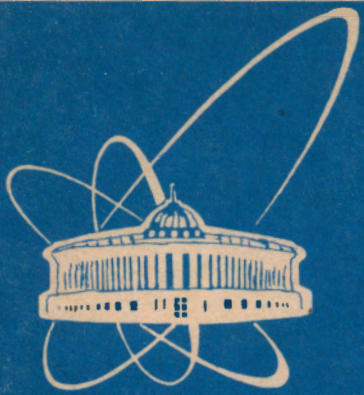

СООБЩЕНИЯ
ОБЪЕДИНЕННОГО
ИНСТИТУТА
ЯДЕРНЫХ
ИССЛЕДОВАНИЙ

Дубна

Р10-95-52

К.Ф.Окраинец

IP-МОНИТОРИНГ

1995

1 Введение

Появление у ОИЯИ среднескоростных (до 512 Кбод) коммуникационных линий, используемых в качестве носителей сетевого трафика (в основном протокольного семейства TCP/IP), требует решения ряда проблем.

В первую очередь, это организация эффективного мониторинга внешних каналов связи. При возникновении проблем оператор или ответственный должен быть поставлен в известность как можно быстрее. Кроме того, в условиях, когда один физический канал Дубна-Потсдам несет почти всю нагрузку (~75%) информационного и вычислительного взаимодействия с внешним миром, для многих пользователей критически важно знать, что случилось и каковы перспективы.

Далее, весьма важно получать данные о реальной, а не декларируемой производительности внешних линий, об их утилизации, о логической структуре сетевого потока и распределении входящих и исходящих IP-пакетов в терминах сетевой топологии и чисто географически.

Анализ подобных данных позволил бы существенно оптимизировать использование каналов, а при экстенсивном расширении сетевой инфраструктуры связей базировать технико-административные решения на точном знании.

Обычно такого рода задачи решаются с помощью специализированных программных средств (HP OpenView, IBM NetView/6000, SunNet Manager и т.д.). Они предназначены для оперативного управления гетерогенными сетями и нуждаются в особых программно-аппаратных агентах для каждого объекта управления, поддерживающих протокол SNMP и спецификации MIB. Однако сейчас подобных средств в Институте нет. Кроме того, статистические данные, полученные от таких агентов, носят чрезмерно обобщенный характер.

2 Принципы строения и реализация

2.1 Общая схема

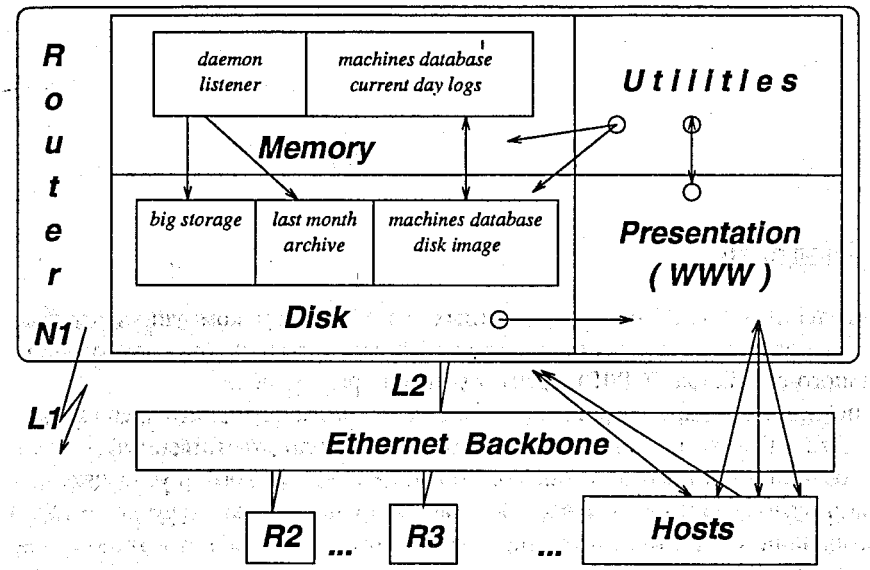

Рис. 1. Общая схема

Ключевой элемент схемы - "слушатель", фоновый процесс (в терминологии Unix [1] "daemon", далее в тексте "демон"), который опирается либо на примитивы ОС, либо на специализированные разработки модулей STREAMS [1], встраиваемые в ядро операционной системы.

