

**Подходы к построению имитационных моделей сети хранения данных
77-48211/638235****# 08, август 2013****Галкин В. А., Миляев Н. А.**Россия, МГТУ им. Н.Э. Баумана
galkin@bmstu.ru

Для имитационного моделирования компьютерных сетей, в частности, сетей хранения данных (СХД) может использоваться специализированное программное обеспечение (ПО) или ПО для моделирования произвольных систем (написанное на специализированном языке моделирования или на языке программирования общего назначения).

Группа китайских авторов из Сингапурского университета [1] представила программу SanSim для моделирования СХД на основе протокола FC (<http://simsan.storwav.com/>). Однако эта программа требует задания очень большого числа параметров, включая адреса взаимодействующих устройств, и поэтому проведение моделирования с её помощью представляется затруднительным. По результатам моделирования [1] удавалось получить очень маленькую погрешность по сравнению с характеристиками реальной сети (менее 10%, для некоторых характеристик даже менее 3%). Но следует отметить, что на этапе проектирования в большинстве случаев не нужно создавать точные и сложные модели, так как неопределённость значений параметров модели вносит большие погрешности по сравнению с погрешностями самой модели.

Существует большое количество программ для моделирования компьютерных сетей: NetSim, NetCracker, OPNET, ns3 и другие. Некоторые программы представлены в виде приложений с графическим интерфейсом, в рабочем пространстве которых пользователь формирует модель сети и задаёт параметры (например, OPNET, NetCracker). К другому классу приложений для имитационного моделирования относятся библиотеки программ. Для их использования необходимо написать программу на языке программирования общего назначения, используя при этом некоторые объекты из библиотеки. К таким программам относится, в частности, ns3 (в виде библиотеки классов на языке C++).

Большинство программ являются платными, особенно те, которые относятся к классу приложений с графическим интерфейсом. NetCracker является самым доступным средством (как по цене, так и по простоте использования), но он выполняет

моделирование с большим числом допущений, чем более дорогие пакеты. Он имеется в наличии на кафедре, поэтому можно будет использовать для моделирования этот пакет.

Немногие программы поддерживают моделирование протокола Fibre Channel (FC). Из перечисленных выше программ его поддерживает только OPNET, но он очень дорог (а бесплатная версия для обучения OPNET IT Guru Academic Edition не поддерживает FC). В этом случае придётся внести допущение и заменить FC другим протоколом (с похожими характеристиками).

Можно проводить моделирование при помощи средств общего назначения (предназначенных для моделирования произвольных типов систем, а не только компьютерных сетей). К таким средствам относятся программные пакеты для моделирования (AnyLogic, Arena и др.), в том числе интерпретаторы языков моделирования (GPSS, Simula и др.) [2,3]. Эти программы ничего не «знают» о сетевых протоколах, поэтому необходимо описывать последовательность передачи и размеры кадров самостоятельно.

NetCracker позволяет промоделировать взаимодействие по локальной сети и по СХД. Как отмечалось ранее, NetCracker не поддерживает моделирование каналов Fibre Channel, поэтому решено заменить его на близкий по скорости оптический канал АТМ ОС-48 (STM-16). Скорость передачи по этому каналу составляет 2,4 Гбит/с, однако следует учесть относительно большой размер заголовка ячеек АТМ (5 байт из 53), что приводит к снижению ёмкости канала до 2,17 Гбит/с, что близко к каналу FC 2 Гбит/с. Больших скоростей передачи по оптическому каналу NetCracker не поддерживает (в имеющейся в распоряжении довольно старой версии 4.1).

Схема модели приведена на рис. 1.

