

Network Working Group
Request for Comments: 1546
Category: Informational

C. Partridge
T. Mendez
W. Milliken
BBN
November 1993

Host Anycasting Service

Статус документа

В этом документе содержится информация для сообщества Internet. Документ не задает каких-либо стандартов Internet. Документ может распространяться без ограничений.

Тезисы

В этом RFC описаны anycast-услуги для протокола IP. Основной задачей документа является рассмотрение семантики anycast-услуг в межсетевой среде IP. Там, где это возможно, документ опирается на практику. Документ описывает экспериментальные услуги и не задает какого-либо протокола. Данный документ разработан IRTF¹.

Мотивация

В сети возникает множество ситуаций, когда хост, приложение или пользователь желает найти хост, поддерживающий определенный сервис, но при наличии нескольких серверов, поддерживающих нужную услугу, выбор конкретного сервера не имеет значения. Технология anycast обеспечивает возможность реализации таких услуг в межсетевой среде. Хост передает дейтаграмму по anycast-адресу и межсетевая среда принимает на себя ответственность за попытку доставки этой дейтаграммы по крайней мере (и предпочтительно) одному из серверов, принимающих дейтаграммы для anycast-адреса.

Основным мотивом разработки такой технологии является существенное упрощение поиска подходящего сервера. Например, пользователи вместо поиска серверовarchie и выбора ближайшего могут просто ввести команду

```
telnetarchie.net
```

и подключиться к ближайшему серверуarchie. Преобразователям DNS больше не нужно будет указывать IP-адреса их серверов, они смогут просто направлять запрос по общеизвестному anycast-адресу DNS. Сайты FTP, имеющие зеркала, смогут просто работать по одному anycast-адресу, а пользователи будут просто подключаться к этому адресу по протоколу FTP для соединения с ближайшим сервером.

Архитектурные вопросы

Добавление технологии anycast в спектр услуг IP требует некоторого обсуждения в части согласования архитектурных требований IP с требованиями anycast. Эти вопросы обсуждаются в данном параграфе.

Первым и наиболее важным является согласование услуг IP без поддержки данных о состоянии (stateless) с желанием иметь anycast-адрес, представляющий один виртуальный хост. Лучше всего будет рассмотреть этот вопрос на примерах. В обоих рассматриваемых случаях два хоста (X и Y) используют anycast-адрес, а третий хост (Z) обращается по этому адресу для использования услуг.

В первом примере предположим, что Z передает дейтаграмму UDP по anycast-адресу. Поскольку anycast-адрес логически рассматривается как адрес одного виртуального хоста, возникает вопрос — может ли дейтаграмма быть доставлена на оба хоста X и Y? Очевидно, что ответ на этот вопрос будет положительным — доставка на оба хоста X и Y допускается. Протокол IP разрешает дублировать дейтаграммы и направлять их по разным маршрутам, поэтому очевидна возможность доставки одной дейтаграммы на оба хоста X и Y. Это позволяет определить anycasting в среде IP, как доставку по возможности² anycast-дейтаграммы одному, а возможно и множеству хостов, которые обслуживают данный anycast-адрес.

Во втором примере предположим, что Z передает по anycast-адресу две дейтаграммы и первая доставляется хосту X. Какой из хостов (X или Y) получит вторую дейтаграмму? Для протоколов, поддерживающих состояние (типа TCP) предпочтительна доставка всех дейтаграмм соединения одному хосту. Однако протокол IP, как таковой, не поддерживает информации о соединениях (и, следовательно, не отслеживает, куда были доставлены предшествующие дейтаграммы). Поскольку одной из задач anycast является поддержка реплицированных услуг, представляется очевидным, что вторая дейтаграмма может быть доставлена как хосту X, так и хосту Y. Протоколы с организацией соединений будут реализовать некие дополнительные механизмы для обеспечения доставки последующих дейтаграмм тому же хосту, который получил первую. Предложения по реализации такого механизма в TCP приводятся ниже.

