Анализ трафика TCP/IP на основе методики допустимого порога и отклонения

Г.В. Бабенко, С.В. Белов
ФГОУ ВПО АГТУ, Астрахань

Современный вектор развития общества тесно связан с процессом информатизации и совершенствования инфокоммуникационных технологий. Одновременно с всеобщей информационной интеграцией стремительно возрастает количество информации передаваемой с использованием сетевых технологий, что влечет за собой рост количества угроз информационной безопасности, возникающих при сетевом взаимодействии [1]. Сегодня практически все сетевые системы защиты информации построены на основе двух архитектур: сигнатурной и поведенческой. Данные архитектуры имеют как достоинства, так и недостатки. Так системы, построенные с использованием сигнатур, в качестве индикатора определения нарушений в сети, не могут выявлять нарушения, незаложенные в данные сигнатуры, а поведенческие системы характеризуются частыми ложными срабатываниями. Объектом анализа, в независимости от архитектуры, у систем защиты данного типа выступает сетевой трафик, так как информация, как служебная, так и передаваемая в сетевых пакетах, является источником, характеризующим все сетевые взаимодействия.

Определено, что информация о сетевом трафике имеет статистический характер и представляет собой временные последовательности [4]. Методы статистического анализа сетевого трафика широко освещены при их использовании в качестве инструментов прогнозирования загруженности каналов связи, определения потерь, качества предоставления услуг и т.п. С точки зрения обеспечения сетевой безопасности проводить анализ сетевого трафика с целью выявления аномального поведения (сбоев в работе, негативного внешнего воздействия, непреднамеренных нарушений) системы крайне необходимо, как для решения задач сетевого администрирования, так и для мониторинга корректного функционирования инфраструктуры компьютерных сетей.

Как класс статистический анализ относится к поведенческим методам определения нарушений в сети и основан на сопоставлении текущего состояния сетевой инфраструктуры с некими определенными заранее признаками, характеризующими штатное функционирование сетевой инфраструктуры. Методы статистического анализа имеют различные интерпретации, основанные на различных динамических характеристиках сетевого трафика, однако, базовые принципы практически у всех идентичны. Неоспоримым преимуществом применения методов статистического анализа является потенциальная возможность определения впервые реализовываемых методов негативного воздействия на объект атаки со стороны злоумышленника. Однако для его успешной реализации необходимо определить объект анализа, иметь определенные структурированные характеристики, образующие корректную конфигурацию и меть критерии, по которым можно определить потенциальную угрозу сетевой безопасности. А для уменьшения вероятности получения ошибочных результатов анализа, основного недостатка применения поведенческих методов выявления инцидентов информационной безопасности, при реализации комплексного подхода положенного в основу разработки автоматизированной системы анализа сетевой инфраструктуры (АС2-И) метод статистического анализа использовался совместно с компоненто-независимыми методами нейросетевого и сигнатурного анализа.

Таким образом, основными задачами, решаемыми в работе, являлись определение анализируемых и контролируемых характеристик сетевого трафика, разработка алгоритма процесса анализа, построение шаблона штатного функционирования сетевой инфраструктуры (ШШФС), определение методик сравнения текущего и корректного состояний характеристик сетевого трафика, а также интеграция статистических методов в комплексную систему анализа трафика TCP/IP.

Характеристики объекта анализа и алгоритм процесса

Статистические методы универсальны, поскольку для проведения анализа не требуется знания о возможных атаках и используемых ими уязвимостях и основаны на изменении некоторых статистических характеристик потока пакетов. Для решения задачи применения статистических методов анализа TCP/IP трафика, необходимо выделить основные показатели, характеризующие штатное функционирование сетевой инфраструктуры и осуществлять динамический контроль над их состоянием. В качестве таких показателей должна выступать информация, по которой можно проанализировать историю сетевого взаимодействия. К данным, которые могут быть проанализированы при захвате трафика TCP/IP, относятся поля заголовков протоколов IP, TCP, UDP, ICMP и содержимое полей данных [3]. Для сокращения времени на обработку и последующее извлечение показателей для формирования ШШФС и анализа, из всей поступающей информации необходимо выделить только информативные служебные ее части. В общем виде алгоритм процесса анализа сетевого трафика выглядит следующим образом (рис. 1).

