Уязвимость BGP

Ни один другой протокол не может доставить столько радости и беззаботного веселья, сколько BGP. Это клей, связующий весь Интернет и иногда с его помощью можно уложить добрую часть Глобальной сети одним неловким движением. Давайте вспомним самые яркие и знаковые аварии BGP.

"AS 7007 incident"

Это самая первая зафиксированная глобальная авария BGP. И, пожалуй, она же самая яркая. Оригинальный документ настолько хорош, что я его просто переведу вместо описывания истории своими словами. 1997 год. Это был самый обычный день. Интернет ещё только зародился. Большинство провайдеров полностью доверяли друг другу. На SMTP серверах был открыт релей. Огромное количество WEB-прокси и анонимных DialOut. Люди были озабочены исчерпанием пространства IP-адресов, сетевые операторы - загрузкой процессоров, которые едва выдерживали полную таблицу маршрутизации BGP, составлявшую около 45 000 записей. И вдруг, Интернета не стало. Операторы и провайдеры по всему миру бросились на поиски причины проблемы. И нашли. Весь Интернет 1997-го года сосредоточился в одной точке - дешёвом маршрутизаторе Bay Networks в **AS 7007**. Причина проблемы была устранена быстро - сошедший с ума роутер отключили от сети. Но это не решило собственно проблему - маршрутизаторы по всему Интернету продолжали выходить из строя. Кто посылал анонсы? Как можно было остановить это безумие? Неужели Интернету, державшемуся на IRC и склеенному изолентой, пришёл конец? Через несколько часов всё улеглось, а Сетевые Операторы всей планеты собрались для обсуждения последствий этой аварии, и каким образом её можно было предотвратить. Это был момент, когда Интернет изменился навсегда. Но, в отличие от многих других изменений, это прошло незамеченным для большинства обычных пользователей.

Что же произошло в AS 7007 25 апреля 1997?

Теория

BGP - это протокол, с помощью которого одна сеть в Интернете сообщает другой две вещи: факт своего существования и какие сети могут быть достигнуты через них. Ну, ещё они сами изучают, как добраться до других сетей в Интернете. Маршрутизаторы, получившие BGP-информацию, выбирают наилучший маршрут до сети назначения и добавляют его в свои таблицы маршрутизации. Для определения лучшего маршрута BGP использует несколько параметров. Наиболее очевидный из них - количество сетей до точки назначения - длина так называемого AS-Path. *Чем короче AS-Path*, тем выгоднее маршрут. Это, конечно, только верхушка теории, но вы увидите, что нам этого будет достаточно. Другое общее правило - *Longest Prefix Match*. При передаче пакета маршрут с более длинной, специфической маской предпочтительнее общего. То есть если вы видите анонсы в сеть 130.95.0.0/**16** (то есть 130.95.0.0 - 130.95.255.255) через путь A и анонс в сеть 130.95.0.0/**24** (то есть 130.95.0.0 - 130.95.0.255) через путь B, то трафик, следующий к любому хосту сети 130.95.0.0/24, будет следовать путём B насколько бы ни был короче путь A.

Причины аварии

Итак, значит стоял этот маршрутизатор в AS 7007, изучал себе полную таблицу маршрутизации по BGP. Причины дальнейшего покрыты мраком (поговаривают о программном баге маршрутизатора). В какой-то момент он начал расщеплять все полученные маршруты на /24 и передавать их обратно своим BGP-соседям. Вся таблица маршрутизации Интернет, дезагрегированная до /24, выплеснулась в глобальную сеть. После операции разделения на /24 очистился атрибут AS-Path, и казалось, что все сети находятся в AS7007. Собственно, все в глобальной сети это приняли к сведению и начали пересылать весь свой трафик в эту автономку. Интернет существовал на конце 45-мегабитного канала, подключенного к AS7007. Корень проблемы ликвидировали быстро - отключили порт на сбойном маршрутизаторе и фальшивые анонсы от него прекратились. Но все остальные по-прежнему передавали по кругу полученные прежде анонсы. "Новая" таблица маршрутизации включала 250 000 записей. Маршрутизаторы не справлялись с такой нагрузкой, перезагружались, заново изучали маршруты от соседей, передавали их дальше и снова перезагружались. Анонсы носились по Интернету несколько часов со скоростью света. Магистральные провайдеры решали проблему кто как: выключали всё оборудование, порты, периодически перезагружали, добавляли фильтрацию на входе, чтобы восстановить внутреннюю связность сети. И так, через несколько часов равновесие восстановилось.
Из первых рук
Оригинал