После рассмотрения этих примеров представляется корректным определить IP anycasting, как сервис, обеспечивающий доставку anycast-дейтаграмм по мере возможностей без поддержки информации о соединениях по крайней мере и предпочтительно одному из хостов, обслуживающих данный anycast-адрес. Из этого определения ясно, что anycast-дейтаграммы обслуживаются в сети тем же, как обычные дейтаграммы IP. Хотя определение позволяет доставлять дейтаграммы множеству хостов, ясно, что целью является доставка только одному хосту.

¹Internet Research Task Force — Группа по исследованию Internet.

²В оригинале «best effort delivery». Прим. перев.

Anycast-адреса

Существует множество путей поддержки anycast-адресов — для некоторых требуется выделение небольших частей существующего адресного пространства, для других — выделение специального класса адресов IP.

Основным преимуществом использования существующего адресного пространства является то, что такой подход позволяет упростить маршрутизацию. В качестве примера рассмотрим случай, когда часть каждой IP-сети может использоваться для поддержки anycast (т. е., сайт при желании может выделить некоторое множество своих подсетей для anycast-адресации). Если, как предполагают некоторые эксперты, anycast-маршруты трактуются протоколами маршрутизации просто как маршруты к хостам, для адресов anycast не потребуется специального анонсирования за пределами сайта — маршруты к хостам могут включаться в маршруты к сетям (если anycast-адреса поддерживаются внешними по отношению к сети хостами, такие хосты будут по-прежнему анонсироваться с использованием маршрутов к хостам). Основными недостатками такой модели являются (1) отсутствие простого способа отнесения адресов к типу anycast для основанных на соединениях протоколах типа TCP и (2) сложность поддержки общеизвестных anycast-адресом в межсетевой среде. Причины, по которым протоколу TCP требуется знать, относится ли адрес к типу anycast, более подробно рассмотрены ниже. Вопрос поддержки общеизвестных anycast-адресов требует некоторых дополнительных пояснений. Идея заключается в том, что использовать в сети Internet адреса anycast в качестве логических адресов серверов DNS. Тогда программы на хостах могут быть настроены производителем так, чтобы всегда отправлять запросы DNS по anycast-адресам DNS. Иными словами, технологию anycast можно использовать для поддержки автоматической настройки конфигурации преобразователей DNS.

Основным преимуществом использования отдельного класса адресов является простота идентификации anycast-адресов и поддержки списка общеизвестных адресов. Важнейшим недостатком является усложнение маршрутизации, обусловленное необходимостью отслеживать множество anycast-маршрутов в протоколах маршрутизации.

Промежуточным вариантом является выделение части адресного пространства (скажем, 256 сетей класса C) с использованием адресов сетей в качестве anycast-адресов (и игнорированием связанной с хостом части адреса класса C). Преимущество такого решения заключается в том, что маршруты anycast выглядят, как обычные маршруты в сети (такие маршруты проще в обслуживании для некоторых протоколов маршрутизации). Недостатками являются неэффективное использование адресного пространства и достаточно жесткие ограничения на количество поддерживаемых адресов anycast.

С учетом сказанного выше представляется разумным использование отдельного класса адресов. Выделение адресов anycast из существующего адресного пространства представляется связанным с большим числом проблем, обусловленных тем, что приложения не смогут распознавать anycast-адреса (если адреса anycasts занимают часть адресного пространства каждого сайта) или адресное пространство будет использоваться неэффективно (если в качестве адресов anycast используются адреса сетей). Преимущества использования anycast-адресов для автоматической настройки конфигурации представляются несомненными. Поэтому в данном документе предполагается, что для адресации anycast будет использоваться отдельный класс адресов IP (еще не выделен¹). Поскольку каждый anycast-адрес является адресом виртуального хоста и число хостов anycast представляется не превышающим числа служб, предоставляемых протоколами типа TCP и UDP, размер требуемого адресного пространства достаточно мал, возможно, не более 2^{16} уникальных адресов.

