Изучение связности потоков данных между сетевыми абонентами в целях обеспечения безопасности корпоративной СПД СО РАН*

Ю. И. Шокин, В. С. Никульцев, В. М. Стубарев,
И. В. Шабальников, С. Д. Белов
Институт вычислительных технологий СО РАН, Новосибирск, Россия
e-mail: shokin@ict.nsc.ru, nik@ict.nsc.ru, vikst@ict.nsc.ru,
igor@nsc.ru, belov@nsc.ru

Some approaches to solution of the monitoring problems arisen during data transmission within the SB RAS network are discussed. Authors allocate subsets of parameters accessible to measurements which determine the objects of monitoring. Time space on which these parameters are determined is also established. Technologies of extraction of the necessary characteristics and the complexes of tool programs intended for solution of the problems of network monitoring are considered. Various programming and technological solutions for the estimation of data flows between network users are proposed.

Введение

Решение задач мониторинга телекоммуникационной инфраструктуры сети передачи данных СО РАН (СПД СО РАН) необходимо для поддержки эффективного управления сетью, обеспечения надежности ее функционирования, гарантированного качества обслуживания абонентов и безопасности, а также для сбора статистики и детального контроля выполнения абонентами сети установленного регламента работы. Неотъемлемой частью мониторинга телекоммуникационных сетей является определение связности потоков данных между сетевыми абонентами и контроль за динамикой их поведения.

Авторами выделяются подмножества параметров, доступных для измерений и являющихся объектами мониторинга, а также устанавливается связь с временным пространством, на котором эти параметры и соответствующие объекты определены. Рассматриваются доступные технологии извлечения необходимых характеристик и инструментальные программные комплексы, предназначенные для решения некоторых задач сетевого мониторинга. Предлагается ряд программно-технических средств, обеспечивающих оценку потоков данных между сетевыми субъектами, представленными адресными сегментами или адресами абонентов сети, или агрегатами, состоящими из списков

*Работа выполнена при финансовой поддержке Российского фонда фундаментальных исследований (грант № 06-07-89038), Программы интеграционных фундаментальных исследований СО РАН (проект № 1.7).

адресов и блоков адресов абонентов. Значения параметров, определяющие потоки между сетевыми субъектами, представляются в виде элементов двухмерных матриц.

Комплекс программ включает графические средства отображения и анализа динамики потоков, обеспечивает соответствующей информацией технологические службы сети, использующие результаты подобного мониторинга для последующего принятия управленческих решений.

1. Общая постановка задачи

Одновременно с развитием и внедрением сетевых технологий, развитием тех или иных сетевых решений и устройств, их реализующих, наблюдается непрерывное развитие систем, обеспечивающих мониторинг сети. Наряду с коммерческими продуктами существует большое количество свободно или условно бесплатно распространяемых систем, доступных через сеть Интернет.

Институт вычислительных технологий СО РАН непосредственно вовлечен в работу по созданию и эксплуатации корпоративной СПД СО РАН. При этом широкий спектр задач по управлению центральным коммуникационным ресурсом этой сети решается с использованием результатов мониторинга сети.

Содержание настоящей статьи затрагивает вопросы анализа некоторых этапов решения задач мониторинга в конкретной сетевой инфраструктуре центра управления сетью Новосибирского научного центра и представляет некоторые результаты исследований взаимодействия сетевых абонентов в СПД СО РАН.

2. Задачи сетевого мониторинга

При формализации задач мониторинга сетевой инфраструктуры возникает в первую очередь вопрос об определении списка или набора этих задач, куда входят:

— поддержка эффективного управления сетью;

— обеспечение надежности функционирования сети;

— гарантирование необходимого уровня качества обслуживания абонентов;

— обеспечение безопасности функционирования сети;

— обеспечение сбора статистики;

— решение вопросов биллинга в сети;

— решение фискальных задач, связанных с соблюдением регламента работы абонентов в сети.

Этот перечень может быть расширен и детализирован, однако и в этом виде он указывает на обязательность непосредственной привязки задач мониторинга к конкретным задачам управления сетью.