Принципиальной разницы с точки зрения пассивного мониторинга нет, но следует отметить, что механизм STREAMS более гибок и позволяет легко и естественно реализовать такие потенциально необходимые для маршрутизатора возможности, как резервирование пропускной способности канала (напр. для real-time приложений), сортировку пакетов по приоритетам и защиту внутренней сети Института от несанкционированного вторжения извне (firewall).

Вернемся к системе в целом.

Демон собирает заголовки всех IP-пакетов, поступающие к нему с заданных сетевых интерфейсов, как типа точка-точка, так и CSMA/CD (L1,L2 на рис.1). Заданные интерфейсы вводятся в промискуитетный режим, что означает реакцию системы не только на адресованные ей пакеты, но и на любые другие.

[1] Unix - торговая марка Open Software Foundation

Демон ведет несколько журналов:

кратковременный - содержит подробную информацию за последние 30 секунд, а именно время отклика ближайшей западной точки, обьем и число пакетов с разбивкой по протоколам и список сессий (см. ниже). Предназначен для графического и гипертекстового мониторинга;

долговременный - фиксирует время отклика и обьем передач. Предназначен для архивирования;

"база машин" - **все** машины во всем мире, с которыми общались компьютеры ОИЯИ за большой промежуток времени, обычно месяц. Содержит обьемы обмена с разбивкой по наиболее распространенным сетевым услугам.

Помимо этого, демон принимает заявки от клиентов, желающих получить результат последнего наблюдения, и по истечении времени накопления (30 с) рассылает кратковременный журнал клиентам.

Отдельный набор утилит предоставляет средства для упорядочивания данных, обработки и перевода результатов в гипертекстовый вид, что позволяет интегрировать систему с мощным аппаратом презентации WWW [2] и богатыми графическими возможностями развитых WWW-клиентов.

2.2 Технические подробности демона

2.2.1 Сессии и способ их получения

Элементарная единица информации, с которой имеет дело демон,–это заголовок IP-пакета, инкапсулированный в заголовок уровня LLC (Link Level Control). Последний, среди прочего, содержит физические адреса отправителя и получателя, в простейшем случае уникальные 48-битные адреса адаптеров Ethernet. Обьединив интересующие нас поля обоих заголовков и отбросив ненужное, получим

физ.	адреса
адрес	интерфейса
IP	адреса
порт	отправителя
порт	получателя
IP	протокол
длина	пакета

Поток заголовков, поступающих снизу, от ядра ОС, буферизуется (~200 заголовков/буфер). Если подсчитать, каков будет поток заголовков при максимальной загрузке интерфейса, то становится ясным, что необходима интеграция данных, причем чем раньше, тем лучше. Это диктуется довольно жесткими

требованиями к оперативной памяти. Для сокращения потока данных алгоритм оперирует некоей абстракцией, "сессией", слегка похожей на обычное TCP-соединение[3].
Определим сессию как структуру данных, изображенную ниже.

адрес	отправителя
адрес	получателя
сервис	
число	пакетов
обьем	

На этом этапе отбрасываются чисто локальный трафик и пакеты, принадлежащие неотслеживаемым каналам. В случае, когда каналы не являются локальными для машины, на которой установлен монитор, сочетание логических и физических адресов получателя и отправителя позволяет однозначно идентифицировать внешний канал и, соответственно, определить, нужны ли эти данные. Такой способ удаленного мониторинга не очень хорош, и, вообще говоря, его надо избегать, но практически он вполне работоспособен. Способы получения некоторых полей (неочевидные) таковы:

сервис определен, если минимум по портам получателя и отправителя меньше 1024 (такие номера фиксированы за определенными сетевыми услугами и присваиваются единым комитетом по стандартам в Интернет[5]). Если же это неверно, то полю присваивается значение, символизирующее неизвестность.

Получая из ядра очередную порцию заголовков, демон для каждого из них пытается найти подходящую сессию. В случае успеха такая сессия модифицируется, в противном случае инициализируется новая.

Насколько же эффективен такой алгоритм ? Практика показывает, что демон, получая поток сырых данных, интегрирует их в новый абстрактный тип, сокращая обьем в среднем (максимально) в 200 (10^4) раз практически **без потери информативности.**

2.2.2 Архивация и общие функции

В начале каждого периода наблюдения демон посылает по контрольному пакету (ICMP Echo Request, см. [4]) на ближайшие машины вне Института, связанные с каналами, по которым производится мониторинг. Время отклика измеряется, но если оно превышает 30 с, то измерение потеряно. Впрочем, это совершенно неважно, поскольку с практической точки зрения работать при времени отклика более 3 с просто нельзя.