Передача и прием anycast-дейтаграмм

Исторически услуги IP разрабатывались с учетом сохранения работоспособности при отсутствии маршрутизаторов (например, в ЛВС без маршрутизации). Более того, значительная часть сообщества Internet до сих пор считает, что хостам не следует принимать участия в работе протоколов маршрутизации (см., например, стр. 7 документа STD 3, RFC 1122). Для обеспечения anycast-услуг, совместимых с этими представлениями, обслуживание адресов anycast должно незначительно отличаться для разнотипных сетей, в которых передаются дейтаграммы с anycast-адресами.

В сетях с разделяемой средой типа Ethernet или Token Ring должна обеспечиваться возможность передачи anycast-дейтаграммы серверу той же сети без использования маршрутизатора (например, по причине его отсутствия). Для решения этой задачи существует по крайней мере два пути.

Одним из вариантов является использование ARP для адресов anycast. Серверы, которые поддерживают anycast-адреса, могут отвечать на запрос ARP и передающий запрос хост может связаться с первым ответившим сервером. Этот вариант напоминает расширение ARP (RFC 1027) и, подобно этому расширению, требует, чтобы тайм-аут кэша ARP для адресов anycast был достаточно малым (порядка 1 минуты), чтобы при отключении сервера anycast хосты достаточно быстро сбрасывали для него запись ARP и запрашивали другие серверы, поддерживающие anycast.

Другим вариантом является передача anycast-дейтаграмм с использованием групповой адресации на канальном уровне. Предполагается, что хосты, обслуживающие anycast-адреса, будут прослушивать групповые адреса канального уровня на предмет дейтаграмм, направленных по anycast-адресу. При использовании групповой адресации в локальной сети для доставки anycast-дейтаграмм маршрутизатор не требуется. Одним из достоинств этого варианта является то, что даже при отключении множества серверов оставшиеся продолжают получать все запросы. Другое потенциальное преимущество заключается в том, что не возникает требования быстрого старения записей ARP для адресов anycast, приводящего к росту избыточного служебного трафика. В варианте ARP передача по любому адресу anycast подобна широковещательному запросу ARP, тогда как при использовании групповой адресации передача осуществляется только в одну группу. Очевидным недостатком является то, что при наличии в сети множества серверов, они все будут получать пакеты anycast, тогда как желательна доставка единственному серверу.

На каналах «точка-точка» поддержка anycast реализуется проще. Одна копия anycast-дейтаграммы передается в подходящий канал, ведущий к адресату anycast.

Когда маршрутизатор получает anycast-дейтаграмму, он должен принять решение о целесообразности ее пересылки и при положительном решении передать копию дейтаграммы на следующий интервал маршрута. Отметим, что несмотря на предположение о том, что маршрутизатор всегда знает следующий интервал для пересылки anycast-дейтаграммы и не будет пересылать ее по групповым адресам локальной сети, возможны ситуации, когда в локальной сети будет несколько серверов и для предотвращения передачи дейтаграммы тому серверу, который в данное время не работает, маршрутизаторы могут передавать anycast-дейтаграммы по групповым адресам канального уровня. Поскольку хосты

¹Такой класс не был выделен и от концепции классов адресов к настоящему времени отказались вообще. *Прим. перев.*

могут передавать по групповым адресам любые дейтаграммы, маршрутизаторам следует принимать меры по предотвращению пересылки дейтаграмм, которые предполагаются пересланными другим маршрутизатором.

Хосты, желающие получать дейтаграммы, направленные по определенному anycast-адресу, будут анонсировать маршрутизаторам свою вовлеченность в обслуживание такого адреса. В сетях с разделяемой средой лучшим способом анонсирования возможно является периодическая передача хостом в использовании групповой адресации сведений о поддерживаемых anycast-адресах (возможно, с использованием расширенной версии IGMP). Групповая передача сообщений гарантирует, что все маршрутизаторы в сети будут знать, что адрес anycast поддерживается в локальной подсети и могут анонсировать этот факт соседним маршрутизаторам, когда это целесообразно. Отметим, что при отсутствии в подсети маршрутизаторов групповые сообщения будут просто игнорироваться (групповая адресация предложена по причине ее простоты и более высокой надежности по сравнению с разработкой протокола регистрации, с помощью которого anycast-серверы должны будут регистрировать себя на каждом маршрутизаторе своей локальной сети).