Следующим шагом в постановке задачи сетевого мониторинга является определение набора параметров, описывающих состояние исследуемой сети и обеспечивающих возможности управления сетью как объектом. Множество параметров, характеризующих работу сетевых устройств, в известной мере детерминировано и, как правило, достаточно полно представлено в соответствующих внутренних базах данных этих устройств. В такой ситуации становится необходимым определение совокупности параметров, влияющих на методы решения практических задач мониторинга.

Не вдаваясь в детали сетевой топологии сети Новосибирского научного центра СО РАН (ННЦ), ограничимся некоторыми количественными характеристиками ее масштабов и качественного состава:

— активных коммутирующих устройств — более трех десятков (включая такие устройства, как Cisco 2950, Cisco 4507 Cisco 7206NPE, различные серверы технологических баз данных, например серверы видеоконференцсвязи, VoIP-шлюзы, и т. п.);

— "пассивных" устройств — более 200 (мультиплексоры, модемы, конверторы и т. п.);

— магистральных сегментов кабельных линий связи — более 50, общей протяженностью более 150 км;

— точек соединений — более 4000;

— систем энергоснабжения, аварийного питания с разными уровнями мощности — более 30;

— системы климат-контроля.

Таким образом, число параметров, влияющих на решение тех или иных конкретных задач мониторинга, может быть оценено в 4–5 тыс. Эти параметры характеризуют только физический и канальный уровни сети. В приведенных выше цифрах отсутствуют данные, определяющие региональные структуры СПД СО РАН и канальные структуры, их интегрирующие. Эти объекты обладают относительной автономностью и, как следствие, оказывают малое влияние на работу остальных сетевых структур, в то время как сегмент, расположенный в ННЦ, в полной мере определяет работу всех абонентов СПД СО РАН.

Учет параметров сетевого уровня при решении задач мониторинга приводит, в свою очередь, к необходимости увеличения общего количества параметров, определяющих работу сети, до 5–10 тыс. Это число может быть увеличено, если исходить из потребности привлечения информации уровня приложений. Для оценки, например, количества сессий, которые существуют одновременно в исследуемом сетевом сегменте, приходится учитывать число пользователей этой сети, которое в отдельные временные интервалы достигает 20–40 тыс.

Для решения всего комплекса задач мониторинга необходимо обязательно учитывать и временные масштабы процессов в сети, а также временные масштабы, характеризующие актуальность тех или иных задач. С достаточной определенностью можно говорить о наличии в сети процессов с временными масштабами от единиц секунд до единиц и десятков минут, от десятков минут до часов и более, от часов до суток, недель и месяцев. Очевидно, что эти масштабы зависят от специфики решаемых задач мониторинга.

3. Технологии сетевого мониторинга

Определив список задач мониторинга, представив количественно параметры, на которых они определены, и проведя их временное масштабирование, следует выделить некоторые базовые технологии, лежащие в основе методов и процедур решения этих задач:

— системы, базирующиеся на использовании опроса внутренних информационных баз данных на основе протокола SNMP (приложение MRTG);

— системы, ориентированные на технологию NetFlow, позволяющую проводить исследования сетевого трафика на уровне сеансов;

— системы, обрабатывающие потоки данных, непосредственно проходящих через них, или системы, получающие зеркалированные копии этих потоков данных и сохраняющие ту или иную информацию о них.

Для того чтобы в большей мере оценить масштаб возникающих проблем в процессе мониторинга сети, необходимо привлечь еще один интегральный параметр, влияющий на использование той или иной технологии. Этот показатель связан с величиной пропускной способности внешнего канала связи ННЦ и в разные периоды времени в зависимости от финансовых условий колеблется от 100 до 150 Мбит/с. При этом, например, в системах, использующих технологию мониторинга потока NetFlow, темп записей в базу данных достигает 6–10 тыс. записей в секунду или, соответственно, миллионов записей в сутки. Непрерывное формирование запросов к базе данных на фоне подобных потоков данных требует весьма серьезных ресурсов системы и даже в случае их использования далеко не всегда отвечает характерным временам исследуемых процессов в сети. Визуальное отображение динамики исследуемых процессов в ряде случаев оказывается делом непростым, а порой и вообще невозможным, несмотря на то что перечисленные выше задачи настоятельно требуют оперативного режима мониторинга.