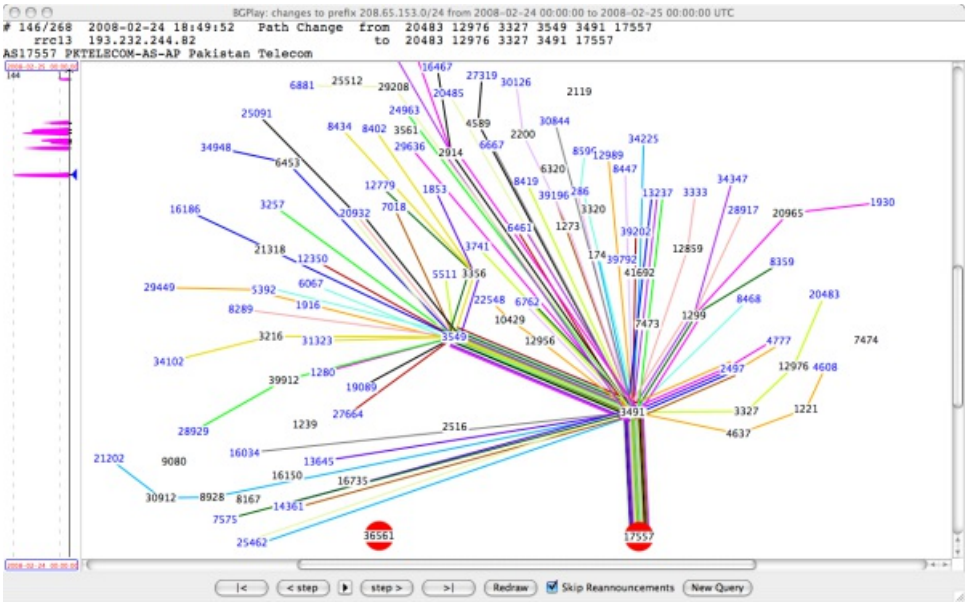
Пакистан заблокировал YouTube по всему миру

А вот событие, которое не прошло незамеченным даже для нового поколения пользователей.

24-го февраля 2008-го года, на полтора часа во всём мире пропал доступ к большей части Youtube. Предыстория такова, что правительство Пакистана постановило заблокировать доступ к этому видеосервису в связи с размещёнными там оскорбительными видеороликами. Естественно, рубанули сплеча - решили блокировать не URL или даже домен - заблокировали сразу блок IP-адресов. Пакистанский провайдер Pakistan Telecom в общем-то пошёл даже дальше, лишив Youtub'a весь мир.

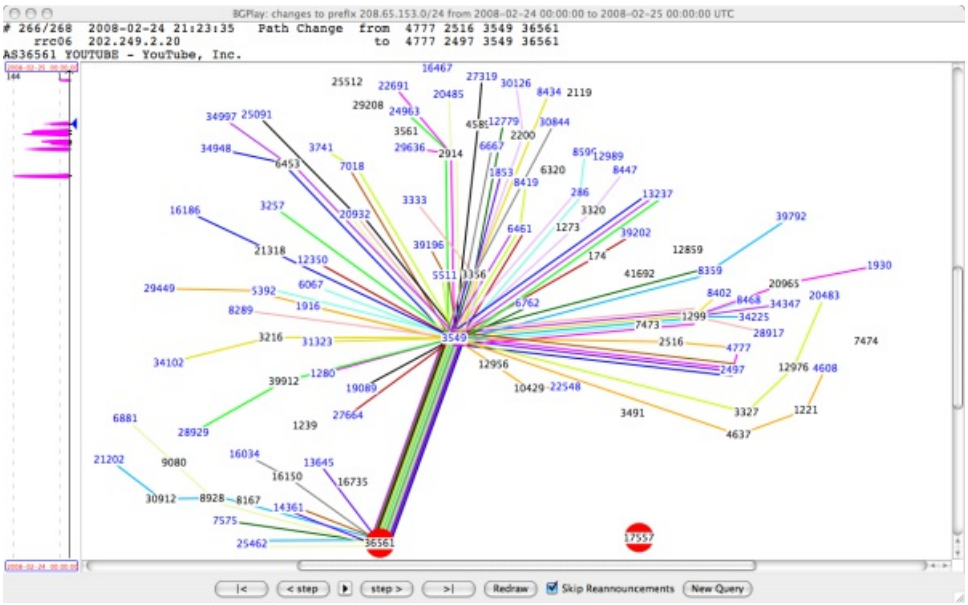
Полагаю, что они хотели завернуть трафик на какие-то свои внутренние ресурсы, но допустили утечку маршрутов в "большой" Интернет - дело-то плёвое. В то время как сам Youtube анонсировал сеть /22, Pakistan Telecom начал передавать своему соседу - магистральному провайдеру PCCW маршруты /24. Поминая все прошлые проблемы BGP, PCCW должен быть ограничивать анонсы своего клиента, но, увы, он этого не делал и фальшивый анонс благополучно мигрировал по сети. В течение трёх минут практически весь Интернет изучил новый маршрут.

