ИНФОРМАТИКА

2006 **июль-сентябрь** **№ 3**

ИНФОРМАЦИОННАЯ БЕЗОПАСНОСТЬ

УДК 681.324

Ю.В. Земцов

ОБНАРУЖЕНИЕ АНОМАЛЬНОЙ АКТИВНОСТИ НА ОСНОВЕ УСЕЧЕННОЙ ПРОЦЕДУРЫ ПОСЛЕДОВАТЕЛЬНОГО АНАЛИЗА

Предлагается новый метод обнаружения аномальной сетевой активности в реальном масштабе времени, основанный на усеченной процедуре статистического последовательного анализа. Разработанный на базе предложенного метода алгоритм реализован программно и испытан в реальных условиях, показав пригодность к быстрому и достоверному обнаружению атак рекогносцировки, а также активности программных сетевых червей.

Введение

Для повышения уровня информационной безопасности в вычислительных сетях правительственных учреждений и различных организаций на современном этапе актуальна разработка методов и алгоритмов для систем обнаружения атак [1]. Подобные системы представляют собой специализированные программные или программно-аппаратные средства, которые позволяют прогнозировать, выявлять, предупреждать и реагировать в реальном масштабе времени на угрозы безопасности [2].

В данной статье предлагается новый метод обнаружения программных сетевых червей и атак рекогносцировки в реальном масштабе времени. Хотя данный метод и относится к классу статистических методов обнаружения аномальной активности, он имеет отличительную особенность, заключающуюся в том, что моделируется не только активность законных пользователей, но также и предполагаемые действия нарушителей. Предпосылкой к подобному моделированию послужило экспериментально подтвержденное предположение о том, что вероятность успешно подключиться к сетевым службам для законных пользователей значительно превышает вероятность подключения злоумышленников. Такое предположение позволяет формально поставить и решить задачу обнаружения аномальной активности на основе усеченной процедуры последовательного анализа. Как показывают испытания, использование статистического последовательного анализа для решения данной задачи дает возможность существенно повысить точность обнаружения нарушителей, а также сократить требуемое для этого время.

1. Современный подход к обнаружению аномальной активности

В настоящее время наиболее эффективный подход к обнаружению аномальной активности заключается в использовании статистического анализа для определения характеристик рассеяния в данных, полученных в результате наблюдения за системой [3]. Статистический анализ может проводиться как для отдельного пользователя, так и для всей системы в целом. Обычно применяются две методики обнаружения нарушителя, использующие статистические методы: определение пороговых значений и собственно выявление аномалий [4].

Анализ статистических параметров работы пользователей, или суммарная статистика, представляет собой одну из самых элементарных форм обнаружения нарушителя, состоящую в определении пороговых значений соответствующих статистических параметров. Цель определения порога – выяснить, когда число появлений отслеживаемого события превысит допустимую границу, ожидаемую при нормальной работе системы. Суть в том, что неестественно большое число событий за короткий период времени может свидетельствовать о появлении нарушителя. Как только пороговое число событий превышено, пороговый детектор либо блокирует источник, либо уведомляет администратора безопасности.

Метод выявления аномалий обеспечивает обнаружение нарушителя без предварительного знания недостатков в защите системы, подвергающейся нападению. Этот подход требует относительно небольшого количества определяемых системой правил, что делает использующие его средства чрезвычайно мобильными. Ниже приведен список нарушений безопасности, а также характеристики этих нарушений, которые могут применятся при определении вторжения на основе выявления аномалий:

– нелегальное проникновение: нарушения часто характеризуются необычным временем и/или способом входа в систему;

– злоупотребление полномочиями: при работе незаконных пользователей может появиться больше, чем обычно, случаев отказа в доступе.

– незаконное распространение информации: незаконную распечатку данных можно выявить по использованию удаленных принтеров или работе в необычное время;

– атаки агрегации и логического вывода: могут быть выявлены по необычно большому числу запросов данных и их поиску;

– троянские кони и вирусы: программы, замещенные троянскими конями, могут отличаться большей загрузкой центрального процессора или интенсивным вводом-выводом; программы, инфицированные вирусом, характеризуются появлением ненормальных модификаций других выполняемых файлов;

– отказ в обслуживании: может характеризоваться ненормально интенсивной деятельностью одного пользователя в отношении атакуемого ресурса, который в то же время ненормально мало используется другими пользователями.

