

Тема 2. Требования, предъявляемые к компьютерным сетям

Соответствие стандартам — это только одно из многих требований, предъявляемых к современным сетям. В этом разделе мы остановимся на некоторых других, не менее важных.

Самое общее пожелание, которое можно высказать в отношении работы сети — это выполнение сетью того набора услуг, для оказания которых она предназначена: например, предоставление доступа к файловым архивам или страницам публичных Web-сайтов Internet, обмен электронной почтой в пределах предприятия или в глобальных масштабах, интерактивный обмен голосовыми сообщениями IP-телефонии и т.п.

Все остальные требования — производительность, надежность, совместимость, управляемость, защищенность, расширяемость и масштабируемость — связаны с качеством выполнения этой основной задачи. И хотя все перечисленные выше требования весьма важны, часто понятие "качество обслуживания" (Quality of Service, QoS) компьютерной сети трактуется более узко: в него включаются только две самые важные характеристики сети — производительность и надежность.

Производительность

Потенциально высокая производительность — это одно из основных преимуществ распределенных систем, к которым относятся компьютерные сети. Это свойство обеспечивается принципиальной, но, к сожалению, не всегда практически реализуемой возможностью распределения работ между несколькими компьютерами сети.

Основные характеристики производительности сети:

- время реакции;
- скорость передачи трафика;
- пропускная способность;
- задержка передачи и вариация задержки передачи.

Время реакции сети является интегральной характеристикой производительности сети с точки зрения пользователя. Именно эту характеристику имеет в виду пользователь, когда говорит: "Сегодня сеть работает медленно".

В общем случае время реакции определяется как интервал между возникновением запроса пользователя к какой-либо сетевой службе и получением ответа на него.

Очевидно, что значение этого показателя зависит от типа службы, к которой обращается пользователь, от того, какой пользователь и к какому серверу обращается, а также от текущего состояния элементов сети —

загруженности сегментов, коммутаторов и маршрутизаторов, через которые проходит запрос, загруженности сервера и т.п.

Поэтому имеет смысл использовать также и средневзвешенную оценку времени реакции сети, усредняя этот показатель по пользователям, серверам и времени дня (от которого в значительной степени зависит загрузка сети).

Время реакции сети обычно складывается из нескольких составляющих. В общем случае в него входит:

- время подготовки запросов на клиентском компьютере;
- время передачи запросов между клиентом и сервером через сегменты сети и промежуточное коммуникационное оборудование;
- время обработки запросов на сервере;
- время передачи ответов от сервера клиенту и время обработки получаемых от сервера ответов на клиентском компьютере.

Очевидно, что разложение времени реакции на составляющие пользователя не интересует — ему важен конечный результат. Однако для сетевого специалиста очень важно выделить из общего времени реакции составляющие, соответствующие этапам собственно сетевой обработки данных, — передачу данных от клиента к серверу через сегменты сети и коммуникационное оборудование.

Знание сетевых составляющих времени реакции позволяет оценить производительность отдельных элементов сети, выявить узкие места и при необходимости выполнить модернизацию сети для повышения ее общей производительности.

Производительность сети может характеризоваться также *скоростью передачи* трафика.

Скорость передачи трафика может быть мгновенной, максимальной и средней:

- средняя скорость вычисляется путем деления общего объема переданных данных на время их передачи, причем выбирается достаточно длительный промежуток времени — час, день или неделя;
- мгновенная скорость отличается от средней тем, что для усреднения выбирается очень маленький промежуток времени — например, 10 мс или 1 с;
- максимальная скорость — это наибольшая скорость, зафиксированная в течение периода наблюдения.

Чаще всего при проектировании, настройке и оптимизации сети используются такие показатели, как средняя и максимальная скорость.