По истечении срока наблюдения демон:

- готовит и записывает на диск краткую сводку;
- формирует расширенную сводку и рассылает ее зарегистрированным клиентам.

Краткая сводка содержит:

- временной штамп;
- входной обьем;
- выходной обьем;
- время отклика.

Расширенная сводка формируется из времени отклика, настроечных параметров и списка сессий, упорядоченного по убыванию обьема передач и обрезанного так, чтобы трансмиссия помещалась в максимально возможный кадр Ethernet.

Краткая сводка сразу же пишется на диск в режиме без буферизации. Архив ведется в виде простых плоских файлов со схемой именования по дате. Раз в сутки демон меняет архивный файл, а раз в месяц производится чистка всего внутреннего состояния демона, а архив сворачивается и компрессируется.

В сжатом виде годовой отчет о работе одного канала займет не более 10 Мбайт (смешная цифра!).

Это, разумеется, не самый лучший способ хранения, но, к сожалению, в ОИЯИ отсутствуют современные реляционные базы класса Sybase,Oracle или Informix, ставшие де-факто стандартом для информационной поддержки сетей масштаба предприятия.

2.2.3 База машин (БМ)

Стартуя, демон вчитывает в себя дисковый образ существующей базы, а умирая, синхронизирует с диском свой внутренний образ.
Что же есть эта база ? Ниже следует внутренняя структура записи:

адрес
канал
обьем
ftp
DNS
mail
gopher
www
telnet

Каждая такая запись соответствует компьютеру, чей пакет, пусть даже всего один, прошел через внешний канал связи. База эта предполагалась **очень** большой - ведь в Интернет порядка 3 миллионов машин. Предполагалось также, что она будет расти, пока позволит память. Однако выяснилось, что существуют стабильные области интересов, своего рода сетевые ареалы, за пределы которых пользователь выходит очень редко. Для сохранения некоторой динамики в данных БМ архивируется и очищается раз в месяц. Общий же пул известных машин, после демонстрации бурного роста в первый месяц, далее практически не увеличивается.

БМ пополняется и модифицируется каждый интервал наблюдения. Материалом служат сессии. Так как на этапе проектирования предполагался большой размер, то поиск по ключу "IP-адрес" (естественный первичный ключ) оптимизирован внутренним представлением записей в виде хешированных двойных списков.

Итак, все сессии находят отражение в БМ с утратой временной составляющей, но с разделением по машинам и сервису.

2.3 Утилиты, HTML-документы и другое

В основном нижеописанные служебные программы предназначены для того, чтобы представить различные выборки из накопленной статистики в виде HTML-документов [7] для удобства доступа через систему WWW. Многие из них генерируют "на лету" простейшие гистограммы в формате GIF87. Термин "CGI-script", применяемый к этим программам, обозначает, что они удовлетворяют так называемому Common Gateway Interface - определению способа взаимодействия WWW с программами-шлюзами, выдающими гипертекст и способными обрабатывать параметры, заданные в формах WWW-клиентов (fill-out forms). Эти CGI-утилиты тесно связаны между собой (и центральным WWW-комплексом ОИЯИ) структурой из нескольких десятков HTML-документов, рассказывающих о сетевых связях Института и многом другом.

