

МАТЕМАТИЧЕСКИЕ МОДЕЛИ КОММУТАТОРОВ ЛОКАЛЬНЫХ ВЫЧИСЛИТЕЛЬНЫХ СЕТЕЙ

Рассмотрены математические модели коммутаторов для моделирования корпоративных сетей на стадии их проектирования, эффективность которых проверена на коммутаторах сети МГТУ им. Н.Э. Баумана.

Ключевые слова: сеть, коммутатор, порт, кадр, пакет, математическая модель.

В настоящее время практически все локальные вычислительные сети (ЛВС) строятся по принципу коммутации. Исключения составляют лишь ЛВС небольших организаций, содержащие от трех до семи компьютеров, которые могут быть связаны либо моноканалом (стандарт 10Base2), либо с помощью концентраторов (стандарты 10Base-T и 100Base-T) [1–3]. В сетях, состоящих из десятков, сотен и тысяч компьютеров, для организации транспортных систем используются коммутаторы, при этом соблюдается технология микросегментации, в соответствии с которой каждое сетевое устройство (компьютер, рабочая станция, принтер, IP-телефон и др.) подключается к отдельному порту коммутатора [2].

Главной характеристикой любого коммутатора является время передачи кадра из входного “провода” в выходной, что при рассмотрении коммутатора как системы массового обслуживания представляет собой время прохождения заявки через систему. От этого времени зависят и производительность коммутатора, и требуемый объем буферной памяти его портов, и вероятность отбрасывания кадра в случае переполнения очередей, организованных в буферах и т.п. Время передачи кадра зависит от многих факторов, наиболее важными из которых являются: число портов коммутатора и соотношение их пропускных способностей; степень загрузки коммутатора, т.е. число активных узлов сети, трафик которых проходит через данный коммутатор; интенсивность потоков кадров от этих узлов и т.д. Сложность выявления степени влияния всех возможных факторов на время передачи кадров, прежде всего, заключается в необходимости измерений очень малых временных интервалов. В коммутаторах фирмы Kalpana значение времени коммутации 40 мкс для ненагруженного коммутатора, т.е. такого, у которого в данный момент активными являются только два узла сети (передающий компьютер и принимающий) [2]. Производительность современных коммутаторов возросла примерно на 2 порядка, т.е. значения времени передачи кадров лежат в наносекундном диапазоне (для Gigabit Ethernet один бит-тайм составляет одну

наносекунду). Сложность измерения столь малых значений усугубляется проблемой синхронизации и определения границ полей внутри кадра, т.е. измерительная аппаратура должна соответствовать протоколам как минимум канального уровня эталонной модели ISO/OSI [3]. В этих обстоятельствах вполне целесообразно использовать для измерений средства самих узлов ЛВС и коммутаторов. Эти идеи реализованы в протоколе SNMP (Simple Network Management Protocol) с помощью объектов MIB (Management Information Base), значения которых запрашиваются у агентов, наблюдающих за устройством [2]. Однако среди групп объектов наиболее современной версии MIB (RMON 2), к сожалению, отсутствуют те, которые позволили бы непосредственно измерять время передачи кадров коммутаторами из одного “провода” в другой. Поэтому в настоящем исследовании было решено определять эту характеристику методом косвенных измерений, а именно:

- предварительно измерять время прохождения кадра между двумя компьютерами, сетевые адаптеры которых соединены кроссовой связью по витой паре, т.е. коммутатор между компьютерами отсутствует;
- подключить сетевые адаптеры двух компьютеров к портам коммутатора и вновь измерять время прохождения кадра между ними;
- определить время передачи коммутатором кадра из одного “провода” в другой по разности между двумя значениями времени прохождения кадра от компьютера–источника к компьютеру–приемнику.

При реализации данной схемы возникают проблемы синхронизации функционирования компьютера–передатчика и компьютера–приемника. Эта проблема может быть решена путем измерений, проводимых только компьютером–передатчиком, т.е. со стороны компьютера–приемника осуществляется прием переданного ему кадра, перенаправление кадра (реализуемого перестановкой MAC-адресов источника и приемника в заголовке кадра и перестановкой IP-адресов в заголовке IP-пакета) и его возврат компьютеру–передатчику. Каждый кадр при этом проходит весь тракт обмена информацией дважды (аналогичная картина имеет место при выполнении команды “ping”, широко используемой при тестировании сетей [2]).