Разумеется, маршрут /24 будет предпочтительнее /22, потому что он более специфичный (с более длинной маской). И почти весь трафик Youtube завернулся по новому маршруту. Вообще говоря - это тип атаки, который по-английски называется HiJacking. С полученным трафиком злоумышленник может делать что угодно: занульрутить, анализировать на предмет определённой информации, модифицировать его нужным образом и вернуть в сеть. Но судя по тому, что в PCCW помимо всех прочих клиентов, звонили и недоумевающие представители самого Pakistan Telecom, это была оплошность, ошибка конфигурации, толстые пальцы.



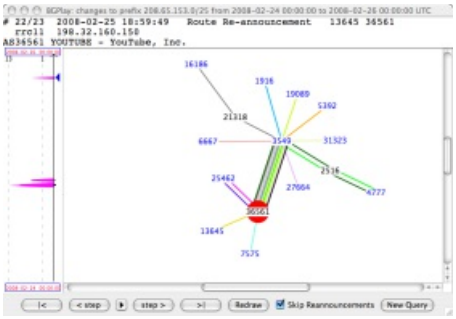
AS 17557 (Пакстан Телеком) анонсировал маршрут 208.65.153.0/24

Инженеры Youtube первыми приняли меры - через 1:20 после начала проблемы они тоже стали анонсировать сеть /24, после чего уже часть аудитории, смогла вернуться к спокойному просмотру роликов. Но, учитывая другую метрику пути в BGP - AS-Path, много где, доступа до сих пор не было. Например, для того же PCCW Pakistan Telecom казался ближе (соседняя AS), чем сервера Youtube, где-то на другом конце материка.



AS 36561 (YouTube) анонсировал маршрут 208.65.153.0/24

Ещё через 10 минут они начали передавать маршруты /25. Это могло бы стать хорошим временным решением. Но эти анонсы далеко не передались, видимо, ввиду того, что многие провайдеры отфильтровывают анонсы с масками длиннее /24. То есть по сути их узнали только ближайшие пиры и ситуация практически не изменилась.



AS 36561 (YouTube) анонсировал маршрут 208.65.153.0/25 и 208.65.153.128/25

Ещё через 20 минут очухались и сами Pakistan Telecom. Сначала они попытались тщетно решить проблему с помощью AS-Path Prepend, добавив к своим анонсам повторно номер своей AS. Это было видно по новым анонсам, пронёсшимся по глобальной сети. Окончательно проблема была решена больше, чем через 2 часа после начала, когда PCCW запретил анонсы данной сети и маршрут постепенно удалился со всех маршрутизаторов Интернета.
Визуализация происшествия
Более подробное описание аварии

Google's May 2005 Outage

Надо заметить, что тремя годами ранее, 7 мая 2005 года очень похожая ситуация с Google уже случалась. Эта авария известна как Google's May 2005 Outage и в Интернете больше связывается с неверной конфигурацией DNS-серверов. Однако инициативные ребята в то время по косвенным признакам решили, что некоторые симптомы характерны скорее для проблемы с маршрутизацией BGP. Они провели анализ данных на основе исторических данных, сохранившихся благодаря проекту RouteViews.

В подготовленном после анализа отчёте указано, что в тот момент Google имел две сети /19, и одну /22. Но, несмотря на это, анонсировал 68 блоков /24. Обычно это делается для того, чтобы путь трафика был как можно более чётким. Анонсы /25 уже блокируются многими операторами.