Средняя скорость, с которой обрабатывает трафик отдельный элемент или сеть в целом, позволяет оценить работу сети на протяжении длительного времени, в течение которого в силу закона больших чисел пики и спады интенсивности трафика компенсируют друг друга. Максимальная скорость позволяет оценить, как сеть будет справляться с пиковыми нагрузками, характерными для особых периодов работы, например в утренние часы, когда сотрудники предприятия почти одновременно регистрируются в сети и обращаются к разделяемым файлам и базам данных. Обычно при определении скоростных характеристик некоторого сегмента или устройства в передаваемых данных не выделяется трафик какого-то определенного пользователя, приложения или компьютера — подсчитывается общий объем передаваемой информации. Тем не менее, для более точной оценки качества обслуживания такая детализация желательна, и в последнее время системы управления сетями все чаще позволяют ее выполнять.

Пропускная способность — максимально возможная скорость обработки трафика, определенная стандартом технологии, на которой построена сеть. Пропускная способность отражает максимально возможный объем данных, передаваемый сетью или ее частью в единицу времени.

Пропускная способность уже не является, подобно времени реакции или скорости прохождения данных по сети, пользовательской характеристикой, так как она говорит о скорости выполнения внутренних операций сети — передачи пакетов данных между узлами сети через различные коммуникационные устройства. Зато она непосредственно характеризует качество выполнения основной функции сети — транспортировки сообщений — и поэтому чаще используется при анализе производительности сети, чем время реакции или скорость.

Пропускная способность измеряется либо в битах в секунду, либо в пакетах в секунду.

Пропускная способность сети зависит как от характеристик физической среды передачи (медный кабель, оптическое волокно, витая пара) так и от принятого способа передачи данных (технология Ethernet, FastEthernet, ATM). Пропускная способность часто используется в качестве характеристики не столько сети, сколько собственно технологии, на которой построена сеть. Важность этой характеристики для сетевой технологии показывает, в частности, и то, что ее значение иногда становится частью названия, например, 10 Мбит/с Ethernet, 100 Мбит/с Ethernet.

В отличие от времени реакции или скорости передачи трафика пропускная способность не зависит от загруженности сети и имеет постоянное значение, определяемое используемыми в сети технологиями.

На разных участках гетерогенной сети, где используется несколько разных технологий, пропускная способность может быть различной. Для анализа и настройки сети очень полезно знать данные о пропускной способности отдельных ее элементов. Важно отметить, что из-за последовательного характера передачи данных различными элементами сети общая пропускная способность любого составного пути в сети будет равна минимальной из пропускных способностей составляющих элементов маршрута. Для повышения пропускной способности составного пути необходимо в первую очередь обратить внимание на самые медленные элементы. Иногда полезно оперировать общей пропускной способностью сети, которая определяется как среднее количество информации, переданной между всеми узлами сети за единицу времени. Этот показатель характеризует качество сети в целом, не дифференцируя его по отдельным сегментам или устройствам.

Задержка передачи определяется как задержка между моментом поступления данных на вход какого-либо сетевого устройства или части сети и моментом появления их на выходе этого устройства.

Этот параметр производительности по смыслу близок ко времени реакции сети, но отличается тем, что всегда характеризует только сетевые этапы обработки данных, без задержек обработки конечными узлами сети.

Обычно качество сети характеризуют величинами максимальной задержки передачи и вариацией задержки. Не все типы трафика чувствительны к задержкам передачи, во всяком случае, к тем величинам задержек, которые характерны для компьютерных сетей, — обычно задержки не превышают сотен миллисекунд, реже — нескольких секунд. Такого порядка задержки пакетов, порождаемых файловой службой, службой электронной почты или службой печати, мало влияют на качество этих служб с точки зрения пользователя сети. С другой стороны, такие же задержки пакетов, переносящих голосовые или видеоданные, могут приводить к значительному снижению качества предоставляемой пользователю информации — возникновению эффекта "эха", невозможности разобрать некоторые слова, вибрации изображения и т. п.

Все указанные характеристики производительности сети достаточно независимы. В то время как пропускная способность сети является постоянной величиной, скорость передачи трафика может варьироваться в зависимости от загрузки сети, не превышая, конечно, предела, устанавливаемого пропускной способностью. Так в односегментной сети 10 Мбит/с Ethernet компьютеры могут обмениваться данными со скоростями 2 Мбит/с и 4 Мбит/с, но никогда — 12 Мбит/с.