Перечисленные выше подходы находят отражение в конкретных схемотехнических решениях, используемых при построении соответствующих систем. Так, в некоторых из них используются функциональные сенсорные компоненты, собирающие определенные статистические данные на канальном или сетевом уровне и складывающие их во внутренние базы данных или генерирующие поток этих данных по какому-либо сетевому адресу.

В качестве функциональных компонентов, выполняющих сбор соответствующих потоков, выделяются такие элементы структуры, как коллекторы, которые обеспечивают накопление и структуризацию данных с последующей их записью в специализированных форматах в файлы или в базы данных. В ряде систем мониторинга функции коллектора могут быть объединены с функциями опроса сетевых устройств. При этом для сбора сетевого трафика могут использоваться различные протоколы (RMON, SNMP, IPFIX).

В качестве примера на рис. 1 приведена схема, в наиболее полном виде обеспечивающая сбор и анализ информации с использованием технологии NetFlow. В ней можно обнаружить наличие всех функциональных элементов, так или иначе присутствующих в различных системах. Если при этом проанализировать системы мониторинга, выполненные на чисто программных компонентах, то можно обнаружить, что здесь все они присутствуют.

Что касается систем отображения полученной информации, то они должны включать в себя также и системы формирования отчетов, и средства анализа и интерпретации результатов с использованием интерактивных графических возможностей. Соответствующие функциональные компоненты могут быть интегрированы в одной вычислительной системе или распределены по различным вычислительным системам, машинам и т. п.

Пользовательский интерфейс систем мониторинга обычно реализуется либо в виде специализированного клиентского приложения для персональных систем, либо посредством доступа через www-интерфейс.

4. Задача анализа связности потоков данных субъектов сети и их реализация

В системе мониторинга сетевого ресурса ННЦ можно обнаружить следы использования всех перечисленных выше технологий мониторинга, включая системы, поставляемые и поддерживаемые производителями сетевого оборудования, программные системы третьих производителей, а также свободно распространяемые по сети Интернет "открытые" системы. Однако зачастую возникают задачи, решение которых с помощью "чужих" систем оказывается затруднительным. Иногда такое решение просто невозможно, иногда оно требует такого количества операций, которое не позволяет получить ответ на сложный вопрос в режиме реального времени.

Примером такого рода "сложных" вопросов сетевого мониторинга является анализ динамики происходящих в сети процессов:

— наблюдение за текущей активностью сетевых субъектов;

— сопоставление поведения различных сетевых субъектов;

— анализ структуры используемых при взаимодействии протоколов;

— выделение в интерактивном режиме взаимодействующих субъектов;

— анализ информации уровня приложений для выбранных объектов.

Во многих случаях обеспечение в автоматическом режиме сбора комплексных параметров, соответствующих процедур анализа и последующей оценки его результатов позволяет в динамике отслеживать возникающие в сети аномальные процессы и принимать адекватные решения по управлению сетевыми ресурсами.

Одной из важных задач мониторинга сети ННЦ является анализ связности потоков данных между абонентами сети. Результаты решения этой задачи могут быть использованы не только в целях оперативного управления сетью, но и для принятия долгосрочных административных мер, таких как выделение тех или иных "устойчивых"

