

Савва Юрий Болеславович

ПРОЕКТИРОВАНИЕ РАСПРЕДЕЛЕННЫХ ИНФОРМАЦИОННЫХ СИСТЕМ

Лекция 15

Тема:

Оптимизация локальных сетей

Основные задачи оптимизации локальных сетей

- Для того, чтобы локальная сеть работала самым эффективным образом, то системному администратору необходимо решить следующие задачи:
 1. Сформулировать *критерии эффективности* работы сети. Чаще всего такими критериями служат производительность и надежность, для которых в свою очередь требуется выбрать конкретные показатели оценки, например, время реакции и коэффициент готовности, соответственно.

2. Определить множество *варьируемых параметров* сети, прямо или косвенно влияющих на критерии эффективности. Эти параметры действительно должны быть варьируемыми, то есть нужно убедиться в том, что их можно изменять в некоторых пределах по вашему желанию.
- Так, если размер пакета какого-либо протокола в конкретной операционной системе устанавливается автоматически и не может быть изменен путем настройки, то этот параметр в данном случае не является варьируемым, хотя в другой операционной системе он может относиться к изменяемым по желанию администратора, а значит и варьируемым.

- Другим примером может служить пропускная способность внутренней шины маршрутизатора - она может рассматриваться как параметр оптимизации только в том случае, если вы допускаете возможность замены маршрутизаторов в сети.

Все варьируемые параметры могут быть сгруппированы различным образом.

- Например, параметры отдельных конкретных протоколов (максимальный размер кадра протокола Ethernet или размер окна неподтвержденных пакетов протокола ТСР) или параметры устройств (размер адресной таблицы или скорость фильтрации моста, пропускная способность внутренней шины маршрутизатора).

Параметрами настройки могут быть и устройства, и протоколы в целом.

- Так, например, улучшить работу сети с медленными и зашумленными глобальными каналами связи можно, перейдя со стека протоколов IPX/SPX на протоколы TCP/IP.
- Также можно добиться значительных улучшений с помощью замены сетевых адаптеров неизвестного производителя на адаптеры BrandName.

3. Определить *порог чувствительности* для значений критерия эффективности.

Так, производительность сети можно оценивать логическими значениями "Работает"/ "Не работает", и тогда оптимизация сводится к диагностике неисправностей и приведению сети в любое работоспособное состояние.

Другим крайним случаем является тонкая настройка сети, при которой параметры работающей сети (например, размер кадра или величина окна неподтвержденных пакетов) могут варьироваться с целью повышения производительности (например, среднего значения времени реакции) хотя бы на несколько процентов.

Как правило, **под оптимизацией сети понимают некоторый промежуточный вариант, при котором требуется выбрать такие значения параметров сети, чтобы показатели ее эффективности существенно улучшились**, например, пользователи получали ответы на свои запросы к серверу баз данных не за 10 секунд, а за 3 секунды, а передача файла на удаленный компьютер выполнялась не за 2 минуты, а за 30 секунд.

Таким образом, можно предложить три различных трактовки задачи оптимизации:

1. Приведение сети в любое работоспособное состояние. Обычно эта задача решается первой, и включает:

- поиск неисправных элементов сети - кабелей, разъемов, адаптеров, компьютеров;
- проверку совместимости оборудования и программного обеспечения;
- выбор корректных значений ключевых параметров программ и устройств, обеспечивающих прохождение сообщений между всеми узлами сети - адресов сетей и узлов, используемых протоколов, типов кадров Ethernet и т.п.

2. Грубая настройка - выбор параметров, резко влияющих на характеристики (надежность, производительность) сети.

Если сеть работоспособна, но обмен данными происходит очень медленно (время ожидания составляет десятки секунд или минуты) или же сеанс связи часто разрывается без видимых причин, то работоспособной такую сеть можно назвать только условно, и она безусловно нуждается в грубой настройке. На этом этапе необходимо найти ключевые причины существенных задержек прохождения пакетов в сети.

Обычно причина серьезного замедления или неустойчивой работы сети кроется в одном неверно работающем элементе или некорректно установленном параметре, но из-за большого количества возможных виновников поиск может потребовать длительного наблюдения за работой сети и громоздкого перебора вариантов.

Грубая настройка во многом похожа на приведение сети в работоспособное состояние.

Здесь также обычно задается некоторое пороговое значение показателя эффективности и требуется найти такой вариант сети, у которого это значение было бы не хуже порогового.

Например, нужно настроить сеть так, чтобы время реакции сервера на запрос пользователя не превышало 5 секунд.

3. Тонкая настройка параметров сети (собственно оптимизация).

- Если сеть работает удовлетворительно, то дальнейшее повышение ее производительности или надежности вряд ли можно достичь изменением только какого-либо одного параметра, как это было в случае полностью неработоспособной сети или же в случае ее грубой настройки.