Пропускная способность и задержки передачи также являются независимыми параметрами, так что сеть может обладать, например, высокой пропускной способностью, но вносить значительные задержки при передаче каждого пакета. Пример такой ситуации дает канал связи, образованный геостационарным спутником. Пропускная способность этого канала может быть весьма высокой, например 2 Мбит/с, в то время как задержка передачи всегда составляет не менее 0,24 с, что определяется скоростью распространения электрического сигнала (около 300000 км/с) и длиной канала (72000 км).

Надежность и безопасность

Одна из первоначальных целей создания распределенных систем, к которым относятся и вычислительные сети, состояла в достижении большей надежности по сравнению с отдельными вычислительными машинами.

Важно различать несколько аспектов надежности.

Для сравнительно простых технических устройств используются такие показатели надежности, как:

- среднее время наработки на отказ;
- вероятность отказа;
- интенсивность отказов.

Однако эти показатели пригодны для оценки надежности простых элементов и устройств, которые могут находиться только в двух состояниях — работоспособном или неработоспособном. Сложные системы, состоящие из многих элементов, кроме состояний работоспособности и неработоспособности, могут иметь и другие промежуточные состояния, которые эти характеристики не учитывают.

Для оценки надежности сложных систем применяется другой набор характеристик:

- готовность или коэффициент готовности;
- сохранность данных;
- согласованность (непротиворечивость) данных;
- вероятность доставки данных;
- безопасность;
- отказоустойчивость.

Готовность или коэффициент готовности (availability) означает период времени, в течение которого система может использоваться. Готовность может быть повышена путем введения избыточности в структуру системы: ключевые элементы системы должны существовать в нескольких

экземплярах, чтобы при отказе одного из них функционирование системы обеспечивали другие.

Чтобы компьютерную систему можно было считать высоконадежной, она должна как минимум обладать высокой готовностью, но этого недостаточно. Необходимо обеспечить сохранность данных и защиту их от искажений. Кроме того, должна поддерживаться согласованность (непротиворечивость) данных, например если для повышения надежности на нескольких файловых серверах хранится несколько копий данных, то нужно постоянно обеспечивать их идентичность.

Так как сеть работает на основе механизма передачи пакетов между конечными узлами, одной из характеристик надежности является вероятность доставки пакета узлу назначения без искажений. Наряду с этой характеристикой могут использоваться и другие показатели: вероятность потери пакета (по любой из причин — из-за переполнения буфера маршрутизатора, несовпадения контрольной суммы, отсутствия работоспособного пути к узлу назначения и т. д.), вероятность искажения отдельного бита передаваемых данных, соотношение количества потерянных и доставленных пакетов.

Другим аспектом общей надежности является безопасность (security), то есть способность системы защитить данные от несанкционированного доступа. В распределенной системе это сделать гораздо сложнее, чем в централизованной. В сетях сообщения передаются по линиям связи, часто проходящим через общедоступные помещения, в которых могут быть установлены средства прослушивания линий. Другим уязвимым местом могут стать оставленные без присмотра персональные компьютеры. Кроме того, всегда имеется потенциальная угроза взлома защиты сети от неавторизованных пользователей, если сеть имеет выходы в глобальные общедоступные сети.

Еще одной характеристикой надежности является отказоустойчивость (fault tolerance). В сетях под отказоустойчивостью понимается способность системы скрыть от пользователя отказ отдельных ее элементов. Например, если копии таблицы базы данных хранятся одновременно на нескольких файловых серверах, пользователи могут просто не заметить отказа одного из них. В отказоустойчивой системе выход из строя одного из ее элементов приводит к некоторому снижению качества ее работы (деградации), а не к полному останову. Так, при отказе одного из файловых серверов в предыдущем примере увеличивается только время доступа к базе данных из-за уменьшения степени распараллеливания запросов, но в целом система будет продолжать выполнять свои функции.