Вполне очевидно, что полученные результаты будут иметь случайный характер. Это объясняется случайностью в смещении интервалов тактовых генераторов компьютеров и коммутаторов как для собственных процессоров, так и для шин, управляющих работой оперативной памяти (буферной памяти сетевых адаптеров компьютеров и портов

коммутаторов), режимом работы операционной системы, установленной на компьютерах (например, наличием или отсутствием фильтрации пакетов на брандмауэрах ОС, приоритетностью обработки прерываний и т.п.), аппаратно-программной реализацией тракта ТСР/IP и т.д. Поэтому результаты измерений могут быть представлены только в традиционной статистической форме (математического ожидания, вторых моментов, плотности распределения и т.п.). Учитывая все изложенное, предлагается следующий план проведения экспериментов.

1. Определение времени прохождения кадров между двумя компьютерами при кроссовом соединении их сетевых адаптеров, что позволит определить все временные задержки прохождения пакетов через ТСР/IP тракты компьютера–передатчика и компьютера–приемника, а также по самой среде передачи физического сигнала. В дальнейшем будем называть этот эксперимент и его результаты опорными или базовыми.

2. Определение времени передачи кадров ненагруженными коммутаторами, к портам которых подключены лишь компьютер–передатчик и компьютер–приемник. В этом случае таблица коммутации минимальна по объему, отсутствуют очереди в буферах портов, что позволяет определять предельные характеристики коммутатора для различных длин пакетов при обмене информацией. Ясно, что опорные результаты (см. п. 1) должны быть получены для пакетов таких же длин.

3. Определение времени передачи кадров каскадом (цепочкой) ненагруженных коммутаторов для тех же длин пакетов и по той же методике, что и п. 2.

4. Определение времени передачи кадров нагруженным коммутатором путем подключения компьютера–передатчика и компьютера–приемника к свободным портам одного из коммутаторов, функционирующих в корпоративной сети МГТУ им. Н.Э. Баумана.

5. Определение времени передачи кадров каскадом нагруженных коммутаторов путем подключения компьютера–передатчика и компьютера–приемника к свободным портам различных коммутаторов корпоративной сети университета.

Приведенные далее результаты получены для компьютеров (передатчик и приемник) с характеристиками: процессор Intel Celeron, 764,2 МГц; частота шины 66 МГц; объем оперативной памяти 256 Мб; объем жесткого диска 20 Гб; тип подключения жесткого диска IDE UDMA/33; сетевая карта, поддерживающая стандарт 100Base-T на базе чипсета Intel 82801BA/CAM Pro 100/Ethernet; операционная система FreeBSD Release 6.1, patch-level 15.

Для исследования использовался модернизированный набор программных продуктов, реализующих эхо-запрос/эхо-ответ утилиты “ping”. Модернизация касалась определения времени прохождения по сети, измеряемого утилитой в миллисекундах. При скорости передачи информации по сети 100 Мбит/с в соответствии со стандартом Fast Ethernet за 1 мс проходит 10^5 bt (бит-таймов), что эквивалентно 12500 байт. Очевидно, что подобная шкала измерения при исследовании задержки на коммутаторе использована быть не может, так как время прохождения кадров максимального размера (~1500 байт) технологии Ethernet составляет всего 15 мкс. Поэтому в качестве единицы измерения было выбрано время одного такта работы процессора компьютера-передатчика. Тактовая частота работы процессора используемых для эксперимента компьютеров составляет 764,2 МГц, что соответствует длительности одного такта $T_1 \approx 1,3$ нс или T_1 составляет 13 % длительности бит-тайма; ($1 \text{ bt} \approx 7,65 T_1$). Программный код, реализующий модернизацию, представлен на рис. 1. Код встраивается в модуль утилиты “ping”.

Эксперименты проводились для кроссового соединения, а также для коммутаторов Catalyst 2950G-24-EI фирмы Cisco, Office Connect Dual Speed Switch 16 Plus фирмы 3Com и DES-1026G фирмы D-Link.