В случае нормально работающей сети дальнейшее повышение ее качества обычно требует нахождения некоторого удачного сочетания значений большого количества параметров, поэтому этот процесс и получил название "тонкой настройки".

Даже при тонкой настройке сети оптимальное сочетание ее параметров (в строгом математическом понимании термина "оптимальность") получить невозможно, да и не нужно. Нет смысла затрачивать колоссальные усилия по нахождению строгого оптимума, отличающегося от близких к нему режимов работы на величины такого же порядка, что и точность измерений трафика в сети.

Достаточно найти любое из близких к оптимальному решений, чтобы считать задачу оптимизации сети решенной. Такие близкие к оптимальному решения обычно называют рациональными вариантами, и именно их поиск интересует на практике администратора сети или сетевого интегратора.

- Поиск неисправностей в сети - это сочетание анализа (измерения, диагностика и локализация ошибок) и синтеза (принятие решения о том, какие изменения надо внести в работу сети, чтобы исправить ее работу).

- **Анализ** - определение значения критерия эффективности (или, что одно и то же, критерия оптимизации) системы для данного сочетания параметров сети. Иногда из этого этапа выделяют подэтап **мониторинга**, на котором выполняется более простая процедура - процедура сбора первичных данных о работе сети: статистики о количестве циркулирующих в сети кадров и пакетов различных протоколов, состоянии портов концентраторов, коммутаторов и маршрутизаторов и т.п.

- Далее выполняется этап собственно анализа, под которым в этом случае понимается более сложный и интеллектуальный процесс осмысления собранной на этапе мониторинга информации, сопоставления ее с данными, полученными ранее, и выработки предположений о возможных причинах замедленной или ненадежной работы сети.

- **Задача мониторинга** решается программными и аппаратными измерителями, тесторами, сетевыми анализаторами и встроенными средствами мониторинга систем управления сетями и системами.
- **Задача анализа** требует более активного участия человека, а также использования таких сложных средств как экспертные системы, аккумулирующие практический опыт многих сетевых специалистов.

- ***Синтез*** - выбор значений варьируемых параметров, при которых показатель эффективности имеет наилучшее значение. Если задано пороговое значение показателя эффективности, то результатом синтеза должен быть один из вариантов сети, превосходящий заданный порог.

Приведение сети в работоспособное состояние - это также синтез, при котором находится любой вариант сети, для которого значение показателя эффективности отличается от состояния "не работает".

- **Синтез рационального варианта сети** - процедура чаще всего неформальная, так как она связана с выбором слишком большого и очень разнородного множества параметров сети - типов применяемого коммуникационного оборудования, моделей этого оборудования, числа серверов, типов компьютеров, используемых в качестве серверов, типов операционных систем, параметров этих операционных систем, стеков коммуникационных протоколов, их параметров и т.д. и т.п.

- Очень часто мотивы, влияющие на выбор "в целом", то есть выбор типа или модели оборудования, стека протоколов или операционной системы, не носят технического характера, а принимаются из других соображений - коммерческих, "политических" и т.п. **Поэтому формализовать постановку задачи оптимизации в таких случаях просто невозможно.**

Критерии эффективности работы сети

- Все множество наиболее часто используемых критериев эффективности работы сети может быть разделено на две группы.

Одна группа характеризует
производительность работы сети,
вторая - надежность.

- **Производительность сети** измеряется с помощью показателей двух типов - **временных**, оценивающих задержку, вносимую сетью при выполнении обмена данными, и **показателей пропускной способности**, отражающих количество информации, переданной сетью в единицу времени.
- Эти два типа показателей являются взаимно обратными, и, зная один из них, можно вычислить другой.