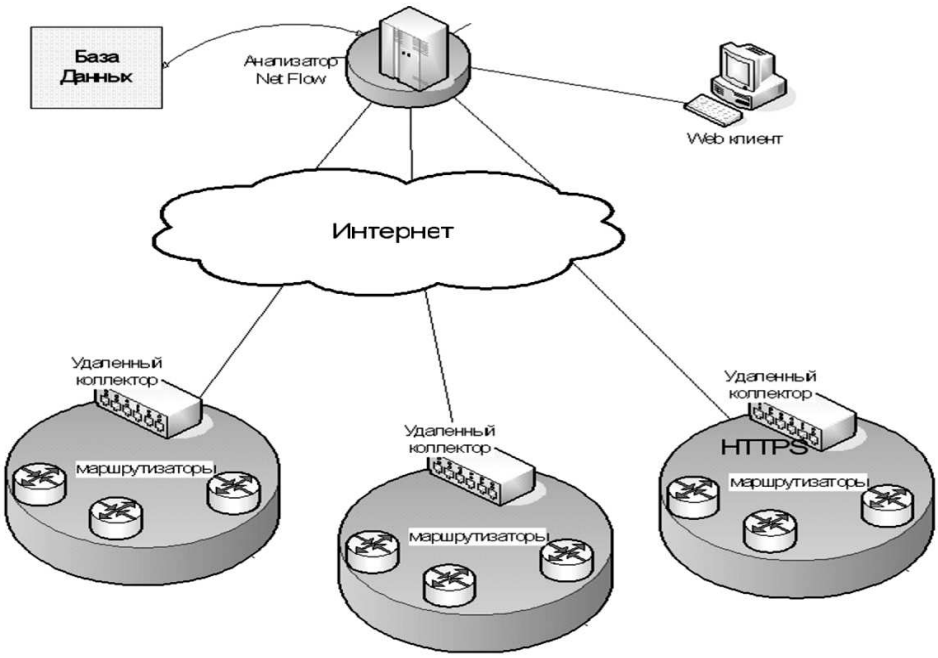

Рис. 1. Схема сбора и анализа информации с использованием технологии NetFlow

внутренних потоков данных в корпоративной сети, а также для детального анализа приоритетных внешних направлений и выделения взаимосвязанных потоков данных с теми или иными корпоративными сетями.

Эффективность решения задач мониторинга потоков данных существенно зависит от возможности гибкого определения сетевого субъекта. В некоторых системах мониторинга такое определение ограничивается неким блоком адресов, в более мощных системах анализа это реализуется на уровне автономных систем. В силу ряда специфических особенностей СПД СО РАН здесь в качестве сетевых субъектов наблюдения (анализа) выступают более сложные образования, агрегаты, представляемые в пространстве сетевых адресов. То есть каждый субъект при анализе потоков между ними определяется некоторым агрегатом либо из списка единичных адресов или списков таких адресов, либо из последовательных блоков адресов или списков подобных блоков адресов. Такое решение позволяет анализировать взаимодействия как отдельных машин или серверов, так, например, и подмножеств организационных структур — институтов (их агрегаций) или совокупностей научных центров СО РАН и их агрегаций.

Стратегия построения комплекса программ для решения рассматриваемого круга задач определяется требованием сохранения неизменности информации, поступающей в соответствующую базу данных от уже использующейся в СПД СО РАН системы сбора данных. Для обеспечения сохранности существующих систем в схему, представленную на рис. 1, включен промежуточный уровень (рис. 2). На этом уровне выделяются такие компоненты, как:

— редиректор, обеспечивающий функции "размножения" потоков данных или каких-либо их частей;

— система постоянных и оперативно задаваемых (выбираемых) фильтров;

— базы данных для оперативного хранения выбранных потоков данных и их последующих сопоставлений и анализа.

