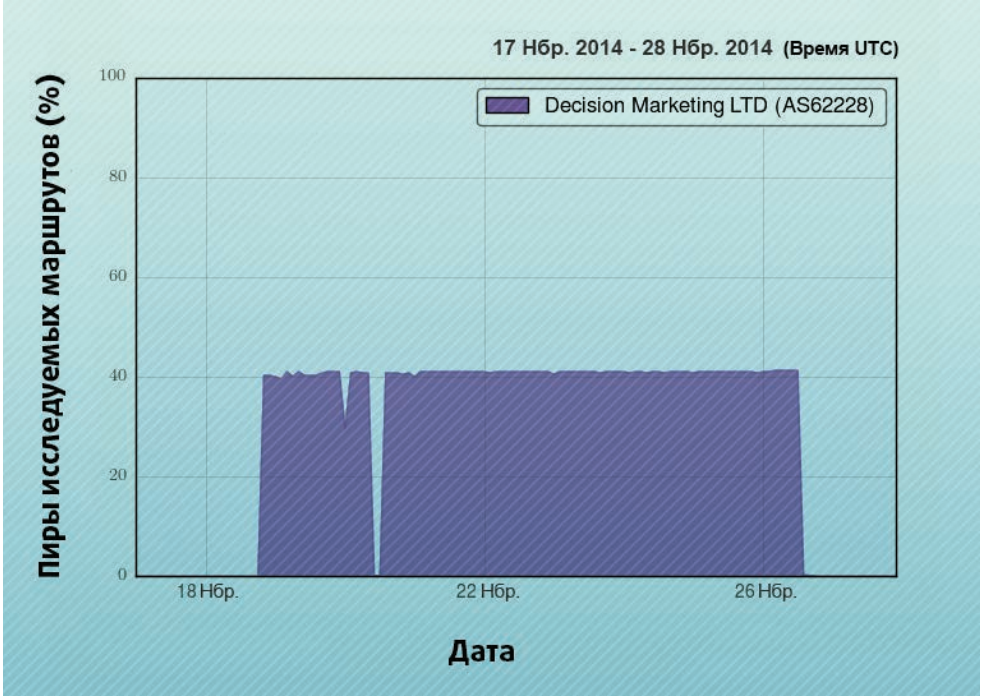
Маршрут, который привлек внимание Нидерландов, был 193.177.64.0/18. Его профиль распространения (рис.1) показан внизу слева – следует отметить, что он никогда не анонсировался более 40% нашей пиринговой базы. Компания Decision Marketing (явно действуя в качестве спамера) своим слоганом «Мы – компания электронной почтовой рассылки» старательно создавала впечатление своей принадлежности к Болгарии.

В следующем месяце швейцарская государственная группа реагирования на компьютерные происшествия объявила о том, что ей удалось (при содействии агентства Spamhaus) возвратить IP-адресное пространство, принадлежавшее региональным государствен-

Таблица 1. адреса маршрутов

159.100.0.0/17	Transport Research Laboratory	GB
171.25.0.0/17	Swisscom IT Services AG Sankt Gallen	CH
193.177.64.0/18	Ministerie van Buitenlandse Zaken	NL
193.201.243.0/24	MA3X Ltd. Sofiya Sofiya-Grad	BG
193.202.128.0/18	Bayer Business Services GmbH Nordrhein-Westfalen	DE
193.243.0.0/17	Cable & Wireless UK P.U.C.	GB
194.38.0.0/18	RIPE Network Coordination Centre	AU
210.79.128.0/18	Mediatti Communications Inc.	JP
210.87.64.0/18	Asia Pacific Network Information Centre	AU
80.114.192.0/18	Ziggo B.V. Amsterdam Noord-Holland	NL
83.175.0.0/18	Telecom Italia S.p.a.	IT

Рис.1. Адрес 193.177.64.0/18 (Ministerie van BZ, Нидерланды), Источник: BGP Data



ным органам Швейцарии, которым пользовались спамеры. На схеме, приведенной ниже (рис.2), показано, как маршрут, использовавшийся в ходе спамерской операции (AS62741) (на графике слева), сначала исчезает 25 июня, а затем возвращается 29 июня к своему законному владельцу, кантону Фрибур.

В защиту голландского министра стоит заметить, что вряд ли возможно полностью предотвратить использование адресного пространства одной сети другой сетью, так как система маршрутизации основана на доверии. Также нужно учесть, что перехват неиспользуемого адресного пространства, несомненно, имеет более низкую приоритетность по сравнению с другими проблемами, с которыми сегодня сталкиваются правительства европейских стран. Возможно, по мере того, как объем доступного адресного пространства IPv4 начнет сжиматься, спамерам станет труднее захватывать незаметно неиспользуемое IP-адресное пространство?

Проблема на Украине

В октябре прошлого года Джим Коуи (Jim Cowie), ученый-эмерит компании Dyn, выступил в качестве основного докладчика на 10-й конференции ENOG, проводившейся в Одессе (Украина). ENOG (Евразийская группа операторов сетей) охватывает Российскую Федерацию, страны СНГ и Восточной Европы. С видеопрезентацией Джима можно познакомиться здесь: https://www.youtube.com/watch?v=cag0A3TH5wU – следует обратить внимание на фрагмент, посвященный мошеннической маршрутизации, которую мы обнаружили исходящей из Украины.

В начале прошлого года мы опубликовали в нашем блоге пост под названием «Бескрайний мир мошеннической маршрутизации», в котором подробно рассмотрели действия шести сетей, преднамеренно анонсирующих адресное пространство, которое им не принадлежит. В случае под номером 5 этого поста мы дали описание преступника, пытавшегося замаскировать свой мошеннический маршрут, создав ложный AS-путь, который содержал свойства, в ином случае принимаемые за вероятный источник объявленного адресного пространства.