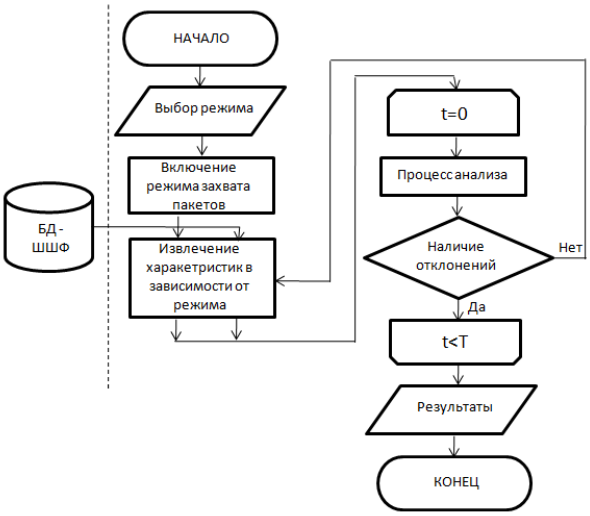

Рис 1. Алгоритм процедуры контроля сети

При включении режима захвата сетевых пакетов, сетевой адаптер переводится в режим «Promiscuous mode» и фиксирует любой пакет, прошедший через интерфейс. Это обусловлено технологией передачи информации в сетях Ethernet. Установив определенные ограничения в режиме анализа, существует возможность уменьшить объем анализируемой информации, отсекая излишки неактуальных данных и увеличив быстродействие системы. Для решения задачи проведения анализа необходимо заблаговременно сформировать актуальный ШШФС. Для этого необходимо выполнить начальные шаги алгоритма, предшествующие процессу анализа и сохранить накопленную информацию. В качестве хранилищ используются как SQL-база данных, так и сформированные двоичные текстовые файлы, хранящие числовые характеристики. Корректность, актуальность данных и четкость временных интервалов ШШФС в данном случае определяются администратором системы и являются одной из основных составляющих получения корректных результатов анализа.

Для определения характеристик потока сетевого трафика, необходимо определить взаимодействующие логические сущности. Установим под логической сущностью потока IP-адреса и порты отправителя и получателя пакетов. Таким образом, методы статистического анализа будут применяться на определенном временном интервале для сетевых пакетов, имеющих одинаковые логические сущности, что позволит более детально проанализировать статистику сетевых взаимодействий. Также для использования в пороговой методике вместо количества полученных или отправленных пакетов будем фиксировать объемы данных, т.е. длину пакета. Это объяснимо стандартом формирования сетевых пакетов RFC, где каждый не фрагментированный кадр (Ethernet-

frame) имеет определенную постоянную величину, что обуславливает использование не количественной характеристики по числу принятых или отправленных пакетов, а объемной по количеству информации обработанной сетевым интерфейсом.
Таким образом, получим следующую последовательность характеристик:

- Логические сущности: адреса источника/приемника – *S/D*.
- Тип и порт пакета - *Tp*.
- Длина пакета - *L*.
- Время фиксации пакета - *Tm*.

Следовательно, любое зафиксированное событие можно описать вектором-объектом события *Tr=<S/D, Tp, L, Tm>*. Время фиксации пакета необходимо для определения выборки событий, удовлетворяющих интервалу анализа. Зададим за *T* некий постоянный временной интервал. Его устанавливает администратор системы в зависимости от преследуемых целей. Анализируемая статистика может быть как текущей или частной (при малых значениях *T*), так и долговременной или глобальной (при значении *T* от нескольких часов или суток). Кроме временной составляющей присутствует и сам объект анализа. Его образует информация о сетевых взаимодействиях: источник, получатель, тип пакета, порт обращения. Таким образом, при анализе имеем выборку из потока зафиксированных событий по детектированию пакета, прошедшего через сетевой интерфейс. Данная выборка позволяет определить активность источника или приемника по количеству обращений, полученным отправленным пакетам, и образовывает объемную характеристику. Также информация о типах пакетов и тем более об используемых портах позволяет более детально персонифицировать процессы, происходящие в сети. Применим пороговую методику к объемной характеристике, а за основу частотного анализа применим критерий среднеквадратического отклонения.