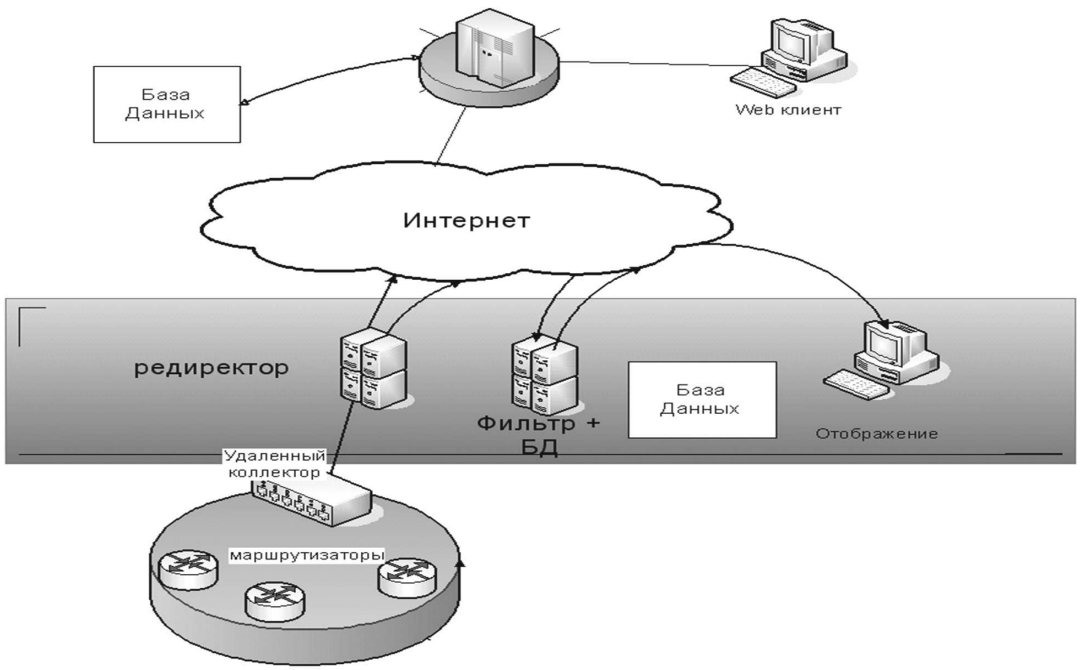

Рис. 2. Модифицированная схема сбора и анализа информации

Одним из важных компонентов этой схемы является система отображений и интерактивного задания параметров отображаемых процессов.

Все эти компоненты в зависимости от специфики решаемых задач могут развертываться как на одной платформе, так и на изолированных и функционально выделенных серверах, поддерживающих необходимые сервисы.

5. Специализированное программно-техническое обеспечение

Обсуждаемый в этом разделе комплекс программно-технических средств предназначен для оценки потоков данных между сетевыми субъектами (агрегатами), представленными адресными сегментами или адресами абонентов сети, или агрегатами, состоящими из списков адресов абонентов. При этом значения, описывающие потоки между сетевыми субъектами, представляются в виде элементов двумерных матриц. Соответствующий комплекс программ включает графические средства отображения и анализа динамики потоков, обеспечивает соответствующей информацией технологические службы сети, использующие результаты такого мониторинга для последующего принятия управленческих решений.

Комплекс содержит набор функциональных модулей, программ (клиентов):

— клиент перенаправления, копирования или порождения новых потоков NetFlow;

— клиент маркирования записей в соответствии с принадлежностью записей сетевым субъектам в смысле определенных выше понятий;

— подсистема накопления, маркирования и фильтрации информации, ее обработки по сетевым субъектам;

— набор графических подсистем отображения в реальном времени данных, полученных на предыдущем этапе, в различных формах;

— другие средства визуализации во времени, в пространствах сетевых адресов, портов и протоколов.

Базовой технологией, используемой при решении обсуждаемого класса задач, является технология NetFlow, которая предоставляет возможность анализа сетевого трафика на уровне сеансов, делая запись о каждой транзакции TCP/IP.