В данном случае, по нашим наблюдениям, применялось неиспользованное адресное пространство компании British Telecom, объявленное под номером AS5400 (номер ASN компании British Telecom), согласно AS-путям в данных BGP. Для обычного наблюдателя это могло показаться законным, однако их выдавало то, что трафик проходил исключительно через мелкого интернет-провайдера в Уфе (Россия) – то есть через город, в котором вряд ли размещается филиал British Telecom.

Деятельность, описанная в случае номер 5, исчезла в ноябре 2014

Рис.2. Адрес 155.288.0.0/16 (Фрибур, Швейцария), Источник: BGP Data

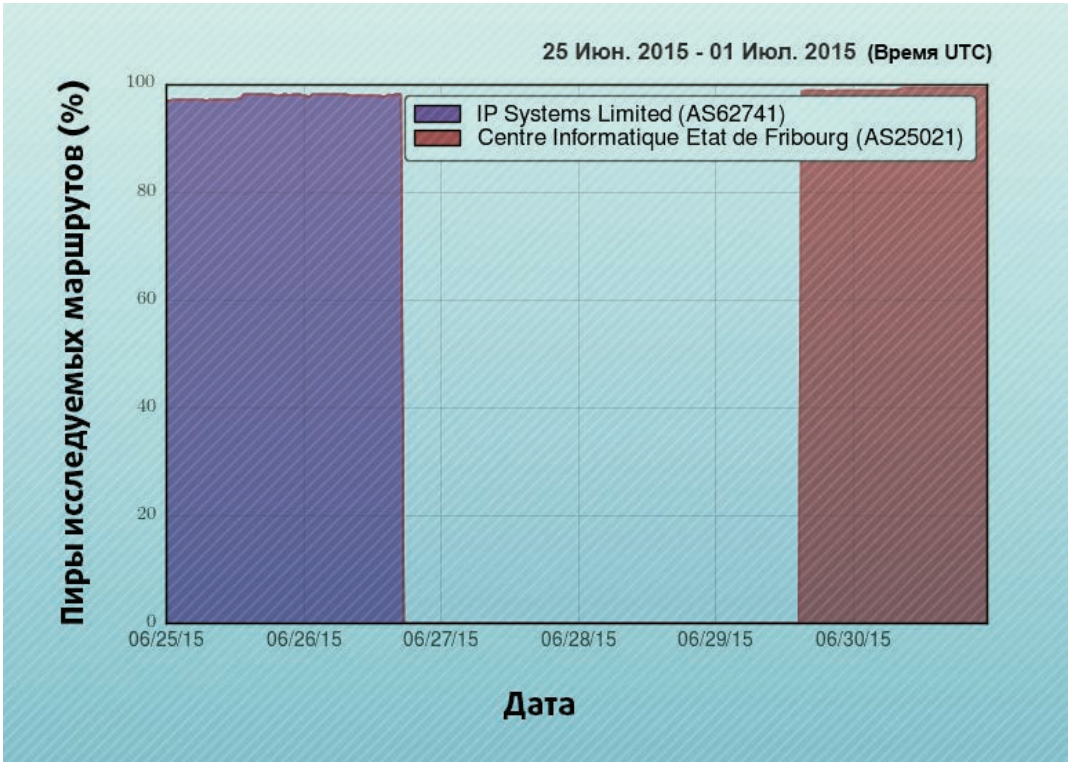


Таблица 2. Ложные маршруты, проходящие через Украину

Prefix	Plausible, but Phoney Origin
187.239.0.0/16 (Uninet, MX)	AS8151 (Uninet, MX)
177.90.0.0/16 (Universidade De Sao Paulo, BR)	AS28571 (Univ De Sao Paulo, BR)
200.200.0.0/16 (Embratel, BR)	AS4230 (Embratel, BR)
181.56.0.0/16 (Telmex Colombia, CO)	AS10620 (Telmex Colombia, CO)
161.255.0.0/16 (Movistar (Telcel), VE)	AS6306 (Movistar (Telcel), VE)
177.21.128.0/20 (Netdigit Telecom, BR)	AS28245 (Netdigit Telecomunicacoes, BR)
196.3.16.0/20 (Net Uno, C.A., VE)	AS11562 (Net Uno, C.A., VE)
186.189.224.0/20 (FastBee Argentina S.A.)	AS28028 (FastBee Argentina S.A)
186.236.240.0/20 (Prefeitura de Cuiabá, BR)	AS263638 (Prefeitura de Cuiabá, BR)
191.102.224.0/20 (DirecTV Colombia)	AS262928 (DirecTV Colombia)
177.8.80.0/20 (Centro Int. de Telemática do Exército,BR)	AS52890 (Centro Int. de Telemática do Exército,BR)
... и еще много других	

Таблица 3. Маршруты из Москвы и Минска

Путь 20 мс из Москвы, RU to 200.202.64.1					
1	*				0.0
2	87.245.229.46	ReTN external interconnections	Москва	Россия	0.478
3	87.245.233.26	ReTN's Backbone	Киев	Украина	19.717
4	*				0.0
5	200.202.64.1	BR HOME SHOPPING LTDA	Белу-Оризонти	Бразилия	20.419
Путь 12 мс из Минска, BY to 200.202.64.1					
1	*				0.0
2	*				0.0
3	93.84.125.194	BELTELECOM	Минск	Белоруссия	4.343
4	93.85.80.54	Republican Unitary Telecommunica	Минск	Белоруссия	4.425
5	93.85.80.126	Republican Unitary Telecommunica	Минск	Белоруссия	0.984
6	87.245.237.21	ReTN external interconnections	Киев	Украина	12.405
7	87.245.232.173	ReTN's Backbone	Киев	Украина	12.511
8	*				0.0
9	200.202.64.1	BR HOME SHOPPING LTDA	Белу-Оризонти	Бразилия	12.67

года, но в следующем месяце, в декабре, мы начали наблюдать схожую активность, идущую из Киева (Украина), то есть мы имеем новый пример ложного, хотя и правдоподобного AS-источника для мошеннической маршрутизации.