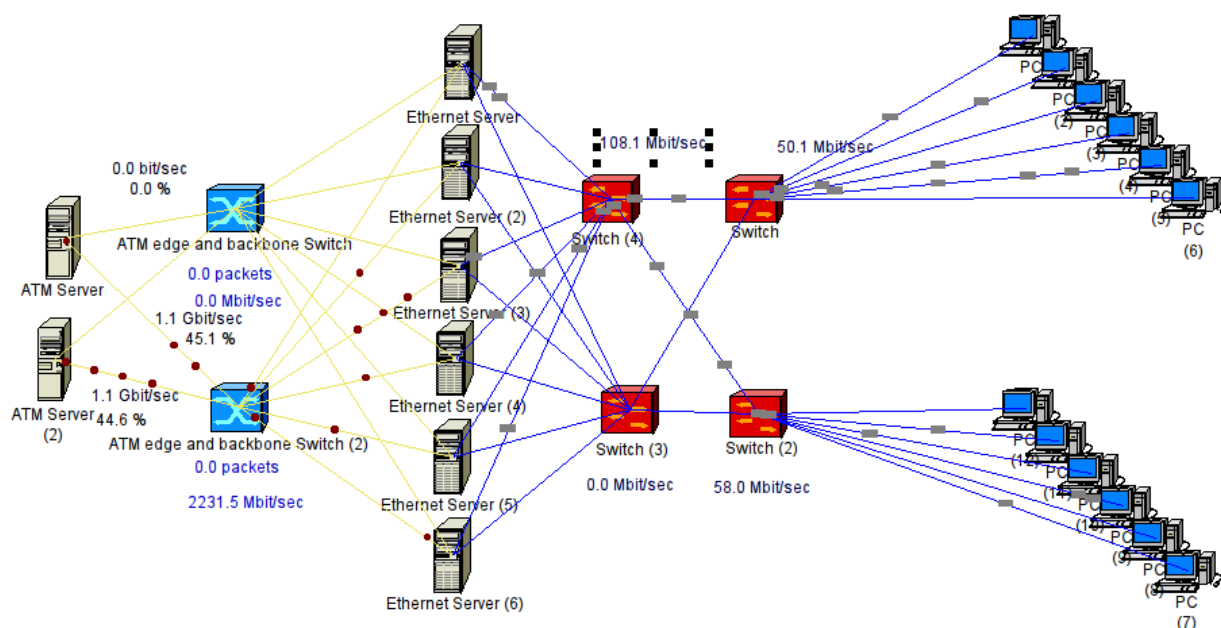


Рис. 1. Схема моделируемой сети в программе NetCracker.

Сеть состоит из нескольких серверов, подключённых через оптические коммутаторы к системе хранения данных (система хранения данных имеет два контроллера с 4 портами FibreChannel на каждом, подключённых к двум коммутаторам, поэтому она представлена в виде двух серверов). К серверам подключены несколько рабочих станций через коммутаторы Gigabit Ethernet.

От рабочих станций поступают запросы к серверам (например, на обращение к базе данных). Серверы обрабатывают эти запросы и формируют запросы к системе хранения. NetCracker не позволяет задать автоматическое порождение запроса к системе хранения при получении сервером запроса от рабочей станции, поэтому интенсивность запросов к системе хранения будет вычисляться с учётом числа рабочих станций и интенсивности запросов к каждой станции.

Каждая рабочая станция посылает запросы на каждый сервер с определённой интенсивностью и определённым размером (в рассматриваемом примере средний размер запроса 10 Кб, средняя интенсивность 20 запросов в секунду от каждой станции). Получив запрос, сервер формирует запросы на чтение или запись блоков БД. Размер блока БД — постоянный, и составляет в рассматриваемом примере 4 Кб. Этот размер может сильно варьироваться в зависимости от настроек СУБД и вариантов использования БД (от того, какие данные хранятся в базе, какой размер записи, какие установлены индексы в БД, какова статистика запросов, поддерживается ли механизм транзакций и др.). Если объекты данных большого размера, или считываются много следующих подряд объектов, то