Главным недостатком современного статистического подхода к обнаружению аномальной активности является использование классических методов стационарной статистики, которые не пригодны для краткосрочного прогнозирования, что не позволяет реагировать на угрозу нарушения безопасности в реальном времени. Достаточно редкое обновление базы параметров нормального поведения позволяет нарушителям адаптировать свое поведение к требованиям системы обнаружения аномальной активности, которая в результате воспринимает его как законного пользователя. Игнорирование специфики действий предполагаемого нарушителя приводит к неадекватной реакции системы, что выражается в большом числе ложных срабатываний [5].

2. Формальная постановка задачи обнаружения аномальной активности

Для ликвидации указанных выше недостатков предлагается использовать математический аппарат последовательной проверки статистических гипотез о состоянии защищаемой системы. Существенной чертой предлагаемого метода, в отличие от ныне существующей методики статистического обнаружения аномальной активности, является то, что количество наблюдений, необходимых для обнаружения злоумышленных действий, зависит в данном случае от исхода самих наблюдений и, следовательно, является не определенной заранее, а случайной величиной.

Очевидно, что заключение о возможном проникновении в систему будет тем надежнее, чем устойчивее экспериментальная оценка соответствующей статистической характеристики. Очевидно также, что хотя надежность статистических оценок увеличивается с ростом объема статистики, подсистема обнаружения должна срабатывать достаточно оперативно и располагать ограниченным объемом данных. Ниже предлагаются теоретические основания методики оценки статистических гипотез по сравнению различных последовательностей наблюдений [7, 8], так как практическое обнаружение атаки состоит именно в выявлении последовательности событий, отличающихся от стандартных.

Рассмотрим предлагаемый метод на примере обнаружения злоумышленника, производящего рекогносцировку в корпоративной сети. Удаленный компьютер (УК) делает запрос на соединение с одним из серверов из защищаемой сети. Результатом данного запроса может быть либо успешное соединение, либо отказ в соединении. Таким образом, i-я попытка соединения представляет собой случайную величину Бернулли X_i, принимающую значение 0, если соединение было успешным, и 1 в противном случае. По мере наблюдения $X_1, X_2, X_3, \ldots$ попытаемся

как можно быстрее и в то же время точнее определить, проводит ли УК несанкционированную рекогносцировку в защищаемой сети. Так как принятие решения на этот счет необходимо производить после каждой попытки соединения и в реальном масштабе времени, то удобно воспользоваться методом последовательной проверки гипотез.

Будем рассматривать две простые гипотезы: пусть гипотеза H_0 (нулевая гипотеза) заключается в том, что проверяемый УК осуществляет легальные попытки подключения, а гипотеза H_1 (конкурирующая гипотеза) указывает на злонамеренную активность УК. Для наглядности демонстрации предлагаемого метода и упрощения вычислений будем считать, что случайные величины $X_i \mid H_j$, где $i = 1, 2, 3, \ldots; j = 0, 1$, независимы и одинаково распределены в соответствии с биноминальным законом распределения (точнее, с его частным случаем – распределением Бернулли):

$$\begin{aligned} &P(X_i = 0 \mid H_0) = \theta_0, \quad P(X_i = 1 \mid H_0) = 1 - \theta_0; \\ &P(X_i = 0 \mid H_1) = \theta_1, \quad P(X_i = 1 \mid H_1) = 1 - \theta_1. \end{aligned} \tag{1}$$

Исходя из того, что попытка соединения законного пользователя, вероятнее всего, будет успешной, можно заключить:

$$\theta_0 > \theta_1. \tag{2}$$

С учетом имеющихся гипотез возможны четыре исхода для процедуры принятия решения:
– обнаружение: выбор гипотезы H_1, когда H_1 истинна;
– упущение (необнаружение): выбор гипотезы H_0, когда H_1 истинна;
– ложное срабатывание: выбор гипотезы H_1, когда H_0 истинна;
– штатное состояние: выбор гипотезы H_0, когда H_0 истинна.