Маршрут 200.202.64.0/19 (компания Brazil Home Shopping Ltd) был одним из таких маршрутов. Маршрут проходил по следующему пути:

... 9002 8438 18739 10495 11295

Если исследовать этот путь, то мы увидим, что он изначально анонсировался из AS11295 (компания Brazil Home Shopping Ltd). Казалось бы, на этом этапе все нормально. Затем он проходит через AS18739 и AS10495, которые тоже относятся к бразильским номерам автономных систем (ASN). Опять-таки кажется правдоподобным? Но затем путь эксклюзивно проходит через украинского провайдера «Гетман Софт» (AS8434) и далее по российской стационарной линии RETN (AS9002). Маршруты с подобными путями анонсируются ограниченному числу преимущественно российских провайдеров.

В прошлом году мы наблюдали за этой компанией, анонсировавшей ложный, хотя и правдоподобный источник маршрута (по всей видимости, она предпочитает ресурсы LACNIC). Смотреть таблицу 2.

Если мы хотим получить дополнительное подтверждение места, где берут начало эти маршруты, то мы можем проследить это адресное пространство с помощью утилиты traceroute и получить временные параметры и пути, которые соответствуют Украине, а не Бразилии. Например, таковым является путь 20 мс из Москвы. Смотреть таблицу 3.

По состоянию на пятницу прошлой недели маршрут 200.202.64.0/19 (компании Brazil Home Shopping Ltd) все еще мошеннически объявлялся исходящим из Украины, хотя путь AS был слегка изменен (вместо AS8434 использовалась AS41331):

... 9002 41331 18739 10495 11295

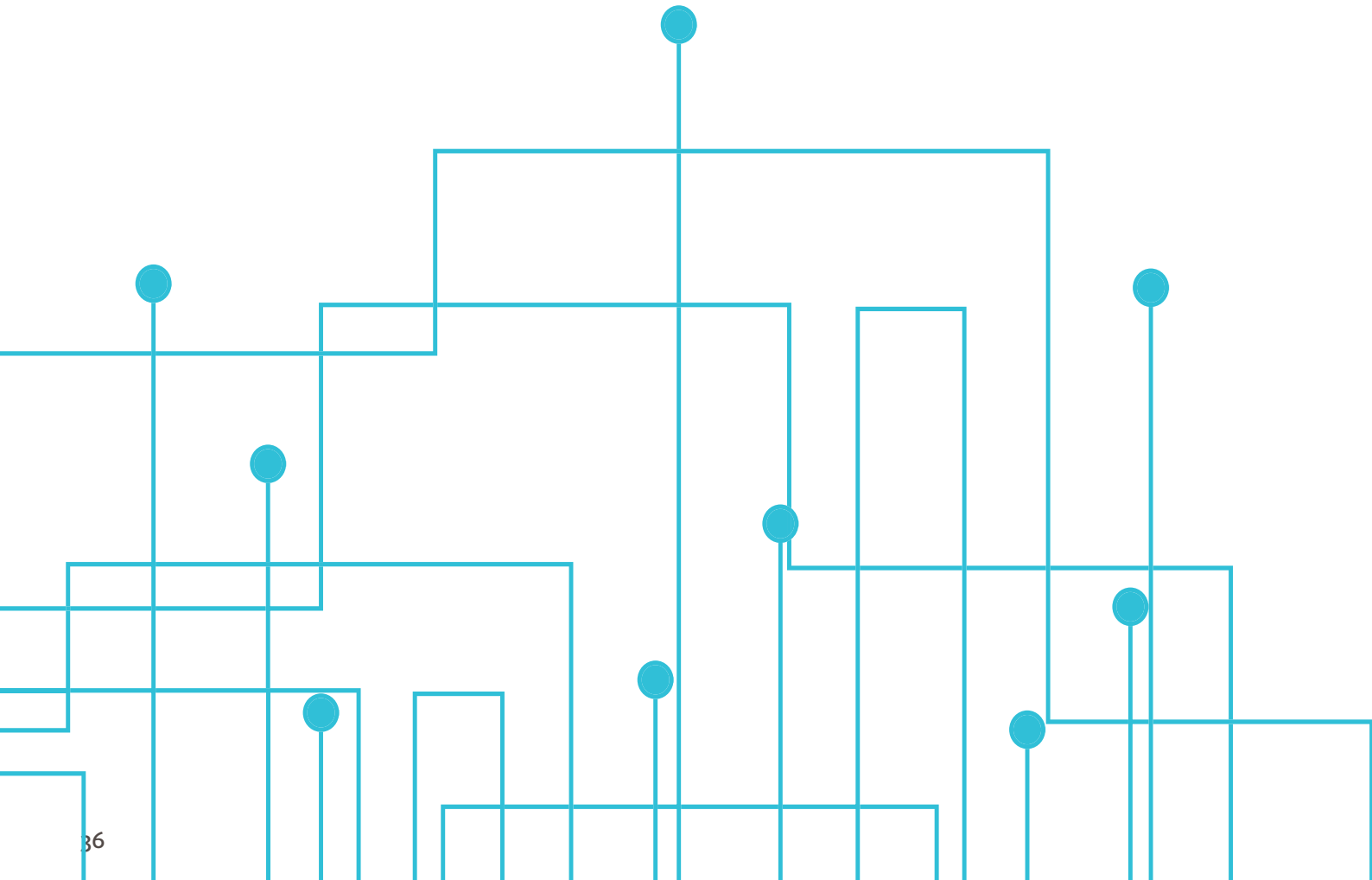
Люди, которые занимаются этой деятельностью, могут вести себя совершенно бесцеремонно. Мало того, что они осмеливаются объявлять адресное пространство, принадлежащее бразильским военным (Centro Int. de Telemáticado Exército), как это видно из примера, приведенного выше, ранее в этом году они провели еще одну операцию по за-

хвату адресного пространства конференции APRICOT 2016 всего за несколько недель до начала конференции. Конференция APRICOT является техническим форумом APNIC, на котором обсуждаются такие вопросы, как безопасность маршрутизации. Мы предупредили об этом организаторов конференции, и им удалось защититься от захвата, перехватив ложные анонсы преступников (AS260) выше по маршруту (GTT, AS3257). Джо Эбли (Joe Ableу), директор Dyn по инфраструктурным вопросам, охарактеризовал этот инцидент как один из самых ярких сюжетов для обсуждения на APRICOT 2016:

Несмотря на то, что GTT заблокировала анонсы конкретных сетей APRICOT 2016 у своего клиента AS260 (Xconnect24), преступники продолжали объявлять ложные маршруты через GTT, маскируя источник этих маршрутов, так, как они это делали в случае с APRICOT.