Рис.2 изображает весь презентационный комплекс в чрезвычайно упрощен-

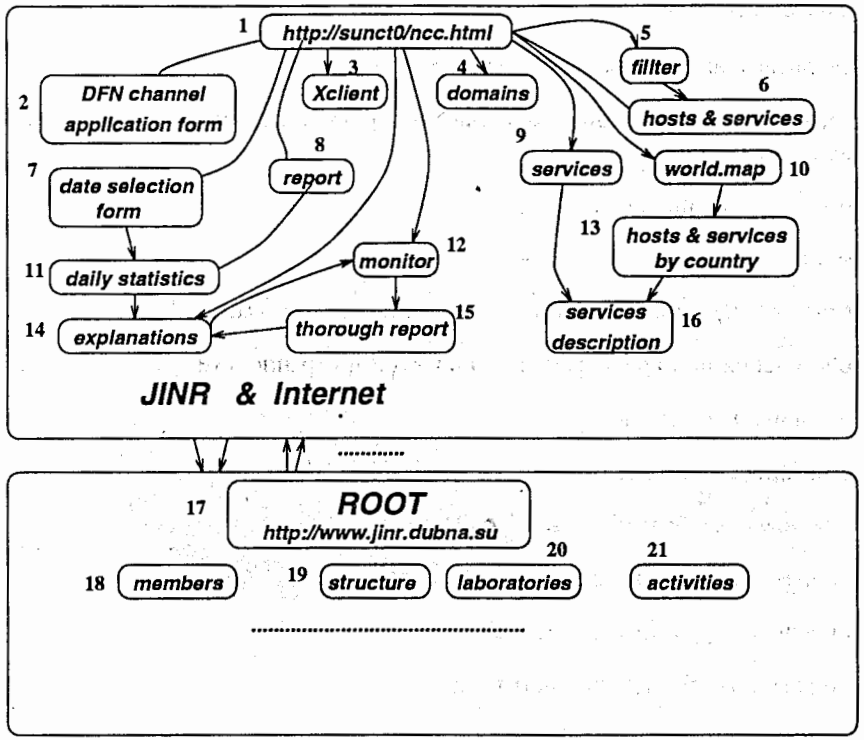

Рис. 2. Представление результатов

ном виде (большинство кросс-связей опущено):

1. центральный входной документ;
2. как воспользоваться каналом Дубна - Потсдам;
3. X-клиент;
4. сводка по адресным доменам высшего уровня;
5. форма-фильтр, в которой задается шаблон для отбора имен машин и набор сетевых услуг;
6. отчет по записям, удовлетворившим условиям фильтра;
7. форма для выбора даты, вызывает утилиту, генерирующую отчет за день со всевозможными статистическими оценками;

8. общая статистика;
9. распределение обмена по сетевым услугам;
10. карта мира, на которой можно указать страну и получить отчет по обменам с ней;
11. отчет за текущий день;
12. монитор текущего состояния;
13. отчет по странам с гистограммой по сетевым услугам;
14. обьяснения по параметрам, терминологии и странностям;
15. расширенная сводка;
16. описание сетевых услуг;
17. официальный WWW-сервер ОИЯИ;
18. список стран - участниц;
19. описание структуры Института;
20. описание деятельности Института;

.....

Многие утилиты должны использовать имена машин, а так как внутреннее представление имеет дело исключительно с адресами, то есть необходимость в программе перевода. Такой перевод не может быть осуществлен в процессе on-line генерации отчетов, поскольку база соответствия адресов имен - Domain Name Service (DNS) [6] имеет принципиально распределенный характер (например, для всего мира за область jinr.dubna.su имен отвечает машина sundg0, за область kiae.su - ns.kiae.su и т.п.) и опрос тысяч машин может неоправданно затянуться. Во избежание этого и существует еще одна программа, которая заставляет демона синхронизироваться с диском, выясняет имена машин, используя в трудных случаях эвристики и аппроксимации, и пополняет дисковый кэш. Этим кэшем и пользуются другие утилиты.

Кроме того, разработан, распространен и используется X11 Window System клиент - монитор, который графически изображает развитие во времени загрузки каналов и выводит списки сессий с интегральными оценками скоростей обмена.