размер блока, передаваемого за раз, будет большим, если считаются непоследовательные объекты небольшого размера, то размер блока будет небольшим (именно этот вариант предполагается в модели, считается, что размер блока передачи равен размеру блока БД). Учитывая то, что для выполнения одного запроса требуется выполнить много обращений к блокам базы данных даже с учётом кэширования на стороне СУБД, и то, что система хранения состоит из двух дисковых контроллеров на все рабочие станции, получим интенсивность запросов со стороны одного сервера на один контроллер:

$$N = B \cdot \frac{W \cdot S}{K \cdot S} = B \cdot \frac{W}{K} \quad (1)$$

В формуле (1) B — среднее количество блоков на один запрос, W , S и K — соответственно, число рабочих станций, серверов и контроллеров. В данном случае будем считать, что на один запрос сервер считывает 50 блоков, и на один контроллер приходится 6 рабочих станций. В этом случае интенсивность запросов от каждого сервера к каждому контроллеру будет в 300 раз больше, чем интенсивность запросов от рабочих станций, т.е. 6000 запросов в секунду.

В таких условиях, при одновременной работе всех рабочих станций, на канале 2 Гбит/с, при приведённых выше интенсивностях запросов, загрузка каналов СХД составляет около 45%. При этом потерь пакетов не происходит (счётчик потерянных пакетов на коммутаторе равен нулю), поток данных через коммутатор составляет 2,2 Гбит/с. Модель сети в NetCracker работает только в режиме active-standby (с холодным резервом): через один резервный коммутатор трафик не идёт, но при отказе основного коммутатора трафик переключается на резервный.

При увеличении загрузки СХД в 2 раза (до 12000 запросов в секунду с каждого сервера на каждый контроллер) сеть перегружается и начинаются сильные потери пакетов, что недопустимо в реальной системе (за первую секунду теряется 147000 пакетов, загрузка канала 78%).

При уменьшении интенсивности запросов до 10000 запросов в секунду сеть находится в режиме с небольшой перегрузкой. При этом за первую секунду теряется только 15 пакетов, загрузка канала 75%.

При уменьшении интенсивности запросов до 9000 запросов в секунду сеть работает без перегрузки, загрузка канала 68%, поток данных через коммутатор 3,38 Гбит/с. Если считать, что в нормальном режиме работы она не должна превышать 70%, то можно рассчитать максимально допустимый поток данных в сети с одним работающим коммутатором:

$$T_{\max} = 0,7 \cdot \min\{K, S\} \cdot T_k \quad (2)$$

Здесь T_{\max} — максимальный поток данных, K и S — число каналов связи со стороны, соответственно, системы хранения и серверов, T_k — пропускная способность канала связи. В частности, для промоделированного случая $T_k=2,5$ Гбит/с (пропускная способность канала STM-16, используемого в модели), $K=2$, $S=6$ и $T_{\max}=3,5$ Гбит/с. При $T_k=4$ Гбит/с $T_{\max}=5,6$ Гбит/с. при $T_k=8$ Гбит/с $T_{\max}=11,2$ Гбит/с.

Следует отметить, что T_{\max} — это максимальный допустимый в ходе эксплуатации системы поток данных. В ходе эксплуатации средняя нагрузка на сеть должна быть намного меньше T_{\max} . Хотя этот факт уже учтён при моделировании: при реальном применении редки случаи, когда все рабочие станции или серверы используют сеть одновременно. Но в случае каких-то нештатных ситуаций возможно возрастание загрузки сети (например, в случае нештатного отключения и последующего включения серверов СУБД проверяет целостность базы и откатывает незавершённые транзакции, при этом количество считываемых блоков резко возрастает).

NetCracker обладает большим числом упрощений и допущений: в частности, он не имеет возможности моделирования загрузки всех коммутаторов (только в режиме active-standby), не учитывает подтверждения при передаче кадров, и даже, по всей видимости, не учитывает заголовки кадров. Загрузка канала рассчитывается без учёта заголовков кадров; например, загрузка канала к контроллерам в том случае, когда к нему обращаются 6 серверов с 9000 запросов в секунду по 4 Кб с каждого, составляет, без учёта заголовков кадров:

$$l = \frac{6 \cdot 9000 \cdot 8 \cdot 4096}{2,5 \cdot 10^9} = 0,708.$$

Это очень близко к полученному в ходе моделирования значению. При инкапсуляции данных в ячейки АТМ (в промоделированном случае) к каждым 48 байтам данных прибавляется 5 байт заголовка, что увеличит загрузку примерно на 10% от полученной в ходе моделирования. Это даёт большие погрешности.