К сожалению, бороться с этим видом активности очень трудно, так как преступники применяют все более изощренные методы, скрывая свою деятельность от базового BGP-анализа, но также и потому, что в случае обнаружения их мошеннической деятельности и предупреждения провайдеров, они продолжают анонсировать ложные маршруты. Именно поэтому мы поддерживаем проект Internet Society по применению взаимно согласованных норм по безопасной маршрутизации (MANRS) и рекомендуем компаниям и организациям отслеживать свое IP-адресное пространство (анонсируемое и неанонсируемое) при помощи таких инструментов, как, например, Internet Intelligence, предлагаемых Центром Дин. Для получения более подробной информации об инцидентах подобного рода смотрите статьи, посвященные нашему анализу в Washington Post и Wall Street Journal.

Источник: Ukraine Emerges as Bogus Routing Source, http://research.dyn.com/2016/03/ukraine-emerges-as-bogus-routing-source/



IT-конференции

Ольга Александрова-Мясина

Я понимаю, что уже весна и 2016 год, что жить надо будущим, а не прошлым, но не написать об осенне-зимних мероприятиях года ушедшего было бы неправильно. Потому что именно тогда меня преследовала «тень президента». Началось это в ноябре, когда мы готовили мероприятие, посвященное 20-летию MSK-IX, а закончилось в декабре, на католическое Рождество, но обо всём по порядку.

Чтобы устроить праздник для друзей MSK-IX, было принято решение найти какой-нибудь ресторанчик и позвать туда самых лучших и любимых. Что и было сделано. Клуб Русского географического общества с удовольствием предложил нам свои услуги в требуемом формате. И вот, когда все гости были приглашены и письменно, и устно, не ревизор, его на другое число не перенесешь. Пришлось нам собирать волю в кулак, запастись валерьянкой (хотя у меня с тех пор глаз дергается) и перенести мероприятие на другое число! А это кошмар любого организатора, и вот почему. Велика вероятность, что человек, которого предупредили о переносе, всё равно что-то перепутает или забудет, или не так поймет и придет, когда не надо. Или по каким-то причинам кто-то из гостей вовсе не получит информацию о переносе и тоже придет, когда не надо. А еще может быть ситуация, когда люди вроде бы все правильно поняли, но находится кто-то один, кто убедит всех в обратном, и в итоге они все вместе приходят, когда не надо. В общем, это настоящий ужас, и мне выпало это пережить. Как-то разрулили. Всё прошло достойно, и гостям вечер понравился.

а до вечера осталось всего несколько дней, из клуба позвонили и сказали, что принять нас не смогут, так как к ним едет президент России! Ну, тут дело такое — президент

Потом начался декабрь, который традиционно уже открывается Пиринговым форумом MSK-IX. Мы с коллегами готовили его продолжительное время.

Как и в 2015 году, форум проходил в здании правительства Москвы на Новом Арбате. Мы выбрали этот зал второй раз подряд из-за вместимости, хотя лично мне эта площадка никогда не нравилась, так как минусов там больше, чем плюсов. Но с плюсами тоже надо считаться.

Итак, нам повезло, и в малом зале этого же здания в то же время проходила конференция эндокринологов. Почему я говорю, что повезло? Потому что на входе, например, было совсем не скучно. Мы встречали гостей, проверяли паспорт (это было требованием службы безопасности) и задавали еще один дополнительный вопрос: *«Вы на пиринг или на эндокринологию?»*. Вообще, со стороны это звучало невероятно комично, но через час мы опытным взглядом могли различать людей без дополнительных вопросов: угрюмые, сосредоточенные дядьки с большими рюкзаками — это к нам, а все женщины за 60 с бутылками воды — на соседнюю конференцию. Раскидав таким образом на потоки первых посетителей, я побежала в большой зал вести онлайн-трансляцию технической части в социальных сетях.

Начался технический семинар представлением большого исследования, которое провел **Джим Коуи** (DYN Research). В этом исследовании он сравнил две российские национальные доменные зоны — .RU и .РФ, особое внимание уделив вопросам размещения контента в этих доменах. Свой выбор Джим Коуи объяснил тем, что .RU и .РФ являются очень крупными доменными зонами и показывают достаточно типичную для крупных национальных доменов картину. Исследование показало, что большая часть контента хранится локально — в .RU 65%, а в .РФ – 81%. «Задержка сигнала Москва-США-Москва составляет от 100 до 250 миллисекунд (в зависимости от того, с какого побережья США идет сигнал — с Западного или Восточного). Для интерактивного контента это критически много, и может привести к потере пользователей.

Поэтому локальное хранение хостинга важно для работы многих сервисов», – рассказал Коуи.

О возможности применения технологии LookingGlass шла речь в докладе **Егора Дробышева (Sea-IX)**. Он рассказал о том, как использовать эту технологию для наблюдения за работой IX, и привел множество интересных примеров того, как с помощью LookingGlass получить полную картину функционирования точек обмена трафиком. *«Это очень удобный инструмент, который позволяет наблюдать за работой IX в реальном времени»*, – сказал он.