Процесс анализа

На этапе процесса анализа трафика, на основе пороговой методики, из события *Tr* извлекаем объект *X=Tr<L>*, что соответствует длине зафиксированного пакета. Пусть X_i событие из множества событий *X*, в момент времени t_i, при $0 < t_i < T$, где *T* заранее установленный период анализа. Событие определяется изменением состояния сетевого адаптера: прием-отправка пакета с постоянной логической сущностью. Тогда Y_i аналогичный набор событий из множества, составляющего шаблон штатного функционирования сети. Пороговую методику применим в два этапа. На первом этапе сопоставим значения X_i и Y_i в момент времени t_i. При заданном уровне «чувствительности» определяется допустимое значение возможного отклонения. В работе применяется значение коэффициента чувствительности *k=0,8*. Тогда нижний порог определяется как $X_i > k * Y_i$, а верхний как $X_i < Y_i / k$. На втором этапе использования пороговой методики необходимо определить краевые значения допустимых интервалов. Для этого определим выборочное среднее из элементов множества $X_i \in X$ (функция 1):

$$Average(X) = \frac{1}{n}\sum_{i=0}^{n} X_i \qquad (1)$$

Допустимый диапазон определяется следующим неравенством: $3/2 * Average(X) > X_i > average(X)/2$. Аналогичные операции применяются к множеству *Y,* и определяют границы полуинтервалов на основе среднего значения выборки (рис. 2). Основываясь на характеристике потока трафика многократное выявление отклонений от заданного интервала, определенно свидетельствует об изменениях в процессе функционирования сети. Это могут быть, как и легальные изменения в нагрузке на узлы сети, так и иные несанкционированные действия.

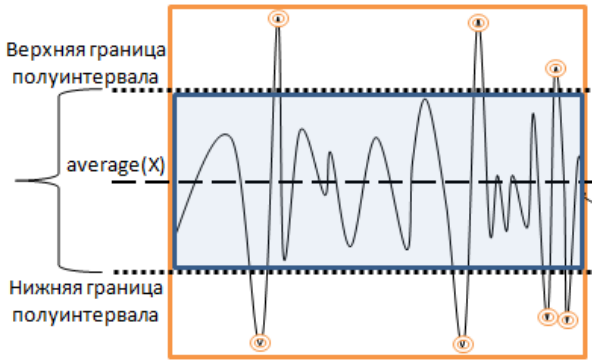

Рис 2. Спектр сетевого потока

Таким образом, имеем три признака: «пороговый на основе ШШФС», «доверительного интервала», «доверительного интервала на основе ШШФС». Использование трех признаков сравнения характеристик позволяет с более высокой степенью вероятности выявить именно действия, направленные на нарушение безопасного сетевого взаимодействия. В результате система оповещает администратора набором следующих сообщений: «В период *T* у сущностей *Tr*<*S/D*> обнаружены отклонения от потока событий <*X*> и/или <*Y*> по признаку [1 (0,1), 2 (0,1), 3 (0,1)]».

Далее рассмотрим иные детектируемые характеристики трафика. Пусть имеется набор типов пакетов $Tr < Tp >= P\{p_1, p_2, ..., p_n\}$, тогда значения X_{p_i} - количество зафиксированных пакетов p_i-типа. При этом $Y = \sum_{i=0}^{n} X_{p_i}$ - общее количество пакетов. Тогда частота фиксации определенного типа пакетов вычисляется как $\rho(X_{p_i}) = X_{p_i} / Y$.

Аналогично определяется частота использования протоколов (служб) при взаимодействии в сети. В общем случае номера портов лежат в диапазоне значений от 0 до 65536. Данная информация составляет характеристику сетевых процессов более детально, т.к. позволяет персонифицировать логические сущности по использованию конкретных служб и приложений. При этом определяются значения, лежащие в интервале $\rho(X_{p_i}) \in [0,1]$.

Частоты признаков

Тип пакета	TCP	UDP	ICMP	ARP	IP
Количество пакетов определенного типа- X_i	578	234	657	35	834
Частота появления- $\rho(X_{p_i})$	0,2	0,15	0,25	0,1	0,3
Идентификаторы источника и приемника являются постоянными					
Порт	HTTP-80	FTP-21	143	408	RDP-1389
Количество пакетов определенного типа- X_i	14566	3321	455	4235	354
Частота появления- $\rho(X_{p_i})$	0,9	0,03	0,02	0,04	0,01

В данном случае в качестве критериев пороговой методики используем математическое ожидание и среднеквадратичное отклонение. При этом критерий математического ожидания определяется по функции 2:

$$M(X) = \sum_{i=0}^{n} X_{p_i} * \rho(X_{p_i}) \quad (2)$$

,

где при достаточно большом размере выборки математическое ожидание стремится к среднему значению. Также критерий среднеквадратичного отклонения выборки определяем с использованием функции 3, для событий текущего потока:

$$\sigma = \sqrt{M(X^2{}_{p_i}) - (M(X_{p_i}))^2} \quad (3)$$

.