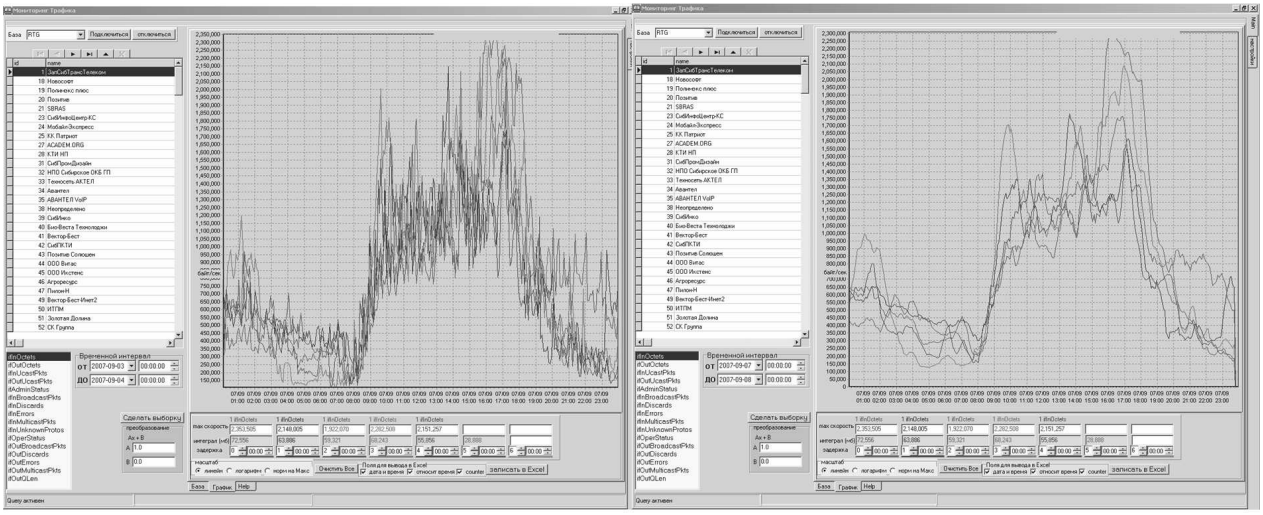

Рис. 3. Пример анализа поведения ежедневного трафика

На рис. 3 приведен пример анализа поведения ежедневного трафика на одном из выбранных сетевых агрегатов с учетом сдвига на 24 ч и с разной степенью усреднения методами скользящего окна (соответственно 5 и 40 мин). На рис. 4 представлено окно вывода трафика через набор выбранных интерфейсов сети. Для каждого выбранного интерфейса задаются выводимые параметры и интервал вывода. Информация в окне обновляется и сдвигается каждые 5 или 1 мин (темп обновления базы данных).

На рис. 5 представлено окно интерфейса, предназначенного для функции размножения NetFlow-потока. Эта функция получает поток с NetFlow-коллектора и копирует

