МОДЕЛИРОВАНИЕ КОРПОРАТИВНОЙ СЕТИ, ИСПОЛЬЗУЮЩЕЙ VPN. РЕКОМЕНДАЦИИ ПО ПАРАМЕТРИЧЕСКОЙ И СТРУКТУРНОЙ ОПТИМИЗАЦИИ

А. Е. Журавлев, Н. В. Крупенина (Санкт-Петербург)

В настоящее время увеличивается число организаций, где всё шире используются малые и средние локальные вычислительные сети, а также распределенные структуры корпоративных сетей, связывающие в единую систему территориально распределенные подразделения (подсети). Часто корпоративные сети открыты для выхода в глобальную сеть (Internet). Управление такой сетью становится важной, актуальной и очень непростой задачей. Для мониторинга и управления работой подобных сетей используются различные по функционалу и, соответственно, по стоимости программные и аппаратные средства сетевого администрирования. При проектировании различных работ по модификации сети возникают более широкие задачи управления оборудованием и информационными процессами, среди которых можно выделить основные:

- ✓ оценка возможности выполнения новых задач при сохранении приемлемых рабочих характеристик сети;
- ✓ выявление проблем (в т.ч. потенциальных) пропускной способности каналов и устройств;
- ✓ рациональное размещение общесетевых ресурсов (например, серверов баз данных);
- ✓ оптимальная модернизация сети с точным обоснованием приобретения дорогостоящих программных комплексов и сетевого оборудования.

Особенно важно это для вычислительных центров и научных структур, планирующих новые вычислительные проекты и желающих повысить эффективность своей деятельности, а также для сетей образовательных учреждений и библиотек (в том числе внутри образовательных учреждений), имеющих жёсткие финансовые границы с одной стороны и стремление к непрерывному информационному развитию с другой.

На практике аналитические оценки даже при небольшом усложнении структуры сети становятся очень громоздкими и приближёнными. Точно проверить различные варианты сетевых решений позволяет только система имитационного моделирования.

Одним из решений некоторых из перечисленных проблем может стать система моделирования, которая проводит имитацию работы сети с заданными видами работ и их интенсивностями для каждого узла за выбранный интервал времени и выдаёт промежуточные и окончательные оценки рабочих характеристик сетевого решения, которые отражаются для каждого узла и сети в целом.

Итак, объектом исследований в рамках работы является ЛВС Ethernet, оборудование, входящее в ее состав, процессы взаимодействия компонентов локальной сети, а предметом – математическое и программное обеспечение имитационного моделирования и синтеза оптимальной структуры сети Ethernet.

Целью работы является выбор математического и программного обеспечения имитационного моделирования и формирование модели исследуемой сети с последующим выполнением параметрической и структурной оптимизации, а также выработка практических рекомендаций по возможностям долгосрочной модернизации программно-аппаратного комплекса сети.

Используя экспериментальную корпоративную сеть, приложения для сбора статистических данных и приложение для анализа данных на основе нейронных сетей различных типов, при помощи МГУА (метод группового учета аргументов) построим функцию отклика и полиномиальную модель пропускной способности сети.

Метод группового учета аргументов (МГУА) — семейство индуктивных алгоритмов для математического моделирования мультипараметрических данных. За параметрический данные возьмем тип операционной системы сервера VPN (OS), количество узлов сети VPN (C), длину используемого ключа шифрования (KEY, b), максимальный размер блока передаваемых данных (MTU, b). Результатом же определим собранные статистические данные по пропускной способности сети (S, mbps) и нагрузке на процессор сервера VPN (L, %). Сокращенный вариант выборки, представляющий собой совокупность кортежей (S, L, OS, KEY, C, MTU) приведен в таблице.