Утилита “ping” реализуется во всех современных ОС протоколом ICMP (Internet Control Message Protocol), который отслеживает передвижение пакета по сети и при его отбрасывании передает сообщение узлу-источнику. ICMP-пакет инкапсулируется в IP-пакет и включает в себя 8-байтный заголовок и тело ICMP-сообщения. Размер ICMP-пакета в байтах (включая 8-байтный заголовок) указывается среди ключей утилиты “ping”. Структура ICMP-пакета приведена на рис. 2.

```

186c186
int interval = 1000; /* interval between packets, ms */
214a215,217
u_int32_t low, high, llow, lhigh;
u_int64_t v, lv;
251,252c254,255
// u_int32_t low, high, llow, lhigh;
// u_int64_t v, lv;
797c800
if (preload == 0) {
799c802,803
} else {
866a871,872
__asm __volatile("rdtsc" : "=a" (low), "=d" (high));
v = (low | ((u_int64_t)high << 32));
872,873d877
__asm __volatile("rdtsc" : "=a" (low), "=d" (high));
v = (low | ((u_int64_t)high << 32));
1376d1378

```

Рис. 1. Модернизированный программный код



Рис. 2. Структура ICMP-пакета

Таким образом, длина ICMP-пакета L в совокупности с 20 байтами заголовка IP-пакета дают размер “полезной” информации в Ethernet-кадре (от 46 до 1500 байт), следовательно, длина кадра составляет $L_k = L + 20 + 18 = L + 38$ байт без учета длины преамбулы кадра размером 8 байт. В предположении зависимости задержки коммутатора от длины кадра опорный эксперимент для кроссового соединения компьютеров был проведен для нескольких длин ICMP-пакетов. В табл. 1 приведены результаты усредненного времени оборота (RTT – Round Trip Time) для 10 000 экспериментов для каждого значения L .

Таблица 1

№	L , байт	L_k , байт	RTT, такты процессора	RTT, бит-тайм	RTT, байт
1	64	102	219082	28676	3584
2	128	166	230165	30126	3766
3	256	294	248608	32540	4067
4	512	550	285873	37418	4577
5	1024	1062	354832	46444	5805

Статистическая обработка результатов позволила определить уровень среднеквадратического отклонения $\sigma \approx 1700$ бт.

Из данных табл. 1 видна явная зависимость значений RTT опорного эксперимента от длины кадра, что объясняется многократной буферизацией пакета в оперативной и буферной памяти сетевых адаптеров аппаратно-программных трактов ТСП/IP компьютера-источника и компьютера-приемника.

В табл. 2 приведены результаты измерения RTT для тех же длин пакетов при соединении компьютеров через коммутатор Catalyst 2950G-24-EI для 10 000 экспериментов. Задержка на коммутаторе T_S вычислялась как разница между значениями RTT для коммутатора и RTT опорного эксперимента.

Таблица 2

№	L_k , байт	RTT, такты процессора	RTT, бит-тайм	T_S , бит-тайм	T_S , байт
1	102	240960	31531	1428	179
2	166	260243	34054	1964	246
3	294	293602	38419	2940	368
4	550	355940	45557	4080	510
5	1062	486099	63609	8583	1073

Результаты, приведенные в табл. 2, получены при подсоединении компьютеров к 1-му и 2-му портам коммутатора. В целях выявления влияния конструктивных особенностей коммутатора были проведены аналогичные измерения для подключения компьютеров к 1-му и 24-му (последнему) портам. Изменение значения T_S оказалось менее 5 % для каждой из длин пакетов.

Результаты соответствующих экспериментальных исследований для коммутатора D-Link представлены в табл. 3.

Таблица 3

№	L_K , байт	RTT, такты процессора	RTT, бит-тайм	T_S , бит-тайм	T_S , байт
1	102	247376	32371	1847	231
2	166	266091	34820	2348	294
3	294	300137	39275	3367	421
4	550	359351	47090	4836	604
5	1062	492356	64428	8992	1124

Аналогичные результаты получены и при коммутировании через 1-й и 24-й порты.

В табл. 4 приведены результаты измерений RTT для коммутатора фирмы 3Com (1-й и 2-й порты).