В сетях «точка-точка» хост может просто анонсировать свои anycast-адреса маршрутизатору на другом конце канала.

Отметим, что протоколы анонсирования являются формой протоколов маршрутизации и может быть осмысленным требование участия серверов anycast (по крайней мере, частично) в обмене обычными маршрутными сообщениями.

Когда хост получает дейтаграмму IP, направленную по поддерживаемому им адресу anycast, этому хосту следует трактовать такую дейтаграмму IP, как обычную дейтаграмму, направленную по одному из обычных (не-anycast) IP-адресов этого хоста. Если хост не поддерживает anycast-адрес, ему следует просто отбросить такую дейтаграмму.

Хостам следует принимать дейтаграммы, отправленные с anycast-адресов, хотя некоторые транспортные протоколы (см. ниже) могут требовать отказа от приема таких дейтаграмм.

Как UDP и TCP используют anycast

Важно помнить, что услуги anycast предоставляются без поддержки информации о состоянии. Межсетевая среда не берет на себя обязательств по доставке двух последовательных пакетов, направленных по одному anycast-адресу, одному и тому же хосту.

Поскольку протокол UDP, как и anycast, не поддерживает данных о состоянии, UDP может трактовать адреса anycast, подобно обычным адресам IP. Дейтаграмма UDP, переданная по anycast-адресу, подобна обычной (unicast) дейтаграмме с точки зрения протокола UDP и приложений. Дейтаграмма UDP с anycast-адреса подобна дейтаграмме с обычного адреса. Более того, дейтаграмма с одного anycast-адреса на другой может трактоваться протоколом UDP, как обычная (unicast) дейтаграмма (семантика использования таких дейтаграмм не вполне понятна).

Использование технологии anycast в TCP менее очевидно, поскольку протокол TCP основан на соединениях. Сложно представить, как может поддерживаться состояние TCP для anycast-партнера, если два последовательных сегмента TCP, переданные по anycast-адресу, могут попасть на совершенно разные хосты.

Решением проблемы может послужить использование anycast-адресов получателей только для сегментов TCP SYN (без флага ACK). Тогда TCP может инициировать соединение с anycast-адресом. После передачи получившим anycast-сегмент хостом ответного сегмента SYN-ACK, вызывающему модуль TCP следует заменить anycast-адрес партнера адресом из принятого пакета SYN-ACK (вызывающий модуль TCP может распознать соединение, к которому относится пакет SYN-ACK, трактуя anycast-адрес, как шаблон, которому соответствует любой входящий сегмент SYN-ACK с корректным портом и адресом получателя, а также портом отправителя, что дает полный адрес SYN-ACK, включая адрес отправителя, не соответствующий другим соединениям, а порядковые номера в пакете SYN-ACK будут корректными). Такой вариант позволяет TCP после получения SYN-ACK продолжить обмен данными с одним хостом.

Приложения и anycast

В общем случае приложения используют адреса anycast, подобно обычным адресам IP. Беспорядок могут вызывать только приложения, пытающиеся организовать поддержку состояния для UDP, и приложения, которые пытаются поддерживать состояния для множества соединений TCP. Поскольку технология anycast не поддерживает данных о состоянии и не гарантирует доставки множества anycast-дейтаграмм одной системе, приложение не может быть уверенным в том, что оно взаимодействует с одним партнером для двух последовательных передач UDP или двух последовательных соединений TCP по одному и тому же адресу anycast.

Очевидным решением этих проблем является требование к приложениям, желающим поддерживать данные о состоянии, определять обычный (unicast) адрес своего партнера при первом обмене дейтаграммами UDP или в первом соединении TCP и в дальнейшем использовать именно этот адрес.

Anycast и групповая адресация

Часто предлагают использовать для определения местоположения ресурсов групповую адресацию IP, поэтому полезно будет сравнить групповую адресацию и технологию IP anycast.