S	L	OS	KEY	C	MTU
44,38	18	1	128	10	1500
51,89	23	1	128	20	1500
57,69	38	1	128	30	1500
43,53	19	1	512	10	1500
21,13	25	1	512	20	1500
52,61	39	1	512	30	1500
12,81	14	2	128	10	1500
23,08	19	2	128	20	1500
68,06	40	2	128	30	1500
12,61	14	2	512	10	1500
23,11	22	2	512	20	1500
63,78	43	2	512	30	1500

В таблице S – в kbps (кбит/с) – общая нагрузка на канал сети (сумма общего входящего и исходящего трафика сервера VPN), рассчитанная как среднее из N = 30 испытаний, проведенных при фиксации набора P = 12 прочих независимых параметров сети:

$$S_i = \frac{1}{N}\sum S_{ij}, i = 1..P, j = 1..N;$$

L – в % – общая нагрузка на центральный процессор сервера VPN,

$$L_i = \frac{1}{N}\sum L_{ij}, i = 1..P, j = 1..N;$$

OS – условное обозначение типа операционной системы сервера VPN (*1*- MS Windows Server, *2* – Linux);

KEY – в bit (бит) – длина ключа, установленного для используемого протокола шифрования. В работе использована группа протоколов AES-xxx-CBC (где xxx – длина ключа) с длиной ключа 128, 256 и 512 бит. Обозначения: 1 – AES-128-CBC, 2 – AES-512-CBC;

C – количество узлов, одновременно осуществляющих обмен данными с сервером VPN;

MTU – в байтах – максимальный размера блока, который может быть передан на канальном уровне коммуникационного протокола. Опытным путем был выявлен оптимальный размер блока, его изменения приводят к общему ухудшению показателей работы сети вне зависимости от иных параметров.

Сбор статистических данных проводился приложениями Perfomance Monitor v5.0 и кроссплатформенным пакетом iPerf v2.0 (среда Sun JVM 6u10). Данные собраны на основе запросов к сформированной базе данных под управлением СУБД MS SQL

2005 SP2, объем выборки составил 90-100 Мб. Анализ полученных данных проведен пакетами NeuroShell 2 и Statistica 6.0. Для построения графиков использован OOp Calc 3.0.

Всего было проведено A=N*P=360 опытов, и на их основе построен сводный график (рис. 1), наглядно представляющий общую зависимость пропускной способности и загрузки CPU сервера от параметров сети.