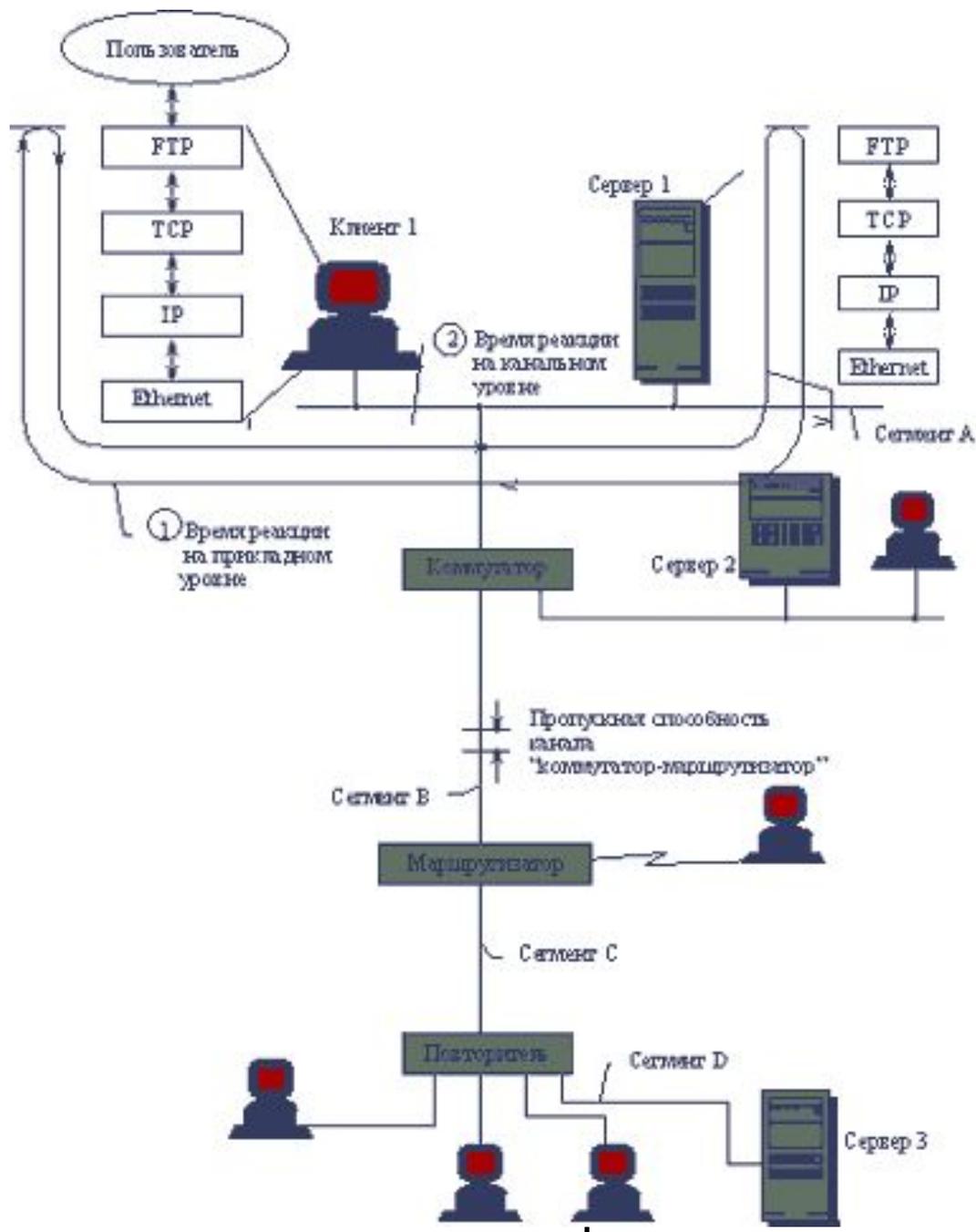
Время реакции

- Обычно в качестве временной характеристики производительности сети используется такой показатель как время реакции.
- Термин "время реакции" может использоваться в очень широком смысле, поэтому в каждом конкретном случае необходимо уточнить, что понимается под этим термином.

- В общем случае, время реакции определяется как интервал времени между возникновением запроса пользователя к какому-либо сетевому сервису и получением ответа на этот запрос (рис. 1).



Рисунок 1



- Рассмотрим несколько примеров определения показателя "время реакции", иллюстрируемых рисунком 2.

Рисунок 2

В примере 1 под временем реакции понимается время, которое проходит с момента обращения пользователя к сервису FTP для передачи файла с сервера 1 на клиентский компьютер 1 до момента завершения этой передачи.

Очевидно, что это время имеет несколько составляющих. Наиболее существенный вклад вносят такие составляющие времени реакции как:

- *время обработки запросов на передачу файла на сервере,*
- *время обработки получаемых в пакетах IP частей файла на клиентском компьютере,*
- *время передачи пакетов между сервером и клиентским компьютером по протоколу Ethernet в пределах одного коаксиального сегмента.*

Можно было бы выделить еще более мелкие этапы выполнения запроса, например, время обработки запроса каждым из протоколов стека ТСР/ІР на сервере и клиенте.

- Для конечного пользователя таким образом определенное время реакции является понятным и наиболее естественным показателем производительности сети (размер файла, который вносит некоторую неопределенность в этот показатель, можно зафиксировать, оценивая время реакции при передаче, например, одного мегабайта данных).

- Однако, сетевого специалиста интересует в первую очередь производительность собственно сети, поэтому для более точной ее оценки целесообразно вычленить из времени реакции составляющие, соответствующие этапам несетевой обработки данных - поиску нужной информации на диске, записи ее на диск и т.п. Полученное в результате таких сокращений время можно считать другим определением времени реакции сети на прикладном уровне.