Семантически различие между этими вариантами адресации заключается в том, что адрес anycast является адресом одного (виртуального) хоста и межсетевая среда будет пытаться доставить anycast-дейтаграммы одному хосту. Это различие влечет за собой два следствия. Во-первых, приложениям, передающим пакеты по anycast-адресам, не нужно заботиться о контроле значения TTL в дейтаграммах IP. Приложения, использующие для поиска сервиса групповую адресацию, должны балансировать значением TTL для повышения вероятности нахождения сервиса и снижения вероятности передачи дейтаграммы множеству серверов одновременно. Во-вторых, организация соединения TCP с anycast-адресом вполне осмысленна, тогда как организация соединений TCP с групповым адресом не вполне очевидна (соединение TCP с групповым адресом, по сути, является попыткой организации соединения со множеством партнеров одновременно, а протокол TCP не рассчитан на поддержку таких соединений).

С практической точки зрения основное различие заключается в том, что технология anycast использует обычную (unicast) адресацию, а для поддержки групповой адресации ведет к усложнению маршрутизации. Важно отметить, что множество маршрутов к anycast-адресату маршрутизатор воспринимает как множество маршрутов к индивидуальному получателю и может использовать стандартные алгоритмы выбора лучшего маршрута.

Другим отличием является то, что для поиска ресурсов с использованием групповой адресации обычно требуется передать большее число дейтаграмм. Предполагается, что для поиска сервера с использованием групповой адресации приложение будет передавать дейтаграммы, увеличивая значение IP TTL. Начальное значение TTL устанавливается достаточно малым, чтобы ограничить число потенциально доступных серверов. Однако при отсутствии отклика от серверов значение TTL должно быть увеличено, исходя из предположения о том, что доступные серверы оказались за пределами радиуса начального TTL. В результате для поиска ресурсов с использованием групповой адресации требуется передать одну или более дейтаграмм по групповым адресам в направлении множества серверов. При этом часть дейтаграмм не достигнет серверов по причине малого значения TTL. При использовании технологии anycast изменение TTL не требуется и поэтому (без учета возможных потерь) для нахождения сервера достаточно передать одну дейтаграмму. Более того, эта дейтаграмма пойдет по единственному пути.

Небольшое различие между рассматриваемыми вариантами заключается в том, что технология anycast может оказаться менее устойчивой к отказам по сравнению с групповой адресацией. При отказе anycast-сервера некоторые дейтаграммы могут ошибочно маршрутизироваться неработающему серверу, а при использовании групповой адресации дейтаграммы получат другие серверы.

Смежные работы

Протокол ARPANET AHIP-E Host Access, описанный в RFC 878, поддерживает логическую адресацию, которая позволяет нескольким хостам использовать общий логический адрес. Такая схема может использоваться для поддержки anycast в подсети PSN.

Вопросы безопасности

Существуют по крайней мере две проблемы безопасности, связанные с технологией anycast. Эти проблемы просто указаны ниже без рассмотрения путей их разрешения.

Во-первых, очевидно, что враждебные хосты, могут несанкционированно обслуживать адреса anycast, перехватывая anycast-дейтаграммы, направленные легитимным серверам.

Во-вторых, «подслушивающие» хосты могут отвечать на запросы anycast, передавая ложную информацию. Поскольку не существует способа проверки принадлежности anycast-адреса, отсутствует возможность достоверного обнаружения фактов получения ответа от подслушивающего хоста вместо ответа от легитимного сервера.

Благодарности

При создании этого документа важную роль сыграли комментарии Steve Deering, Paul Francis, Christian Huitema, Greg Minshall, Jon Postel, Ram Ramanathan и Bill Simpson. Однако всю ответственность за изложенные здесь идеи авторы принимают на себя.

Адреса авторов

Craig Partridge

Bolt Beranek and Newman
10 Moulton St
Cambridge MA 02138
EMail: craig@bbn.com

Trevor Mendez

Bolt Beranek and Newman
10 Moulton St
Cambridge MA 02138
EMail: tmendez@bbn.com

Walter Milliken

Bolt Beranek and Newman
10 Moulton St
Cambridge MA 02138
EMail: milliken@bbn.com

Перевод на русский язык

Николай Малых
nmalykh@gmail.com