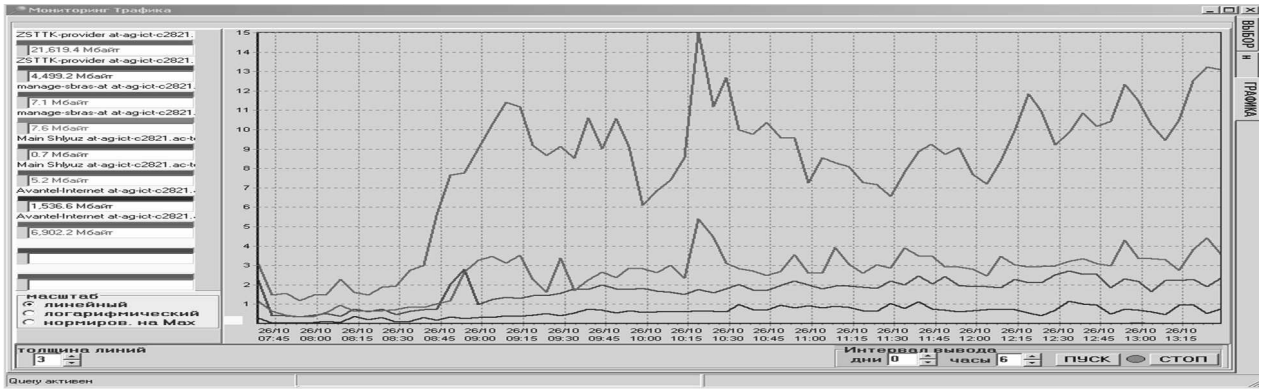

Рис. 4. Окно вывода трафика

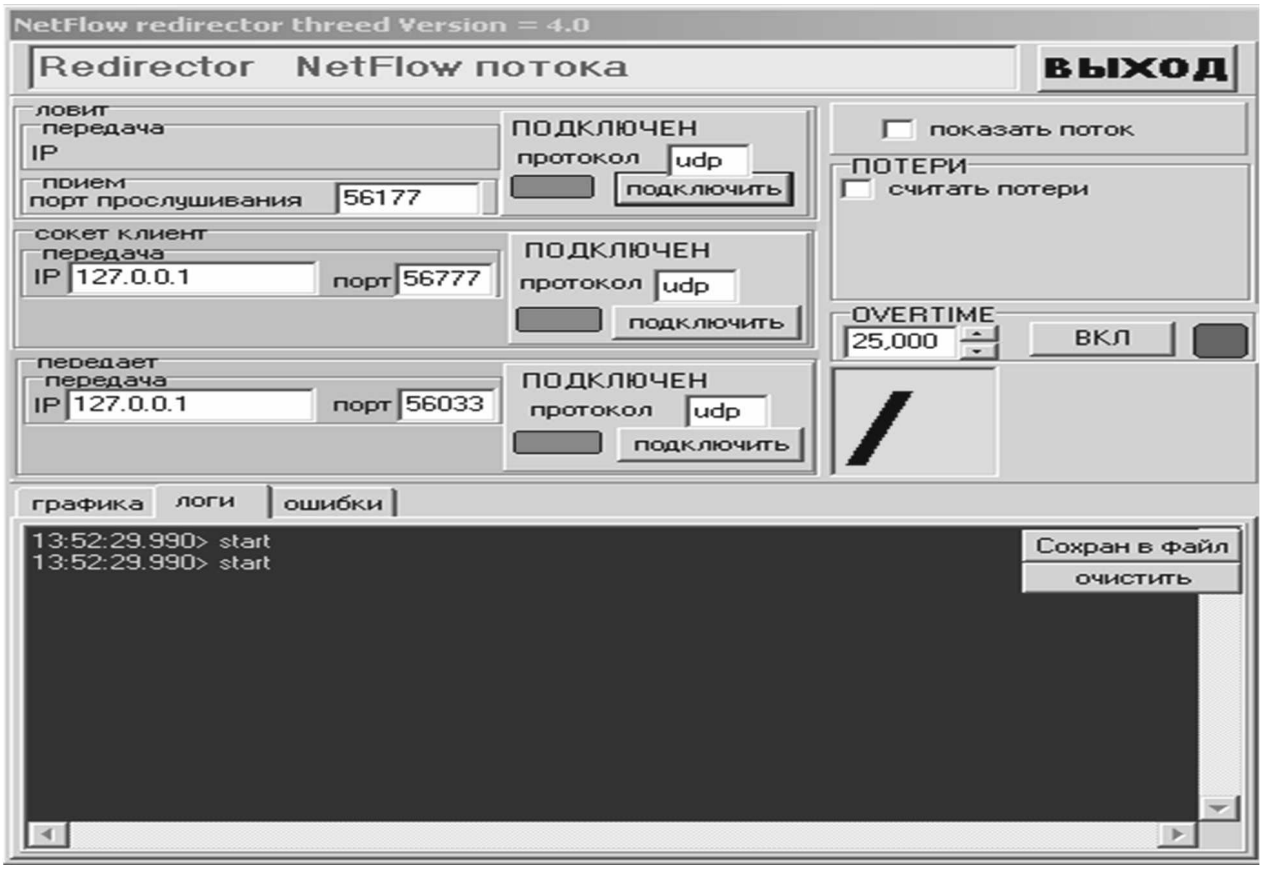

Рис. 5. Окно интерфейса функции размножения NetFlow-потока

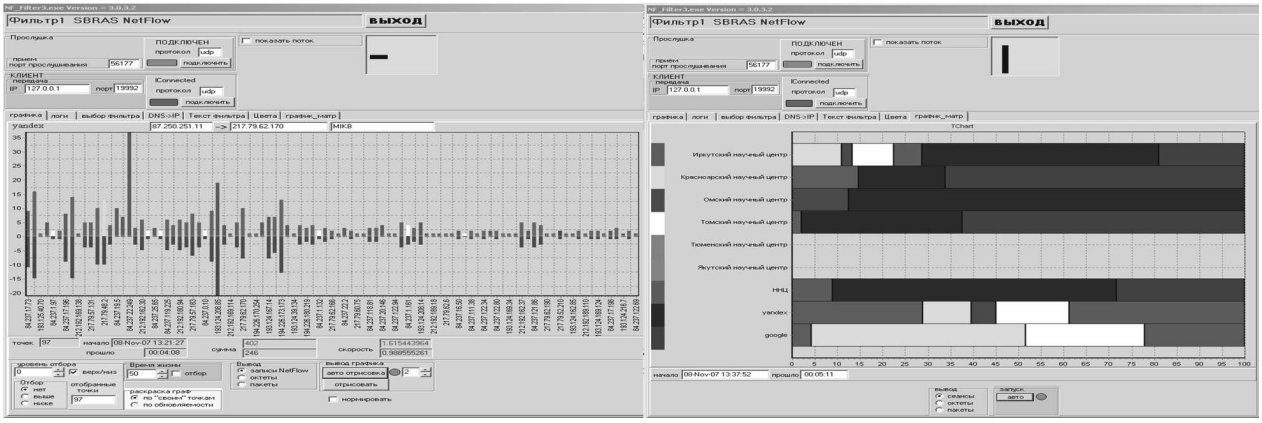

Рис. 6. Примеры интерфейсов для потоковой фильтрации NetFlow-потока

его в несколько потоков для дальнейшей распределенной обработки данных. Программа позволяет:

— выбрать значение порта, на который принимается UDP NetFlow-поток,

— выбрать значение IP и портов-приемников NetFlow-потоков,

— подсчитать потери во входном NetFlow потоке.

На рис. 6 представлены примеры интерфейсов, предназначенных для потоковой фильтрации NetFlow-потока. Соответствующая функция получает этот поток от программы NFRedirtct или напрямую с коллектора NetFlow, производит предварительную обработку в соответствии с выбранным режимом и описанием фильтра из выбранного XML-файла, выводит результат в графическом виде на экран монитора или в выходной поток. Программа NF-Filter 3 позволяет выбрать:

— порт, на который поступает NetFlow-поток;

— IP и порт, на который будет направлен выходной поток;

— вид фильтрации и графического вывода, в том числе вставить в записи выходного потока информацию результата фильтрации для использования ее в дальнейшей обработке и вывести в графическом виде результаты фильтрации.

К результатам фильтрации относятся матрица связанности потоков и график распределения сетевого параметра: сессии, пакеты, от выбранной точки (отдельный IP-адрес или группа IP-адресов (например, некоторое предприятие)) по IP-адресам партнеров по сетевому взаимодействию.

Заключение

При исследовании динамики поведения потоков данных существенной является возможность обнаружения и наблюдения корреляционных связей между параметрами процессов. Интересные результаты просматриваются, например, при анализе матрицы взаимодействия потоков. В тех случаях, когда задаются конкретные сетевые адреса, удается обнаруживать ряд процессов, которые маскируются при использовании других средств анализа. После выявления таких взаимосвязей они могут быть включены в процесс управления потоками данных, в общую систему управления сетью передачи данных.

Решение задач, упомянутых в настоящей статье, необходимо для поддержки эффективного управления сетью, обеспечения надежности ее функционирования, гарантированного качества обслуживания абонентов и безопасности, а также для сбора статистики и детального контроля выполнения абонентами сети установленного регламента работы.

Авторы благодарят Л.Б. Чубарова за полезные обсуждения результатов работы и помощь в подготовке статьи.

Поступила в редакцию 14 марта 2008 г.