Так как NetCracker не даёт большой точности, его можно использовать для получения приближённых результатов, уточняя их при помощи других средств. Также NetCracker не даёт возможности моделировать устройства, работающие со своими особенностями (например, дисковые накопители требуют определённого времени на позиционирование головки чтения-записи диска, а накопители на ленте должны получать поток данных фиксированной скорости). По этой причине принимая во внимание дороговизну пакетов моделирования сетей, можно провести моделирование с использованием ПО для моделирования систем общего назначения. В качестве такого ПО

возьмём GPSS World [3].

Составим модель на языке GPSS для аналогичной рассмотренной в предыдущем разделе сети (моделироваться будет только сеть хранения данных). При моделировании будет учтено следующее:

- Дисковый массив состоит из двух контроллеров, к каждому из которых подключено по 24 диска. В сети имеются 6 серверов и 2 коммутатора FibreChannel (как и в предыдущем случае). Число серверов, контроллеров (или отдельных дисковых массивов) и коммутаторов можно менять. Скорость работы диска составляет 125 Мбайт/с, время позиционирования головки чтения-записи 10 мс.
- Передача идёт с подтверждением каждого кадра, при передаче учтён размер заголовка кадра.
- Перед выполнением операций чтения или записи данных сервер передаёт на систему хранения данных команду на чтение или запись, получает подтверждение, передаёт данные (или принимает) с подтверждением каждого пакета, и посылает команду на завершение (тоже с подтверждением). Каждый сервер обращается с одинаковой интенсивностью к каждому контроллеру и диску по каждому направлению (на чтение и на запись).
- Кроме серверов, работающих с приложениями, к системе подключён сервер, управляющий процессом резервного копирования на ленту, а также накопитель на магнитной ленте. Накопитель на ленте работает со скоростью 40 Мбайт/с (стандарт LTO-2).

Интенсивность запросов составляет 72000 запросов в секунду (по 6000 между каждым сервером и контроллером).

Во время снятия резервной копии с диска приложения не обращаются к нему, чтобы не вызвать сбоя потока данных, поступающего на ленту. Если скорость потока данных недостаточная, лента останавливается до тех пор, пока в буфере накопителя не накопится определённый объём данных, после чего привод ленты опять включается и производится запись данных. Такие остановки сильно сокращают ресурс ленты. В реальных условиях для работы с приложениями может использоваться копии дисков в RAID-массивах, в то время, пока с зеркальной копии снимается копия на ленту.

При моделировании ленточного накопителя учитывается то, что данные с диска поступают на сервер архивации, который управляет работой ленточного накопителя. Для снятия копии используется постоянно работающий диск, с которого не идёт чтение данных пользовательских приложений. Команды на открытие соединения не предусмотрено, данные поступают сплошным потоком. На каждый блок данных получатель отвечает подтверждением. Ленточный накопитель имеет буферную память, из

которой берутся данные для записи на ленту. В начале работы системы буфер считается заполненным. При опустошении буфера фиксируется ошибка, и блок данных пропускается. При нормальной работе системы должен обеспечиваться поток данных с требуемой пропускной способностью. При этом очевидно, что скорость чтения данных с диска должна быть больше скорости записи на ленту.

Очевидно, что при наличии задержек на перемещение головки чтения-записи при чтении или записи данных небольшими непоследовательными блоками производительность будет очень низкой. Поэтому будем считать, что размер блока достаточно велик и на каждое позиционирование головки диска считывается, например, 1 Мб данных. Также следует отметить, что протокол FC позволяет не передавать подтверждение на каждый кадр, а работать с сериями (последовательностями) кадров. Предположим, что блок данных уместится в одну последовательность из 512 кадров (она может состоять из многих тысяч кадров). В этом случае требуется изменить параметры модели (увеличить параметры модели, отвечающие за размер заголовка, размер запроса, размер поля данных и время коммутации пакета, в 512 раз, чтобы они соответствовали всей последовательности), а интенсивность запросов при этом уменьшить в соответствующее число раз. В модели учтено то, что коммутатор не дожидается приёма всего блока или последовательности, а начинает передавать его после получения первого кадра.