Таблица 4

№	L_K , байт	RTT, такты процессора	RTT, бит-тайм	T_S , бит-тайм	T_S , байт
1	102	234363	30668	996	124
2	166	253504	33172	1523	190
3	294	288589	37764	2612	327
4	550	354516	46390	4486	560
5	1062	486600	63674	8615	1077

В отличие от предыдущих коммутаторов для коммутатора фирмы 3Com использование 1-го и 16-го (последнего) порта для подключения компьютеров существенно меняет результаты экспериментов, что видно из сравнения табл. 4 и 5.

Таблица 5

№	L_K , байт	RTT, такты процессора	RTT, бит-тайм	T_S , бит-тайм	T_S , байт
1	102	244918	32049	1686	211
2	166	263580	34491	2182	273
3	294	300298	39296	3378	422
4	550	360560	47181	4882	610
5	1062	496298	64943	9250	1156

Анализ зависимости задержки T_S на свободных коммутаторах от длины кадра позволяет сделать вывод о том, что во всех трех типах коммутаторов используется коммутация с буферизацией, при этом явно прослеживается почти линейная зависимость T_S от L_K , поэтому в дальнейшем будем считать, что

$$T_S = AL_K + B, \quad (1)$$

где A и B могут быть найдены по методу наименьших квадратов в результате решения для данных каждой таблицы следующей системы уравнений:

$$\begin{cases} A \sum_{i=1}^5 L_{K_i}^2 + B \sum_{i=1}^5 L_{K_i} = \sum_{i=1}^5 L_{K_i} T_{S_i}; \\ A \sum_{i=1}^5 L_{K_i} + B \cdot 5 = \sum_{i=1}^5 T_{S_i}. \end{cases} \quad (2)$$

В результате обработки получены следующие значения A и B для исследуемых коммутаторов:

$A = 0,922$; $B = 87$ — для коммутатора Cisco Catalyst 2950G;

$A = 0,918$; $B = 135$ — для коммутатора D-Link DES-1026G;

$A = 0,987$; $B = 27$ — для коммутатора 3Com Office Connect Dual Speed Switch 16 Plus (1-й и 2-й порты);

$A = 0,973$; $B = 118$ — для коммутатора 3Com Office Connect Dual Speed Switch 16 Plus (1-й и 16-й порты).

Значения B приведены в байтах, соответственно, при расчетах значения задержки на коммутаторах длину кадра в формулу (1) следует подставлять также в байтах. Для пересчета значения T_S в бит-таймы следует умножить полученные значения на 8, а для дальнейшего перевода в секундный диапазон — на время одного бит-тайма выбранной скорости обмена информации.

Таким образом, двучленные математические модели (T_S в байтах) исследуемых коммутаторов для ненагруженных режимов получаются в следующем виде:

$$T_S = 0,922L_k + 87 \quad (3)$$

— для коммутатора Cisco Catalyst 2950G;

$$T_S = 0,918L_k + 135 \quad (4)$$

— для коммутатора D-Link DES-1026G;

$$T_S = 0,987L_k + 27 \quad (5)$$

— для коммутатора 3Com Office Connect Dual Speed Switch 16 Plus (1-й и 2-й порты);

$$T_S = 0,973L_k + 118 \quad (6)$$

— для коммутатора 3Com Office Connect Dual Speed Switch 16 Plus (1-й и 16-й порты).

Отличие значений A от единицы во всех математических моделях объясняется частичной конвейеризацией обработки кадров, при которой передача в выходной порт начинается до завершения полного

приема кадра в буфер входного порта. Заметим также, что явная зависимость задержки кадров в коммутаторе фирмы 3Com от номеров коммутируемых портов свидетельствует о двухступенчатой иерархии конструкции этого коммутатора, в соответствии с которой все порты делятся на группы. Внутри каждой группы коммутация осуществляется быстрее, чем при коммутации портов, входящих в разные группы.

Проверка достоверности полученных математических моделей ненагруженных коммутаторов проводилась экспериментальными исследованиями каскадов коммутаторов.

Последовательное соединение коммутаторов Catalyst и 3Com (1-й и 2-й порты) для длин кадров 550 и 1062 байт дает расчетное значение T_S , равное 1163 и 2140 байт соответственно. Полученные значения для 10000 экспериментов составляют 1132 и 2160 байт, т.е. погрешность не превышает 3%. При последовательном соединении коммутаторов Catalyst и 3Com (1-й и 16-й порты) и для тех же длин кадров расчетные значения равны 1247 и 2217 байт, а экспериментальные значения – 1191 и 2253 байт, т.е. погрешность менее 5%.