Среднеквадратичное отклонение является наиболее распространённым показателем рассеивания значений случайной величины относительно её математического ожидания. Получаем, что любое анализируемое событие является аномальным, если оно не попадает в границы ожидаемых значений (рис. 3).

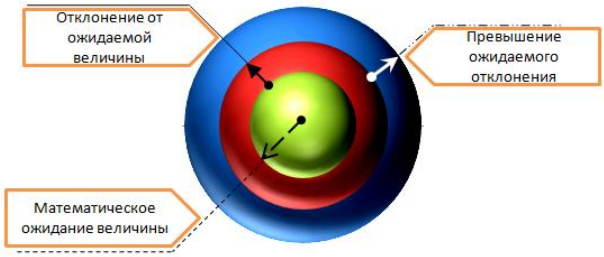

Рис 3. Области ожидаемых значений

В случае использования ШШФС необходимо анализировать характеристики потока с данными заложенными в шаблон. В этом случае среднеквадратическое отклонение определяется по функции 4.

$$\sigma = \sqrt{\frac{1}{n}\sum_{i=0}^{n}(X_{p_i} - Y_{p_i})^2} \qquad (4)$$

Обычно в качестве Y_{p_i} выступает некоторое ожидаемое значение, определяемое как выборочное среднее, а именно $Y_{p_i} = Average(X_{p_i})$. Однако, имея ШШФС, ожидаемым значением определяется характеристика потока сети в момент времени соответствующий характеристике X_{p_i} из шаблона. Таким образом, имеем два признака: «отклонение потока», «отклонение потока от ожидаемого ШШФС».

При статистическом анализе необходимо учитывать, что трафик компьютерных сетей обладает свойством масштабной инвариантности - имеет особую фрактальную (самоподобную) структуру, сохраняющуюся на разных масштабах [5]. В процессе передачи возникают большие всплески при относительно низком среднем уровне трафика, что крайне важно учитывать при статистическом анализе.

Анализ результатов указывает на то, что значения критерия математического ожиданий отдельных характеристик для выборок, характеризующих штатное функционирование, принимают меньшие значения по сравнению с данными, содержащими отклонения, как и значения выборочных средних. Аналогично, анализ значений среднеквадратических отклонений, для данных, содержащих отклонения, показывает преобладание больших по величине, чем для значений, характеризующих штатное функционирование. Это указывает на то, что при появлении отклонения к общему штатному режиму функционирования вливаются значения, большие по абсолютной величине, что вызывает колебания уровней характеристик и является причиной увеличения дисперсии и отклонений.

Производя периодический или постоянный анализ состояния сетевой инфраструктуры, появляется возможность идентифицировать действия злоумышленников, направленные на подготовку к проведению атак: сканирование портов, службы, получение информации об используемом прикладном, системном программном обеспечении и системах защиты (на этапе рекогносцировки). Четко зафиксировать время начала негативного воздействия на объект атаки (вторжение), из-за его краткосрочности, достаточно затруднительно (что не исключает такой возможности), однако, определить наличие последующего воздействия на объект при помощи методов реализованных в системе является возможным, после чего по отношению к источнику угрозы необходимо применять иные меры воздействия. В итоге даже визуальный анализ построенных характеристик позволяет выделить информацию об изменениях в долях используемых протоколов (служебных, прикладных, пользовательских), фиксируются изменения в нагрузке на узлах сети (ICMP-Flood,

запросы), по количеству зафиксированных пакетов, определяется объем переданной/полученной информации, регистрируется теневое использование ресурсов сети и т.п.

При реализации вышеописанных методов в автоматизированной системе были созданы хранилища данных, содержащие характеристики сетевого потока - ШШФС. Для улучшения визуального восприятия получаемых результатов анализа разработана двумерная динамическая характеристика сетевой активности. У администратора системы имеется по два динамических поля отображающих как текущую характеристику сети, так и информацию загруженную администратором из ШШФС (рис. 4).