Оценивать качество метода обнаружения будем по двум параметрам: вероятности обнаружения P_O и вероятности ложного срабатывания $P_{ЛО}$ [6]. В частности, потребуем, чтобы для выбранных значений a и b выполнялось

$$\begin{aligned} &P_{ЛО} \le a; \\ &P_O \ge b. \end{aligned} \tag{3}$$

На практике обычно выбирают значения $a \in [0{,}001; 0{,}1]$, $b \in [0{,}95; 0{,}999]$.

Таким образом, необходимо по мере поступления потока случайных событий (запросов на соединение) как можно быстрее выбрать одну из двух гипотез (H_0 или H_1), причем с условиями качества (3).

В соответствии с методикой статистического последовательного анализа [8] для каждого наблюдаемого события вычисляется отношение вероятностей:

$$\frac{P(X \mid H_1)}{P(X \mid H_0)} = \prod_{i=1}^{n} \frac{P(X_i \mid H_1)}{P(X_i \mid H_0)},$$

где X – вектор произошедших на момент вычисления событий; $P(X \mid H_i)$ – функция условной вероятности последовательности событий X при условии справедливости гипотезы H_i. Переход к произведению возможен благодаря предположению о том, что случайные величины $X_1, X_2, X_3, \ldots$ независимы и одинаково распределены. На практике более удобным является вычисление логарифма отношения вероятностей, так как это позволяет перейти от произведения к сумме:

$$\ln \frac{P(X \mid H_1)}{P(X \mid H_0)} = \sum_{i=1}^{n} \ln \frac{P(X_i \mid H_1)}{P(X_i \mid H_0)} \equiv K(X). \tag{4}$$

Далее рассчитанное отношение вероятностей необходимо сравнить с нижним и верхним пороговыми значениями η_0 и η_1 соответственно. Если окажется, что $K(X) \le \ln\eta_0$, то принимается гипотеза H_0. Если $K(X) \ge \ln\eta_1$, то принимается гипотеза H_1. Если же $\ln\eta_0 < K(X) < \ln\eta_1$, то необходимо ждать следующего события, чтобы заново рассчитать $K(X)$.

Пороговые значения η_0 и η_1 должны выбираться таким образом, чтобы выполнялось условие (3). При последовательном анализе гипотез принципиальным является тот факт, что для снижения вычислительной сложности пороговые значения должны быть в простой зависимости от a и b, а также не зависеть от распределений (1). Важно отметить, что хотя эти распределения и не влияют на выбор пороговых значений η_0 и η_1, от них зависит количество наблюдений N, необходимое для выбора одной из гипотез и завершения процедуры последовательного анализа.

Можно показать [8], что для η_1 (η_0) существует верхний (нижний) предел, зависящий от P_O и $P_{ЛО}$. Это позволяет аппроксимировать предельные значения η_0 и η_1 посредством замены P_O и $P_{ЛО}$ на a и b соответственно. Пусть имеется n наблюдений X_1, X_2, X_3, …, X_n, причем на n-м событии достигнуто условие $K(X) \ge \ln\eta_1$, тогда

$$\frac{P(X_1,...,X_n \mid H_1)}{P(X_1,...,X_n \mid H_0)} \ge \eta_1 . \tag{5}$$

Таким образом, для любой последовательности наблюдений справедливо следующее утверждение: вероятность $P(X_1,\ldots,X_n \mid H_1)$ по крайней мере в η_1 раз больше вероятности $P(X_1,\ldots,X_n \mid H_0)$, причем это справедливо для произвольной последовательности наблюдений, приведшей к выбору гипотезы H_1 независимо от количества наблюдений n. Таким образом, вероятность всех последовательностей наблюдений, при которых выбирается гипотеза H_1, когда H_1 действительно истинна, будет по крайней мере в η_1 раз больше вероятности всех последовательностей наблюдений, при которых выбирается H_1, но действительно истинна гипотеза H_0. Первая из этих вероятностей представляет собой вероятность обнаружения P_O, вторая – вероятность ложного срабатывания $P_{ЛО}$. Поэтому справедливо следующее выражение:

$$\eta_1 \le \frac{P_O}{P_{ЛО}} . \tag{6}$$

Аналогично можно получить выражение для η_0:

$$\eta_0 \ge \frac{1-P_O}{1-P_{ЛО}} . \tag{7}$$

Предположим, что предельные значения выбираются равными значениям, получаемым заменой P_O и $P_{ЛО}$ на a и b соответственно, тогда

$$\eta_0 = \frac{1-b}{1-a}; \quad \eta_1 = \frac{b}{a} . \tag{8}$$

Так как выражения для граничных условий (6), (7) верны для любых предельных значений η_0 и η_1, то справедливо

$$\frac{b}{a} \le \frac{P_O}{P_{ЛО}}; \quad \frac{1-b}{1-a} \ge \frac{1-P_O}{1-P_{ЛО}} . \tag{9}$$