Максим Раевский (IVI), в шутливой форме критикуя операторов, дал им несколько «вредных советов о том, как неправильно подключаться к IX». Именно так он назвал свою презентацию, в которой рассказал об опыте работы с точкой обмена трафиком со стороны крупного клиента. Максим Раевский перечислил множество ошибок, которые совершают неопытные менеджеры при подключении к IX.

Максим Каминский (Brain4Net) и **Сергей Монин** (НП «ЦПИКС») рассказали о технологии SDN/NFV с разных точек зрения: оператора, абонента и IX. Для каждого типа пользователей SDN/NFV имеются свои особенности и нюансы ее применения, поэтому Максим Каминский предположил, что массовое внедрение этой технологии в России начнется не ранее, чем через три года. А Сергей Монин поделился опытом практического внедрения и использования SDN/NFV в точке обмена трафиком.

Одной из важных тем технического семинара стала тема контроля и мониторинга сетей. Денис Матоушек (Netcope Technologies) рассказал об использовании сетевых интерфейсов FPGA для задач анализа и мониторинга трафика.

Павел Храмцов (MSK-IX) в своем выступлении коснулся атаки на root-серверы DNS по всему миру и в связи с этим рассказал о работе **международного проекта Yeti**, в котором участвует MSK-IX и который посвящен изучению функционирования DNS-сервиса в среде IPv6. *«Участие в проекте Yeti позволяет нам повышать качество оказываемых услуг и квалификацию персонала»*, – сказал Павел Храмцов.

Лев Бокштейн (Extreme Networks) наглядно продемонстрировал, как происходит управление локальными точками доступа и какие инструменты можно для этого применять. Он проанализировал то, как пользовались Wi-Fi-доступом участники сегодняшнего форума, и прямо на экране показал, как распределялся трафик и кто из участников стал чемпионом по трафикопотреблению в первые три часа работы. Персональные данные участника разглашены не были.

После технической части было заседание неформальной **рабочей группы «BGP: к лучшему протоколу и практикам»**. Организаторы заседания и его соведущие — **Андрей Робачевский** (ISOC) и **Александр Азимов** (Qrator Labs) — пригласили участников форума, интересующихся вопросами работы протокола BGP, к коллективному поиску решения насущных проблем протокола BGP. *«Протокол BGP не содержит бизнес-правил, и сети многих операторов испытывают трудности. Чужие ошибки влияют на нас, это может приводить к глобальным инцидентам по нарушению маршрутизации. Инициатива MANRS (Коллективная ответственность и сотрудничество для устойчивой и защищённой системы маршрутизации) помогает избежать этих проблем и сделать работу сети более устойчивой»*, — рассказал **Андрей Робачевский**.

Рабочая группа вызвала такой интерес, что мы изначально не рассчитали количество стульев, и некоторым опоздавшим пришлось переминаться с ноги на ногу в проходе, чтобы ухватить хоть что-то из сказанного. Я, конечно, переживала. Во-первых, потому что у меня в принципе повышенная тревожность, а во-вторых, всегда чувствую неловкость, когда гостям не на что сесть и нечего есть. Хотя многие меня уверяют, что это всё не главное, и самое важное в мероприятии — это атмосфера. Только я не понимаю, как может быть хорошая атмосфера, если есть нечего и сесть не на что.

Потом было официальное открытие форума, и на сцену вышла директор MSK-IX **Елена Воронина**. Она рассказала о музее MSK-IX, который был открыт недавно.

Круглый стол «Будущее IXP» вёл **Константин Чумаченко** (MSK-IX), он собрал известных отраслевых экспертов и руководителей, поделившихся своим видением направления, в котором развиваются сегодня точки обмена трафиком, и перспектив развития точек. Участники круглого стола поговорили о роли, задачах и приоритетах IXP. Так, Елена Воронина главными приоритетами в деятельности точек обмена трафиком назвала стабильность, безупречное качество услуг и высокую пропускную способность. *«IXP должны не просто постоянно находиться на рынке; они должны работать так, чтобы к ним хотели подключаться все бизнесы — и интернет-ориентированные, и не имеющие к технологиям никакого отношения. Поэтому нам нужно сделать так, чтобы наши услуги были доступны и удобны для всех потенциальных клиентов»*, — сказала Елена Воронина. Также участники обсудили роль IXP в развитии Интернета и интернет-сообщества и выразили уверенность в том, что значение IXP, причем не только для сети в техническом смысле, но и для Интернета как социального явления, в ближайшие 10 лет существенно возрастет. Очень скоро, по их мнению, все без исключения коммуникации будут осуществляться только через Интернет. Закончилась конференция фуршетом с праздничным тортом и танцами под духовой оркестр.

Это было 10 декабря, а буквально за несколько дней до Пирингового форума нас поставили в известность, что MSK-IX, Технический центр Интернет и Координационный центр национальных доменов .RU/.РФ будут участвовать со стендом и секцией в новой конференции «Интернет-экономика», куда приедет президент. Когда нам это сказали, я как-то призадумалась... Всякое бывало в моей жизни, но вот чтобы за 10 дней найти контент для секции и построить стенд

для визита президента — такое случилось впервые. Узнали мы это на общей встрече с коллегами из РОЦИТ и РАЭК, проходила она в башне Москва-Сити. Вышли мы оттуда с Мишей Анисимовым, и пошли пить кофе на первый этаж. Я склонилась над чашкой и мрачно сказала: «Нам конец». Миша залился звонким смехом, на что я даже не подняла головы. Было страшно.

«Не парься, — сказал Миша. — *Мы не одни такие, а значит, есть механизмы для того, чтобы это сделать».* И помолчав, добавил: *«Ну, если, конечно, повезёт...»*

Нам, благодаря какому-то чуду, действительно повезло, и всё успели подготовить при помощи коллег из РОЦИТа и РАЭКа; и секцию, и ролик для объединенного стенда трёх компаний, а сам стенд мы делили вместе с Фондом информационной демократии.