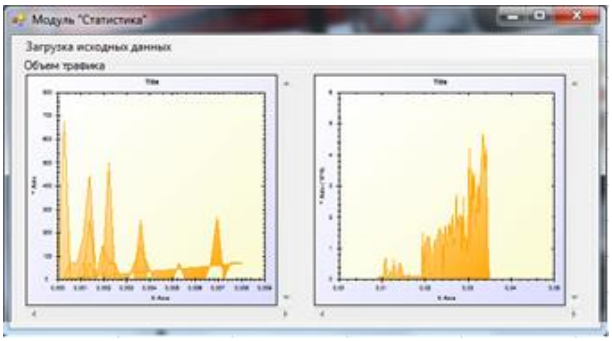

Рис 4. Характеристика TCP/IP: объемная

Так же имеется аналогичная частотная характеристика детектирования типов пакетов и служб активации (рис. 5).

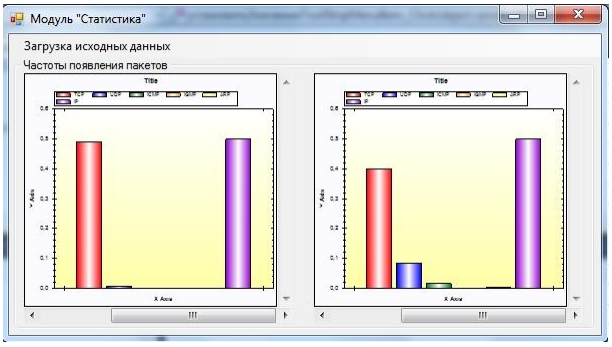

Рис 5. Характеристика TCP/IP: частотная

Описанные выше методы обнаружения попыток нарушения безопасности в сети основаны на том обстоятельстве, что в этом случае могут изменяться некоторые статистические характеристики потока пакетов. Например, в случае определенного вида воздействия изменяется отображаемая характеристика сетевого трафика. Так при сканировании сети, увеличивается доля пакетов типа ICMP, для получения доступных компонентов сети, а также пакетов TCP с предустановленными флагами контроля и т.п. В этом случае обнаружение нарушений основывается на визуальном сравнении текущих характеристик потока пакетов, с характеристиками положенными в ШШФС, а также анализе сообщений автоматического контроля сетевой активности. Необходимо также учитывать актуальность и корректность построения ШШФС, т.к. ошибки в его построении могут оказывать влияние на процесс анализа, непосредственно связанного с сопоставлением характеристик из ШШФС.

Заключение

Сопоставляя результаты, можно сделать вывод, что диапазоны значений характеристик штатного функционирования сетевой инраструктуры, имеют относительное постоянство уровней значений математических ожиданий и дисперсий. При возникновении нештатной ситуации аналогичные оценки по различным характеристикам изменяются значительно.

Стоит отметить, что эффективность применения данных методов зависит от временного периода анализа, составляющего порядка 30 минут на итерацию для анализа малых выборок, и порядка суток для получения глобальной картины сетевой активности. С целью уменьшения вероятности появления ошибок первого рода, характерных для такого рода процессов анализа, рекомендуется применять иные компонента независимые методы анализа трафика не приводящие к наследованию некорректных результатов. Как описано выше, для решения задачи интеграции в АС2-И применены в совокупности методы статистического, нейросетевого и сигнатурного анализа с сетевым трафиком в качестве объекта анализа. Выбор данных методов обусловлен их «компоненто-независимостью» при анализе данных, что уменьшает вероятность наследования ошибок 1-ого и 2-ого рода [6]. Иной особенностью применения данных методов является возможность определения проблем не только внешнего негативного воздействия, но и вопросов корректного функционирования компонентов сетевой инфраструктуры.

Литература

1. Бабенко Г.В. Анализ современных угроз безопасности информации, возникающих при сетевом взаимодействии. // Вестник Астраханского государственного технического университета. Серия: "Управление, вычислительная техника и информатика", 2010. №2. –С. 149-152.
2. Андронов А.М., Копытов Е.А. Теория вероятностей и математическая статистика. СПб.: «Питер», 2004. – 461 с.
3. В.Г. Олифер. Н.А. Олифер. Компьютерные сети. Принципы, технологии, протоколы. Учебник для вузов. 4-е изд. – СПб.: Питер, 2010. – 944 с.
4. Иванов В.В. Статистическая модель сетевого трафика: автореферат диссертации / В.В. Иванов. – Дубна., 2009. – 30 с.
5. Петров В.В. Структура телетрафика и алгоритм обеспечения качества обслуживания при влиянии эффекта самоподобия: автореферат диссертации / В.В. Петров. – М., 2004. – 20 с.
6. Бабенко Г.В. Систематизация методов анализа сетевых взаимодействий. // Наука-Поиск. - 2010. –С. 56-58.