Но с 7 по 9 мая наблюдалась ситуация, когда один маршрут 64.233.161.0/24, который на самом деле принадлежит Google (AS 15169), был известен также через другую AS - 174, принадлежащую компании Cogent (никакого отношения к Google не имеющую). Большая часть AS, включая AS, непосредственно подключенные к Автономной Системе 15169, предпочли новый маршрут в AS 174.

Для пользователей это выглядело как редирект с сервисов google.com на другие, не имеющие отношения к Google ресурсы.

Инициативная группа, начавшая это расследование (*School of Computer Science Carleton University,*

Ottawa, Canada) связалась с Google и получила от них некоторые полезные данные. Во-первых, Google сообщил о своих проблемах с DNS, которые действительно имели место. Во-вторых, в связи с некорректной работой DNS, запросы на домены верхнего уровня перенаправлялись на другие. В-третьих, представители Google заявили, что они никак не связаны с Cogent и действительно испытывали проблемы с сетью 64.233.161.0/24 в тот период. В-четвёртых, инженерному персоналу Google было сообщено об уязвимостях BGP и возможных последствиях таких аварий.

В то же время Cogent очень неохотно шёл на контакт, и по сути ничего, кроме AS-Path, вовлечённых в инцидент, получено от них не было по предлогом соглашения о нераспространении приватной информации. Так или иначе, в этом случае установить точную причину не удалось. Несмотря на то, что возможны две причины: неверная конфигурация и намеренная атака - группа склоняется ко второй по нескольким косвенным причинам. Подробнейший отчёт о проблеме вы можете прочитать здесь.

Неудачная попытка бразильского провайдера CTBC анонсировать Full View от своего имени

Это пример удачного стечения обстоятельств.

Провайдер Companhia de Telecomunicacoes do Brasil Central – CTBC - вдруг решил своим двум пирам раздавать Full View, якобы от себя. Это могло вылиться в ещё одну проблему такого же масштаба, как инцидент с AS7007, обвалив весь Интернет. Но здесь отработали вышестоящие провайдеры, которые не пропустили эти анонсы дальше.

Кроме того отлично сработала система предупреждений об утечке маршрутов ("route leaks" and "route hijacks"), и Сетевые операторы быстро были в курсе проблемы. Однако совсем не замеченной эта проблема не прошла.

Авария на Яндексе

Нельзя не вспомнить и аварию на Яндексе двухлетней давности, но она коснулась исключительно Яндекса и исключительно по вине самого Яндекса.

В их случае из-за ошибки ПО или, что вполне вероятно, некорректной конфигурации, например, редистрибьюции, полная таблица BGP перкочевала в БД OSPF. Разумеется, при обсчёте SPF вся память устройства утекла, и оно вышло из строя. Маршрутная информация распространилась по сети и быстро уложила всю внутреннюю сеть Яндекса.

Сервисы восстанавливали шаг за шагом, подбираясь всё ближе к источнику проблемы. Комментарии представителя компании:

Заключение

Было ещё несколько крупных событий, которые не причинили сильно большого урона в масштабах Интернета. Было большое количество инцидентов среднего размера и бесчисленное множество мелких неприятностей с неправильными анонсами, которые никогда не выходили за пределы BGP-пира соседней AS.

Что изменилось после всех этих аварий? Провайдеры стали фильтровать анонсы от своих соседей и клиентов. От конечных заказчиков принимались только их собственные подсети, от равных по рангу соседей анонсы начали ограничиваться по определённым критериям. Многие вендоры начали встраивать в свои маршрутизаторы "магические возможности", позволяющие администратору тюнить BGP, например, сколько анонсов можно получить от соседа, прежде чем разорвать с ним сессию и заблокировать. Появились специальные организации и сервисы, которые занимаются мониторингом сети Интернет:

- BGPmon
- Renesys
- RouteViews

Сейчас Сетевые Операторы могут получать уведомления о каких-то важных событиях в Интернет, если они связаны с их префиксами или AS, как и произошло в случае с CTBC или недавний инцидент с Китаем.

Ещё одна ветвь развития - разработка новой редакции BGP, лишённой такого серьёзного недостатка. Такой подход подразумевает проверку того, что анонсы, получаемые от данного соседа, действительно от него должны быть получены. То есть подсети Apple едва ли должны анонсироваться из какой-то провинции Китая.

На сегодняшний день существует две конкурирующие реализации: **S-BGP** (Secure-BGP) и **soBGP** (secure origin BGP). Но это уже совсем другая история. Чудесный документ на тему безопасности BGP. Берегите свой BGP.