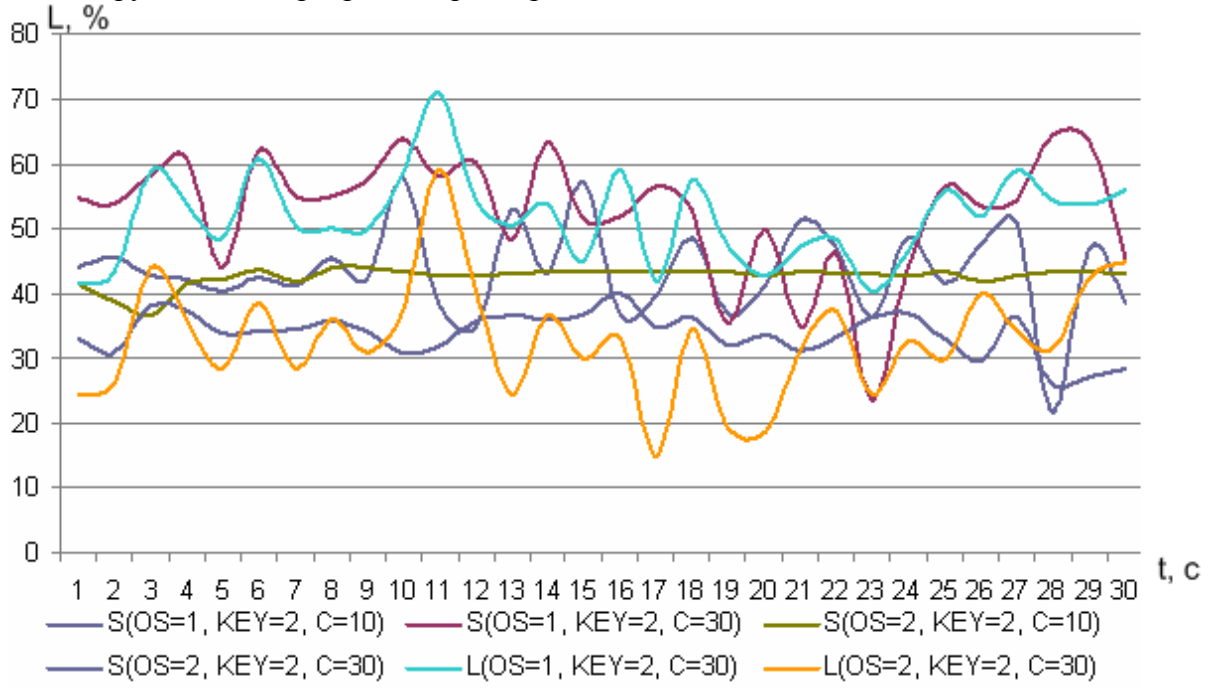

Рис. 1

Из собранных данных и представленного графика можно сделать несколько предварительных общих выводов:

1. При прочих равных условиях операционные системы семейства Linux показывают большую производительность (меньшую нагрузку на центральный процессор), особенно при увеличении количества узлов сети.
2. Для ОС семейства MS WinServer характерны достаточно большие нагрузки на аппаратные ресурсы сервера, но зависимость нагрузки от количества узлов сети достаточно мала.

Далее проведем математический анализ данных, используя МГУА и NeuroShell. Исходную выборку разбиваем на две подвыборки As и Bs. Подвыборка As используется для определения коэффициентов модели, а подвыборка Bs – для определения качества (среднеквадратического отклонения). При этом соотношение количества данных в обеих выборках составляет 80%/20%.

Далее выбираем общий вид перебираемых моделей, так называемые опорные функции. Например, полином Колмогорова-Габора:

$$Y(x_1,\ldots,x_n) = a_0 + \sum_{i=1}^{n} a_i x_i + \sum_{i=1}^{n}\sum_{j=i}^{n} a_{ij} x_i x_j + \sum_{i=1}^{n}\sum_{j=i}^{n}\sum_{k=j}^{n} a_{ijk} x_i x_j x_k + \cdots$$

Сложность модели в таком случае определяется количеством коэффициентов.

Используя опорные функции, строим вариант модели для всех аргументов – полином со всеми переменными. Для каждой модели определяются её коэффициенты методом регрессионного анализа.

$Y=-0.42-0.14*X_1-8.8E-002*X_2+0.4*X_3+0.36*X_3^2+6.7E-002*X_1*X_2+0.26*X_1*X_3-2.6E-002*X_2*X_3$, (*)

где X_1 = OS, X_2 = KEY, X_3 = C, Y = S.

На модели можно опробовать влияние всплесков широковещательных запросов. В процессе моделирования выясняются следующие параметры:

- ✓ предельные пропускные способности различных фрагментов сети и зависимости потерь пакетов от загрузки отдельных станций и внешних каналов;
- ✓ время отклика основных серверов в самых разных режимах, в том числе таких, которые в реальной сети крайне нежелательны;
- ✓ тип внутреннего протокола маршрутизации и его параметров (например, метрики), протокола шифрования и его параметров (при использовании VPN);
- ✓ предельно допустимое число пользователей того или иного сервера;
- ✓ влияние мультимедийного трафика на работу локальной сети.

Перечисленные задачи предъявляют различные требования к программам. В одних случаях достаточно провести моделирование на физическом (MAC) уровне, в других нужен уже уровень транспортных протоколов (например, UDP и TCP), а для наиболее сложных задач нужно воспроизвести поведение прикладных программ. Все это должно учитываться при выборе или разработке моделирующей программы. Ведь нужно учесть, что система должна в той или иной мере воспроизвести действия всех машин в моделируемой сети. Результаты моделирования должны иметь точность 10-20%, так как этого достаточно для большинства целей и не требует слишком много машинного времени.