Вариантами этого критерия могут служить времена реакции, измеренные при различных, но фиксированных состояниях сети:

А) Полностью ненагруженная сеть. Время реакции измеряется в условиях, когда к серверу 1 обращается только клиент 1, то есть на сегменте сети, объединяющем сервер 1 с клиентом 1, нет никакой другой активности - на нем присутствуют только кадры сессии FTP, производительность которой измеряется. В других сегментах сети трафик может циркулировать, главное - чтобы его кадры не попадали в сегмент, в котором проводятся измерения.

- Так как ненагруженный сегмент в реальной сети - явление экзотическое, то данный вариант показателя производительности имеет ограниченную применимость - его хорошие значения говорят только о том, что программное обеспечение и аппаратура данных двух узлов и сегмента обладают необходимой производительностью для работы в облегченных условиях. Для работы в реальных условиях, когда будет иметь место борьба за разделяемые ресурсы сегмента с другими узлами сети, производительность тестируемых элементов сети может оказаться недостаточной.

В) Нагруженная сеть. Это более интересный случай проверки производительности сервиса FTP для конкретного сервера и клиента.

- Однако при измерении критерия производительности в условиях, когда в сети работают и другие узлы и сервисы, возникают свои сложности - в сети может существовать слишком большое количество вариантов нагрузки, поэтому главное при определении критериев такого сорта - проведение измерений при некоторых типовых условиях работы сети.

- Так как трафик в сети носит пульсирующий характер и характеристики трафика существенно изменяются в зависимости от времени дня и дня недели, то определение типовой нагрузки - процедура сложная, требующая длительных измерений на сети. Если же сеть только проектируется, то определение типовой нагрузки еще больше усложняется.

- **В примере 2** критерием производительности сети является время задержки между передачей кадра Ethernet в сеть сетевым адаптером клиентского компьютера 1 и поступлением его на сетевой адаптер сервера 3. Этот критерий также относится к критериям типа "время реакции", но соответствует сервису нижнего - канального уровня. Так как протокол Ethernet - протокол дейтаграммного типа, то есть без установления соединений, для которого понятие "ответ" не определено, то под временем реакции в данном случае понимается время прохождения кадра от узла-источника до узла-получателя.

- Задержка передачи кадра включает в данном случае время распространения кадра по исходному сегменту, время передачи кадра коммутатором из сегмента А в сегмент В, время передачи кадра маршрутизатором из сегмента В в сегмент С и время передачи кадра из сегмента С в сегмент D повторителем. Критерии, относящиеся к нижнему уровню сети, хорошо характеризуют качества транспортного сервиса сети и являются более информативными для сетевых интеграторов, так как не содержат избыточную для них информацию о работе протоколов верхних уровней.

- При оценке производительности сети не по отношению к отдельным парам узлов, а ко всем узлам в целом используются критерии двух типов: *средно-взвешенные* и *пороговые*.
- Средно-взвешенный критерий представляет собой сумму времен реакции всех или некоторых узлов при взаимодействии со всеми или некоторыми серверами сети по определенному сервису, то есть сумму вида

$$(\sum_i \sum_j T_{ij}) / (n \times m)$$

- где T_{ij} - время реакции i -го клиента при обращении к j -му серверу, n - число клиентов, m - число серверов. Если усреднение производится и по сервисам, то в приведенном выражении добавится еще одно суммирование - по количеству учитываемых сервисов.
- Оптимизация сети по данному критерию заключается в нахождении значений параметров, при которых критерий имеет минимальное значение или по крайней мере не превышает некоторое заданное число.

- Пороговый критерий отражает наихудшее время реакции по всем возможным сочетаниям клиентов, серверов и сервисов:

$$\max_{ijk} T_{ijk},$$

где i и j имеют тот же смысл, что и в предыдущем случае, а k обозначает тип сервиса. Оптимизация также может выполняться с целью минимизации критерия, или же с целью достижения им некоторой заданной величины, признаваемой разумной с практической точки зрения.

- Чаще применяются пороговые критерии оптимизации, так как они гарантируют всем пользователям некоторый удовлетворительный уровень реакции сети на их запросы.
- Средне-взвешенные критерии могут дискриминировать некоторых пользователей, для которых время реакции слишком велико при том, что при усреднении получен вполне приемлемый результат.

- Можно применять и более дифференцированные по категориям пользователей и ситуациям критерии.
- Например, можно поставить перед собой цель гарантировать любому пользователю доступ к серверу, находящемуся в его сегменте, за время, не превышающее 5 секунд, к серверам, находящимся в его сети, но в сегментах, отделенных от его сегмента коммутаторами, за время, не превышающее 10 секунд, а к серверам других сетей - за время до 1 минуты.

Пропускная способность

- Основная задача, для решения которой строится любая сеть - быстрая передача информации между компьютерами. Поэтому критерии, связанные с пропускной способностью сети или части сети, хорошо отражают качество выполнения сетью ее основной функции.