В первый день было много людей в фойе и в залах, но все они были свои, из одной интернет-отрасли, случайных людей я там не заметила, и про услуги наших компаний рассказывать было некому. Мы общались с коллегами, слушали секции, обсуждали следующий день, когда должен был приехать **Владимир Путин**.

Вечером позвонил мой старший брат, сообщил, что наш стенд и меня вместе с директором Координационного центра **Андреем Воробьевым** показывали по Первому каналу — вот она, слава! Хоть на несколько секунд, но зато на всю страну! И родственники заметили!

Во второй день на форуме была значительно усилена охрана, и посетителей пропускали внутрь только по специальной голограмме на бейдже, которую надо было получать накануне у представителей спецслужб.

Меня как-то легко пропустили и даже паспорт не стали проверять, видимо, я внушала доверие охране президента, и это сильно облегчало процесс попадания внутрь.

В зоне кофе-брейка было изобилие плюшек, бутербродов и конфет, а людей почти не было — что очень нетипично для IT-мероприятий. Я быстро выпила чаю и поднялась на второй этаж, где отраслевые эксперты в парадных одеждах замерли в ожидании...

Ровно в 11 утра все двери в здание закрыли, а участников попросили пройти в большой зал, что все и сделали. Мы расселись по местам согласно отметкам на бейджах, и всё...

Это продолжительное ожидание превратилось в мучение. К сожалению, я не догадалась взять с собой воду с бутербродами и заранее сходить в туалет, потому что из зала больше не выпускали, а президент появился только в 4 часа дня. Наверное, я еще никогда и никого не ждала так в жизни, как его в этот день! Потому что появление Путина было для меня приравнено к скорому получению некоторых бытовых мелких радостей, на которые мы обычно в повседневной жизни не обращаем внимания... Сзади кто-то сказал: *«Мне кажется, что я еду в сидячем поезде в Екатеринбург»*.

Люди знакомились, делали селфи, публиковали разные статусы в социальных сетях и размещали фото друг друга. Зато закончилось потом всё очень быстро. Президент поднялся на трибуну, сказал слова про важность интернет-экономики и вклад Интернета в развитие страны, назначил **Германа Клименко** своим советником по Интернету и уехал. В принципе, на этом конференция закончилась, потому что потом все только это и обсуждали с небольшими перерывами на фрукты и шампанское. В общем, день удался. Мы с коллегой пошли ужинать в соседний ресторан и радовались, что всё наконец-то позади!

Сетевая общественность долго потом не могла успокоиться и брызгала слюной по поводу назначения Германа. Эта тема была трендом многих месяцев, и тогда я с грустью думала, что человеческая реакция на чужие успехи всегда бывает негативной. В разные времена и в разных странах только немногие искренние друзья могут реально радоваться чьим-то успехам. Когда речь идет о массовом психозе, то добра не жди – обязательно найдутся люди, которые всё и всем припомнят.

Когда отшумели новогодние праздники и рынок начал постепенно оживать, мы принялись к приготовлению **Cyber Security Forum**. Нет, организаторами был РОЦИТ — мы только помогали, чем могли и участвовали со стендом. Так как стенд был единственный в фойе, все подходили с прикладными вопросами; где туалет, почему кофе не наливают, и когда будет выступать тот или иной докладчик. Нас это не бесило, наоборот, мы с радостью размахивали руками и давали ответы на разносторонние вопросы. На открытии выступали чиновники и политики, говорили о том, как страшно жить. В частности, жить страшно пользователям Интернета, которые не пользуются головой и оставляют финансовую информацию на фишинговых сайтах.

Одно из центральных событий форума – подписание меморандума о развитии российского сегмента доменного пространства сети Интернет между Координационным центром национального домена сети Интернет, Техническим центром Интернет и точкой обмена трафиком MSK-IX. Документ подписали директор Координационного центра **Андрей Воробьев**, генеральный директор ТЦИ **Алексей Платонов** и генеральный директор MSK-IX **Елена Воронина**. Меморандум послужил еще одним примером сотрудничества между организациями, которые отвечают за бесперебойную стабильную работу российской интернет-инфраструктуры.

«Подписание меморандума — это элемент доверия, без которого Интернет существовать просто не может. Этот документ определяет пути развития российских доменов верхнего уровня до 2020 года — причем не только существующих, но и тех, которые могут появиться в ближайшие годы: в 2017 году ICANN собирается объявить о старте второго этапа программы по регистрации новых доменов верхнего уровня», — сказал Андрей Воробьев.

Также одной из тем Cyber Security Forum стали вопросы цифровой грамотности, просвещения специалистов и пользователей, формирования позитивной интернет-среды. На форуме прошел всероссийский чемпионат по интернет-игре **«Изучи интернет – управляй им!»** Так как до старта чемпионата еще два месяца, это был специальный турнир для учащихся 8-11 классов, приуроченный к Неделе безопасного Рунета.

Потом наступил март — и я уехала в Таллин на 19-ую встречу маркетологов организации **CENTR** (объединяет европейские регистратуры верхнего уровня). Никогда до этого не была в Прибалтике, потому широко раскрытыми глазами смотрела по сторонам. Когда мне было лет 10, мама приезжала сюда на экскурсию, а я получила вельветовые джинсы и свитер с зелеными ромбиками. Носила я их потом еще несколько лет. Так как с красивыми вещами в Советском Союзе было туго, то откуда только можно привозили обычно одежду на вырост — края брюк подворачивали на несколько сантиметров, а в пояс вшивали резинку. Таким образом, вместе с чадом росла и вещь, постепенно отпускались края штанин и ослаблялась резинка.

Выйдя из самолета, я почему-то сразу вспомнила об этих детских брюках и свитере в ромбик, наверное, потому что только это и связывало меня с этим городом. Решили с коллегой взять такси до отеля, несмотря на курс евро, который был беспощадным и медленно толкал туристов начинать возить с собой кофе и консервы в заграничные поездки.

Отель наш был в центре города, мои окна выходили на небольшую площадь напротив Старого города, и я залипала в окне, пробегая из ванной к письменному столу. Надо было быстро переодеться и спешить погулять по городу до приветственного ужина.