Степень соответствия полученной модели можно наблюдать на графике (рис. 2)

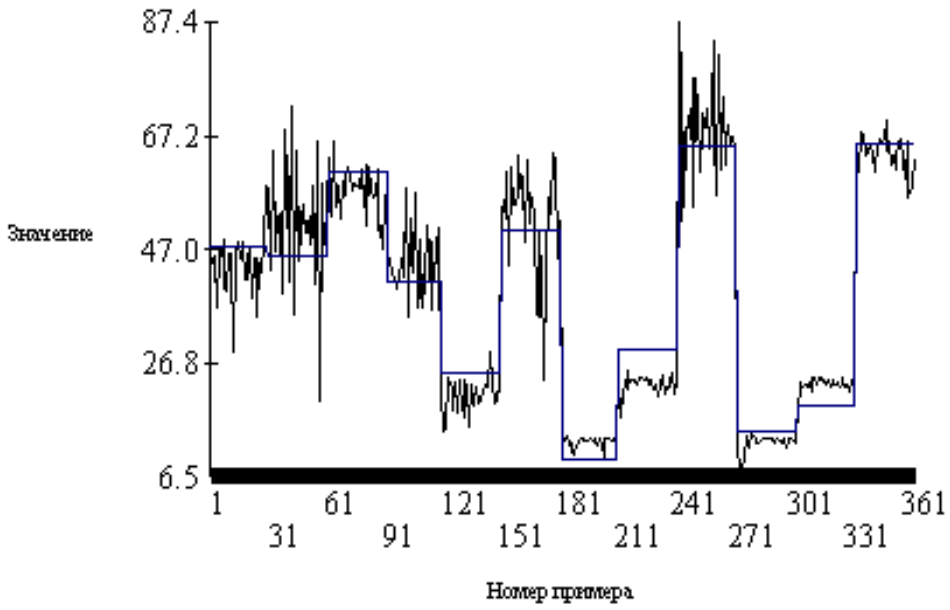

Рис. 2

Среди всех моделей выбираем наилучшую, используя приложение NeuroShell (более подробно процесс рассматривается в руководстве пользователя к программе NeuroShell). Качество моделей определяется среднеквадратическим отклонением ошибки и корреляцией Y и исходных данных (рис. 3).

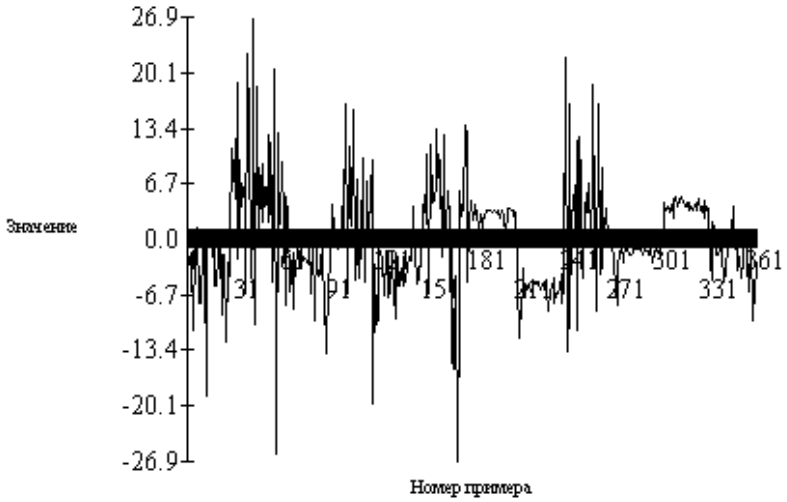

Рис. 3

Таким образом, на основе полученной модели можно предсказать оптимальные параметры корпоративной сети при заданной максимальной пропускной способности, а также определить информационную нагрузку S на сетевой канал при предопределенных параметрах.

Литература

1. **Бутов А. С., Гаскаров Д. В., Егоров А. Н., Крупенина Н. В.** Транспортные системы: моделирование и управление / Под общ. ред. проф. Бутова А. С. СПб.: Судостроение, 2001. 552 с.
2. **Семенов Ю. А.** Сетевое моделирование (ГНЦ ИТЭФ) / http://book.itep.ru/;
3. **Малышев Ю. В.** Применение моделирования в решении проблемы развития сети единого информационного образовательного пространства – Архив материалов конференции ИТО-99, http://www.ito.su/;
4. **Скуднев Д. М.** Математическое и программное обеспечение имитационного моделирования и синтеза оптимальной структуры сети Ethernet / Автореферат диссертации на соискание ученой степени кандидата технических наук, Рязань. 2009.