Для каскадного соединения трех коммутаторов эксперименты из 10 000 замеров (в коммутаторе 3Com подсоединялись 1-й и 2-й порты) позволили установить значения задержки в 1755 и 3348 байт при длинах кадров 550 и 1062 байт. Расчетные значения для тех же длин кадров составляет 1803 и 3250 байт соответственно. При коммутации 1-го и 16-го портов коммутатора 3Com экспериментальные задержки оказались равными 1795 и 3407 байт, а расчетные значения – 1887 и 3327 байт при тех же длинах кадров, т.е. во всех случаях каскадирования различия между экспериментальными значениями и значениями, рассчитываемыми по формулам (3)–(6), не превосходили 5%.

Размеры таблиц коммутации с указанием MAC-адресов и номеров портов, к которым подключены соответствующие хосты сети, влияют на время коммутации незначительно. Если в таблицах памяти входного порта отсутствует информация о том, на какой порт следует передать пришедший от источника кадр, то происходит обращение к таблице коммутации системного блока коммутатора. При отсутствии нужной информации и в этой таблице осуществляется приоритетная “заливка”, т.е. кадр из входного порта передается вне очереди на все остальные порты коммутатора. Подобная ситуация возможна лишь при включении коммутатора и его инициализации либо по истечении времени “жизни” записей в оперативной памяти входного порта и системного блока. При наличии записей “заливка” отсутствует и кадр передается на соответствующий порт. Поэтому увеличение времени задержки передачи информации при поиске в таблицах коммутации

максимальных размеров не превосходит времени задержки при передаче первого (после включения коммутатора) кадра серии, которое по результатам проведенных экспериментов не превышает 10 %. При моделировании функционирования корпоративных сетей этот фактор можно не учитывать.

Полученные в результате проведенных стендовых испытаний математические модели коммутаторов были проверены на коммутаторах реальной сети МГТУ им. Н.Э. Баумана. Подключение компьютеров к свободным портам коммутаторов в различных сегментах распределенных сетей показало полную идентичность времени задержки одного коммутатора (и их каскадов) при отсутствии перегрузок в сети значенням, рассчитанным по соотношениям (3)–(6). Наибольшее отклонение от расчетных значений не превысило 6,8 % (для серий из 10 000 посылок ICMP-пакетов утилитой “ping”).

Вместе с тем более существенные отличия имели место при тестировании этой утилитой реально работающих хостов, что позволяет сделать вывод о большем влиянии на показатели функционирования корпоративных сетей источников и приемников информации по сравнению с влиянием режимов работы коммутаторов. Вполне обоснованным можно считать утверждение о полной пригодности установленных математических моделей для моделирования корпоративных сетей на стадиях их проектирования.

СПИСОК ЛИТЕРАТУРЫ

1. Б о й ч е н к о М., И в а н о в И., К о л о б а е в Л., Л о х т у р о в В. Разработка математических моделей локальных вычислительных сетей, построенных по принципу коммутации // Отчет о НИР 01.07.05. – Москва.: МГТУ им. Н.Э. Баумана, 2006.
2. О л и ф е р В., О л и ф е р Н. Компьютерные сети. Принципы, технологии, протоколы. – СПб.: Питер, 2006.
3. Т а н е н б а у м Э. Компьютерные сети. – СПб.: Питер, 2003.

Статья поступила в редакцию 30.03.2009

Игорь Потапович Иванов родился в 1955 г., окончил МВТУ им. Н.Э. Баумана в 1979 г. Канд. техн. наук, проректор по информатизации МГТУ им. Н.Э. Баумана, заведующий кафедрой “Теоретическая информатика и компьютерные технологии” МГТУ им. Н.Э. Баумана. Автор более 30 научных работ в области информационно-коммуникационных технологий.

I.P. Ivanov (b. 1955) graduated from the Bauman Moscow Higher Technical School in 1979. Ph. D. (Eng.), vice-rector on informatization of the Bauman Moscow State Technical University, head of “Theoretical Informatics and Computer Technologies” of the Bauman Moscow State Technical University. Author of more than 30 publications in the field of information and telecommunication technologies.