На улице светило солнце и было -10. Мысленно я благодарила себя, умную и любимую, за кеды на меху и зимний пуховик. В Старом городе продавалось много вещей из шерсти: платья, тапочки, платки, одежда и всё, что только можно придумать. Даже магнитики на холодильник были меховые. Моя коллега помнила по прошлым визитам стилизованный под старину ресторан без электричества, где продавали ручной работы стаканы из дутого стекла. Зашли туда.

Нас встретила миловидная местная девушка, было видно, что она знает понемногу разные языки и старается сделать приятное иностранцам.

— Нам бы стаканчики купить, — сказали мы.

— По стаканчику? — обрадовалась девушка, — это можно! Садитесь за столик...

Признаться сразу в том, что мы, в общем-то, за другим пришли, не рискнули. Решили выпить домашнего пива на травах, а уже потом выяснять на английском про покупку. В итоге, прекрасные разноцветные стаканы были мной тщательно упакованы в дорожную сумку.

На ужине я оказалась за столиком с дамами из датской и бельгийской регистратуры. Мы поделились советами и ситуациями на наших доменных рынках, пожаловались друг другу на регистраторов и заели наше горе вкусным крем-брюле.

На следующий день был первый день весны, Москву завалило снегом по самые уши, а я рассказывала про маркетинговые программы для регистраторов коллегам из европейских регистратур. Обычно я всегда готовлюсь к выступлениям, даже если выступать на русском, если на английском, то я готовлюсь дольше и тщательнее. Так было и в этот раз, поэтому сильно не волновалась.

Всего в работе встречи приняли представители 25 регистратур. Один из главных вопросов, обсуждаемых на встрече, — взаимодействие национальных регистратур и регистраторов доменных имен.

Представители регистратур отметили, что на сегодняшний день на доменном рынке сложилась ситуация, при которой повышение стоимости доменных имен для регистраторов становится необходимым условием для дальнейшего развития национальных доменов. Большое влияние на рынок оказывают и новые домены верхнего

уровня, которые массово открыли регистрации в прошлом году, и активное развитие социальных медиа, и другие факторы. Поэтому многие из европейских регистратур уже подняли цены для регистраторов – это произошло, например, в национальных доменах Швеции, Великобритании, Бельгии. Участники встречи констатировали, что повышение стоимости доменных имен для регистраторов становится общей тенденцией во всем мире.

Также представители регистратур практически всех европейских национальных доменов жалуются на низкую активность компаний-регистраторов. Это отметил и представитель национального домена Японии (регистратура японского домена .JP является ассоциированным членом CENTR), который рассказал о том, как активность регистраторов влияет на их доходность: 33% японского рынка доменных имен обслуживают всего 5% от общего числа японских регистраторов, но именно они проявляют готовность участвовать в акциях регистратуры. Остальные же регистраторы, слабо реагирующие на предложения регистратуры, получают гораздо меньше выгоды от своей регистраторской деятельности.

Все отметили, что созрела необходимость выработки новых маркетинговых стратегий для повышения этой активности, так как существующие программы очевидно не приносят желаемых результатов из-за изменения пользовательских предпочтений. Пользователи все менее охотно реагируют на текстовую информацию, выбирая визуальный контент и нестандартные решения. Эти тенденции необходимо иметь в виду, планируя различные акции и мероприятия, направленные на привлечение новых пользователей и дальнейший рост национальных доменов.

Второй день воркшопа прошел без представителей из России, так как наша авиакомпания отменила наш рейс, и пришлось срочно паковать вещи для вылета утром. К слову сказать, этот новый рейс всё равно задержали из-за погодных условий, и день прошел в дороге и бестолково. Зато мы купили в зоне вылета вкусную местную еду, которой я и радовала потом домашних.

Когда я ехала в машине от аэропорта до дома, то думала о том, что как же я люблю свой неординарный и странный город Москва.

Календарь событий: 2016 год

Международные события

24-26 апреля 2016
28-й форум EURO-IX,
Люксембург

EURO-IX является ассоциацией точек обмена трафиком (IXP), координирующей различную коллективную деятельность между участниками и предоставляющей информационные услуги, такие как база данных IXP по всему миру. https://euro-ix.net/members/forums/28th-euro-ix-forum/

23-27 мая 2016
RIPE72,
Копенгаген, Дания

Встречи RIPE проводятся два раза в год и собирают около полутысячи частников для обсуждения вопросов политики распределения номерных ресурсов (IP-адресов и номеров автономных систем) в зоне обслуживания RIPE NCC, сотрудничества, а также технических вопросов, связанных с маршрутизацией, DNS, связностью, измерениями и инструментарием. https://ripe72.ripe.net/

13-15 июня 2016
NANOG67,
Чикаго, США

Североамериканская группа сетевых операторов (The North American Network Operators Group, NANOG) является одной из самых активных профессиональных ассоциаций в области сетевой архитектуры, конфигурации и технического администрирования сетей в Интернете. Основной фокус NANOG на технологиях и системах, обеспечивающих работу Интернета: систему глобальной маршрутизации, DNS, пиринг и связность. https://www.nanog.org/meetings/nanog67/home

17-22 июля 2016
IETF96,
Берлин, Германия

IETF (Internet Engineering Task Force) является одной из основных организаций по разработке стандартов в области Интернета. В основном работа в IETF производится в многочисленных списках рассылки, соответствующих различным рабочим группам (этих групп более 100). Три раза в год IETF проводит недельные совещания, на которые приезжают разработчики протоколов, инженеры и операторы со всего мира (в среднем около 1200 участников из более 50 стран мира). http://ietf.org/meeting/upcoming.html