С учетом того, что $P_O \in (0;1)$, возведем обе части первого неравенства выражения (9) в степень минус единица:

$$P_{ЛО} < \frac{a}{b} \equiv \frac{1}{\eta_1}. \tag{10}$$

Рассуждая аналогично, из второго неравенства в (9) получим

$$1 - P_O < \frac{1-b}{1-a} \equiv \eta_0. \tag{11}$$

Анализируя равенство (10), замечаем, что для выбранных предельных значений в соответствии с выражением (8) вероятность ложного срабатывания $P_{ЛО}$ хотя и может превысить верхнюю границу a, но только незначительно, так как нижняя граница b для вероятности обнаружения P_O близка к единице. Например, если a = 0,03, b = 0,98, тогда вероятность ложного срабатывания не превысит 0,031. Аналогичным образом можно показать, что вероятность пропуска атаки $(1 - P_O)$ не может значительно превысить выбранной верхней границы $(1 - b)$. Из неравенств (9) следует еще одно замечательное свойство данного метода [8]:

$$1 - P_O + P_{ЛО} \le 1 - b + a. \tag{12}$$

Смысл этого неравенства заключается в том, что вероятность пропуска атаки $(1 - P_O)$ и вероятность ложного срабатывания $P_{ЛО}$ не могут одновременно быть выше заданных верхних границ $(1 - b)$ и a соответственно.

Для заданного критерия качества обнаружения (3), а также соответствующих предельных значений (8) необходимо оценить количество наблюдений N, требующихся для принятия одной из гипотез и завершения процедуры проверки. Очевидно, что так как решение о прекращении проверки на некоторой стадии эксперимента зависит от исхода предшествующих наблюдений, то количество требуемых наблюдений является случайной величиной. Теория оптимальных правил остановки позволяет показать [9], что

$$E[N \mid H_0] = \frac{a \ln \frac{b}{a} + (1-a) \ln \frac{1-b}{1-a}}{\theta_0 \ln \frac{\theta_1}{\theta_0} + (1-\theta_0) \ln \frac{1-\theta_1}{1-\theta_0}}; \qquad E[N \mid H_1] = \frac{b \ln \frac{b}{a} + (1-b) \ln \frac{1-b}{1-a}}{\theta_1 \ln \frac{\theta_1}{\theta_0} + (1-\theta_1) \ln \frac{1-\theta_1}{1-\theta_0}}, \tag{13}$$

где $E[N \mid H_0]$ и $E[N \mid H_1]$ – математические ожидания количества наблюдений до принятия гипотез H_0 и H_1 соответственно.

Как видно из выражений (13), $E[N \mid H_0]$ и $E[N \mid H_1]$ зависят от четырех параметров a, b, θ_0 и θ_1, т. е. от вероятности ложного срабатывания, вероятности обнаружения аномальной активности и меры отличия активности легальных пользователей от аномальной активности. Таким образом, чтобы определить количество попыток подключения, которое может предпринять нарушитель до обнаружения его данным методом, необходимо установить значения этих параметров.