Проведём моделирование для разных скоростей работы канала FC и для разных потоков данных от серверов. При увеличении скорости канала в 2 раза число дисков также увеличивается в 2 раза, чтобы они могли нагрузить канал связи.

Результаты моделирования представлены на рис. 2–4. Скорость диска во всех случаях — 125 Мбайт/с, объём буфера ленточного накопителя и коммутатора — 1000 блоков (для накопителя) или пакетов (для коммутатора), размер блока, считываемого за одно обращение — 1 Мбайт, время позиционирования головки диска 10 мс, скорость потока данных для ленты — 40 Мбайт/с, время моделирования 60 с. В качестве интегральной характеристики производительности системы, которая зависит от параметров загрузки разных видов оборудования, лучше всего взять среднее время обработки запроса, поэтому на рис. 2–4 показана зависимость этого показателя от величины потока данных с одного сервера. Интерполяция между точками — линейная.

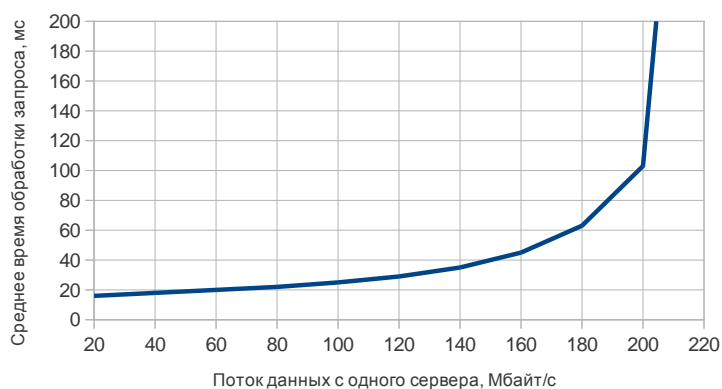


Рис. 2. Зависимость времени обработки запроса от нагрузки на сервер при скорости канала 2 Гбит/с, 24 дисках на контроллере.

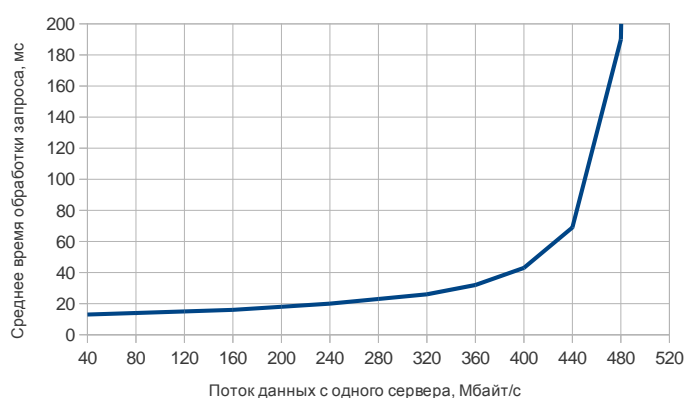


Рис. 3. Зависимость времени обработки запроса от нагрузки на сервер при скорости канала 4 Гбит/с, 48 дисках на контроллере.

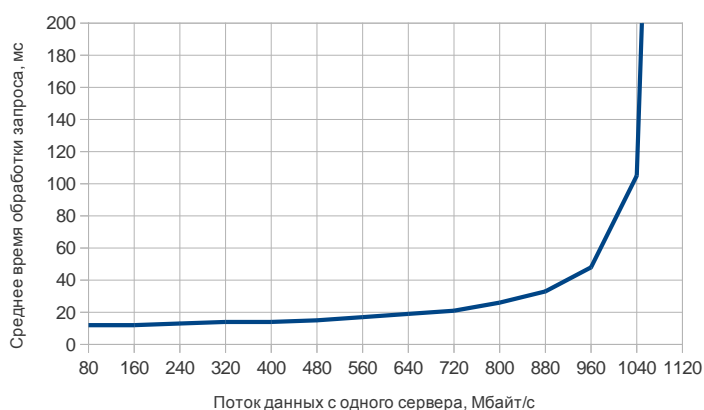


Рис. 4. Зависимость времени обработки запроса от нагрузки на сервер при скорости канала 8 Гбит/с, 96 дисках на контроллере.