2.4 Обзор данных

2.4.1 Распределение обьема передач по сервисам

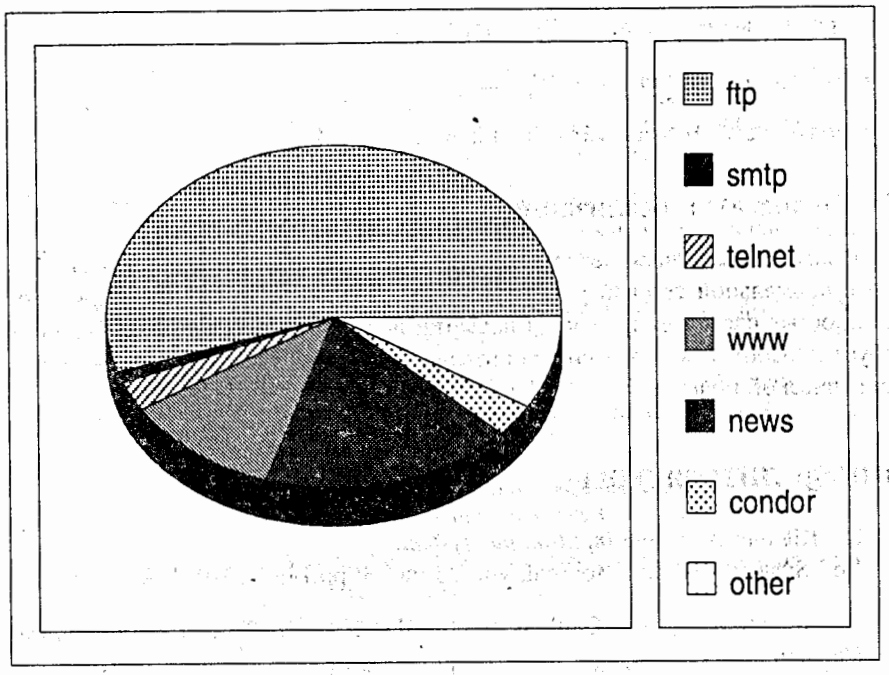

Рис. 3. Сетевые услуги

На рис. 3 изображено распределение обьема переданной информации по сетевым услугам. Видно, что большая часть - файловые передачи. Категория "other" состоит из информационного сервиса, сконфигурированного нестандартно, тестов multimedia и служебных пакетов.

Такая структура потока, хотя и не критична, но должна вызвать некоторые раздумья. Данные свидетельствуют о недостаточной проработке архивной службы и о том, что новые, чрезвычайно удобные формы информационного сервиса по каким-то причинам недоступны или не распространены.

2.4.2 Обобщенные показатели

Ниже приведены данные по каналу на DFN за январь 1995 и декабрь 1994 года:

- общее число перерывов связи 4;

- в работоспособном состоянии 99% времени;
- средняя загрузка 36%;
- среднее время отклика 540 миллисекунд;
- передано 5748 мегабайт (95 в день);
- получено 7294 мегабайт (121 в день).

2.5 Недостатки реализации

Главным структурным недостатком является невозможность хранить результаты в нормальной сетевой реляционной базе данных. Остальные недостатки суть простые следствия первого и нехватки дискового пространства. X-клиент требует доработки в части синхронизации доступа к обратному DNS-кэшу и косметической правки.

Список литературы

[1] D. Ritchie: *A stream input-output system*,
Bell System Technical Journal, vol. 63, no. 8, pp.1897-1910, 1984.

[2] T. Berners-Lee, R. Gailliau et al.: *World Wide Web: the Information Universe*,
Electronic Networking 2(1)52-58, 1992.

[3] J. Postel: *Transmission Control Protocol*,
RFC0793, 1981.

[4] J. Postel: *Internet Control Message Protocol*,
RFC0792, 1981.

[5] J. Postel, J. Reynolds: *Official Internet protocols*,
RFC1011, 1987.

[6] J. Postel: *Domain Name System Structure and Delegation*,
RFC1590, 1994.

[7] T. Berners-Lee, D. Connolly: *HyperText Markup Language (HTML)*,
URL=http://www.w3.org/hypertext/WWW/MarkUp/MarkUp.html, 1993.

Рукопись поступила в издательский отдел
9 февраля 1995 года.

Окраинец К.Ф.
P10-95-52

IP-мониторинг

Данная работа описывает архитектуру и реализацию программного комплекта, предназначенного для мониторинга внешних каналов связи, несущих IP-пакеты. Комплекс также собирает статистику использования и предоставляет средства доступа к данным и их анализа. Программы доступны как http://sundg0.jinr.dubna.su/ncc.html.

Работа выполнена в Лаборатории вычислительной техники и автоматизации ОИЯИ.

Сообщение Объединенного института ядерных исследований. Дубна, 1995

Перевод автора

Ocrainets C.F.
P10-95-52

IP Monitoring

JINR external IP traffic monitoring and usage statistics gathering software package description is given. Presentation issues along with implementation details are discussed. Per-service traffic analysis is done. Results accessible on-line as http://sundg0.jinr.dubna.su/ncc.html.

The investigation has been performed at the Laboratory of Computing Techniques and Automation, JINR.

Communication of the Joint Institute for Nuclear Research. Dubna, 1995