В России

19 мая 2016
Санкт-Петербург

MSK-IX - 2016: Конференция «Российский день IPv6»
MSK-IX при поддержке RIPE NCC и Технического центра Интернет проведет ежегодную конференцию «Российский день IPv6 MSK-IX», посвященную проблемам перехода Интернета на протокол IPv6 и развитию технологий межсетевого взаимодействия. http://www.msk-ix.ru/events/forumspbix2016/

25-27 мая 2016
Сочи

Конференция российских операторов Связи
Конференция российских операторов связи КРОС – это уникальное отраслевое многокомпонентное мероприятие, проводимое в течение 10 лет. На конференцию собираются специалисты и руководители телекоммуникационной отрасли. Участники конференции обсудят передовые технологии и новинки телекоммуникационного оборудования, получат возможность познакомиться и поделиться опытом с коллегами по отрасли из десятков регионов России и стран СНГ, задать острые вопросы представителям государственных органов и профильных ведомств. http://cros.nag.ru/schedule

В Москве

10-13 мая 2016,
ЦВК Экпоцентр

«СВЯЗЬ-2016» 28-я международная выставка
В 2016 году выставка «Связь» впервые пройдет в рамках Российской недели высоких технологий. В рамках «недели» состоятся еще две международные экспозиции: навигационных систем, технологий и услуг «Навитех-2016».
http://www.sviaz-expo.ru/

24 мая 2016,
event-Холл «ИнфоПространство»

Форум «МИР ЦОД. Инфраструктура»
Форум «МИР ЦОД. Инфраструктура» будет посвящен вопросам выбора инфраструктурных решений, проектированию, построению и эксплуатации ЦОД.
http://ospcon.ru/

26-27 мая 2016,
отель «Азимут Олимпик»

VIII Международный бизнес-форум «Wireless Russia Forum: 4G, 5G & Beyond - Эволюция сетей мобильной и фиксированной беспроводной связи в России и СНГ»
Тематика Wireless Russia Forum: 4G, 5G & Beyond сфокусирована на практическом опыте, стратегиях и решениях беспроводных технологий, развертывания мобильных широкополосных сетей LTE, LTE-A, 5G в России и мире.
http://www.comnews-conferences.ru/ru/conference/wireless2016

31 мая 2016,
Swissotel, Конференц-центр

Форум Телеком 2016
Ежегодный форум «Телеком» – это все знаковые представители отрасли на одной площадке, только острые проблемы и лучшие кейсы, конструктивный диалог бизнеса и власти. http://info.vedomosti.ru/events/telekom16/

7-8 июня 2016,
гранд-отель Marriott

ENOG 11
ENOG (Евро-азиатская группа сетевых операторов) представляет собой региональный форум интернет-специалистов, занимающихся важнейшими аспектами работы Интернета.
https://www.enog.org/ru/meropriiatii/enog-11/

7 июня 2016,
Radisson Blu Belorusskaya

Пятый Международный Форум («Будущее Телеком Индустрии: Кросс-Отраслевые Бизнес Модели и Стратегии»)
Основные акценты на мероприятии будут сделаны на развитии новых возможностей телеком-оператора как сервис-провайдера и интеграционной платформы, развитии операторами новых услуг в B2B и B2C сегментах.
http://www.telco-forum.ru/

9 июня 2016,
Holiday Inn Suschevsky

Виртуальные операторы подвижной радиотелефонной связи в Российской Федерации – MVNO Russia 2016
В последние несколько лет в России значительно увеличился интерес к бизнес-модели MVNO, со стороны не только операторов связи, но и корпоративного сектора.
http://www.tmtconferences.ru/mvno2016.html

14 сентября 2016,
Москва

Международная конференция «ЦОД-2016»
Диалог и обмен лучшими практиками между всеми участниками рынка ЦОД – владельцами, операторами, производителями и проектировщиками – с целью решения конкретных практических проблем, возникающих на разных стадиях жизненного цикла дата-центра. http://dcforum.ru/

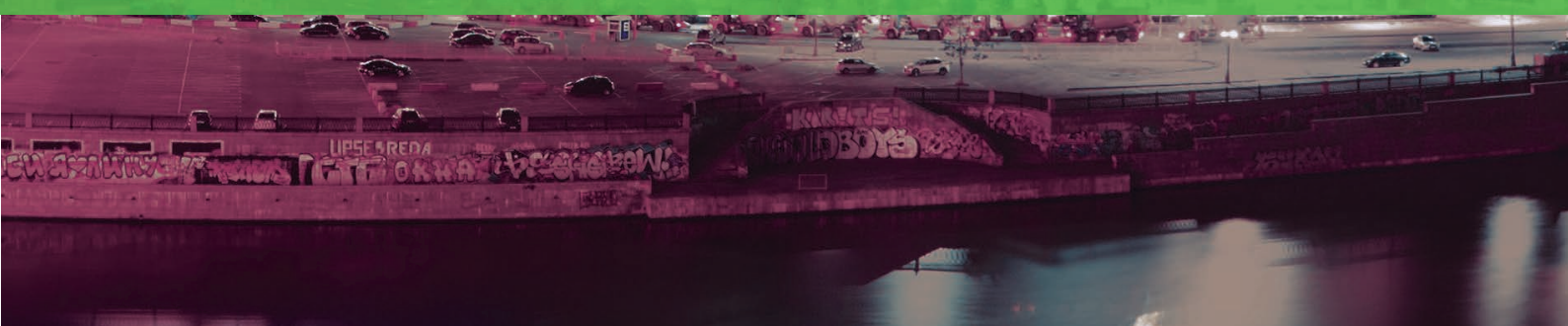
15 сентября 2016,
Москва

Связь в большом городе
Руководители операторов связи из всех крупнейших городов России, представители крупных вендоров, системных интеграторов, поставщиков и разработчиков оборудования обсудят особенности работы и взаимодействия участников телекоммуникационного рынка. http://comnews-conferences.ru/

MSK IX

WWW.MSK-IX.RU
+7 (495) 737-9295

«Московский Internet Exchange» — крупнейшая в России точка обмена интернет-трафиком (IX)



Интернет изнутри

2016