По результатам имитационного моделирования можно сделать вывод, что повышение скорости канала позволяет сократить время обработки запросов и увеличить максимальный поток данных, поступающий с серверов. При этом должно быть соблюдено условие:

производительность дисков должна быть достаточной, чтобы обеспечить этот поток данных, т.е. увеличение скорости канала должно быть связано с увеличением числа дисков или их скорости, в противном случае эффект от повышения скорости канала будет незначителен.

Проведём сравнение двух подходов к имитационному моделированию системы: на основе специализированного ПО (NetCracker) и на основе ПО для моделирования систем общего назначения (см. табл. 1).

Таблица 1

Сравнение различных подходов к имитационному моделированию СХД

Критерий сравнения	Специализированное ПО (например, NetCracker)	ПО для моделирования систем общего назначения (например, GPSS World)
Трудоёмкость разработки модели	Незначительная по сравнению с ПО общего назначения.	Более значительная, так как в специализированном ПО уже разработаны модели отдельных сущностей предметной области (сетевого оборудования).
Точность и стоимость моделирования	Зависит, как правило, от затрат на приобретение программного продукта. Чем дороже продукт, чем сложнее заложенная в него модель и чем лучше описывает она реальный процесс, тем выше точность. NetCracker, как продукт начального уровня, имеет большое число допущений, и поэтому обеспечивает низкую точность по сравнению с системами более высокого класса. Затраты на разработку самой модели низкие.	Зависит в первую очередь от затрат на разработку модели. Само моделирующее ПО, как правило, стоит дешевле, чем программные продукты для моделирования специализированных типов систем. Чем сложнее модель и чем лучше она описывает реальный процесс, тем больше затраты на её разработку и больше точность.
Гибкость модели	Меньше, чем при использовании ПО для моделирования систем общего назначения, так как используются разработанные заранее модели компонентов системы.	Большая гибкость, так как модели компонентов системы разрабатываются самим разработчиком модели и можно менять их свойства произвольно (насколько это допускает моделирующая среда)
Представление модели и выходных данных	Как правило, модель представлена наглядно в виде графической схемы, а выходные данные — в виде индикаторов и надписей на схеме. Поддерживается анимация. NetCracker	Модель представлена в виде исходного кода на языке моделирования или на языке программирования общего назначения. Выходные данные — как правило, таблицы и графики. Анимация

	поддерживает в том числе и звуковое оповещение пользователя.	поддерживается с помощью отдельных пакетов, подключаемых к среде моделирования. Некоторые среды, например, AnyLogic, поддерживают анимацию сами по себе.
--	--	--

Литература

1. Simulation of Fibre Channel Storage Area Network Using SANSim / Chao-Yang Wang, Feng Zhou, Yao-Long Zhu и др., 2004): [Электронный ресурс] — Режим доступа http://uluru.ee.unsw.edu.au/~tim/icon_final/507.pdf, свободный.
2. Черненький В. М., Ревунков Г. И., Постников В. М. Эксплуатация АСОиУ : метод. указания к выполнению курсовой работы; МГТУ им. Н.Э. Баумана. - М. : Изд-во МГТУ им. Н. Э. Баумана, 2009. - 28 с.
3. Черненький В.М., Черненький М.В., Михеев В.А. Имитационное моделирование. Сборник учебно-методических работ кафедры «Системы обработки информации и управления» (бакалавры): Учебное пособие; Выпуск 1. МГТУ им. Баумана, 2013. - 279 с.