Будем полагать, что злоумышленник случайным образом выбирает адреса в защищаемой сети, тогда вероятность того, что подключение будет установлено, зависит от общего количества адресов в сети. На рис.1, *а* показано изменение $E[N \mid H_1]$ при увеличении θ_1. Для a = 0,01, b = 0,99 и θ_0 = 0,75 расчетное значение $E[N \mid H_1]$ = 10 при θ_1 = 0,3 и возрастает до 17 при θ_1 = 0,4. Также из графика на рисунке видно, что при снижении допустимого значения для вероятности ложного срабатывания a растет количество требующихся для принятия решения наблюдений. На рис.1, *б* показана зависимость $E[N \mid H_1]$ от θ_1 при b = 0,99 и θ_0 = 0,95,

по которой хорошо заметно, насколько эффективно работает данный метод при близких к единице значениях θ_0. Напомним, что θ_0 представляет собой вероятность удачных попыток подключения легальных пользователей, а так как в грамотно администрируемых сетях практически все попытки подключения законных пользователей успешны, то обнаружение аномальных подключений в них будет происходить особенно быстро. Ввиду того, что вычислить точные значения параметров θ_0 и θ_1 на практике крайне затруднительно, устанавливать их следует «экспертным образом», по возможности опираясь на данные мониторинга сетевого трафика, журналов аудита операционных систем и сетевого оборудования, а также других систем обнаружения сетевых атак, функционирующих в защищаемой сети. К сожалению, подобным образом определенные параметры неизбежно увеличивают вероятности ошибок. Например, если завысить значение θ_0 относительно реального, то последовательный процесс может завершиться преждевременным принятием гипотезы H_1. В результате этого анализ данных, который, возможно, привел бы к принятию гипотезы H_0, выполнен не будет, что может вызвать ложный сигнал тревоги. Тем не менее, обеспечение робастности [10] предлагаемого метода обнаружения аномальной сетевой активности является задачей будущих исследований в данном направлении.

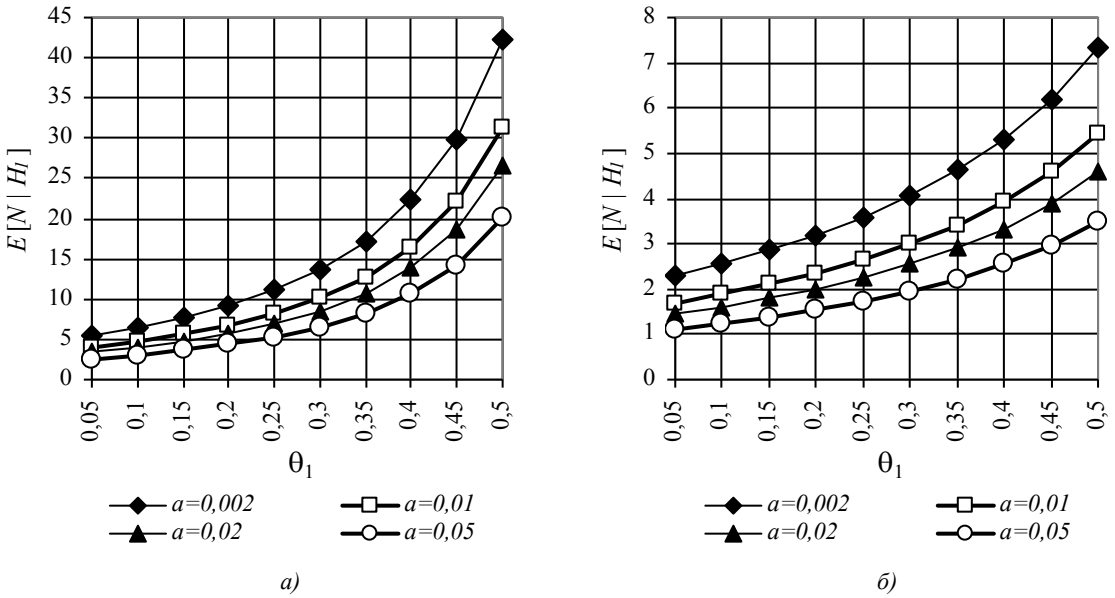

Рис. 1. Графики зависимости $E[N \mid H_1]$ от θ_1 при $a = \{0{,}002;\ 0{,}01;\ 0{,}02;\ 0{,}05\}$, $b = 0{,}99$:
а) для $\theta_0 = 0{,}75$; *б)* для $\theta_0 = 0{,}95$

Для того чтобы данный метод можно было применять для обнаружения аномальной активности, не боясь сделать саму систему обнаружения мишенью для атак отказа в обслуживании, необходимо реализовать возможность установки верхней границы числа наблюдений, например, n_0. Этого можно достигнуть усечением последовательного процесса на $n = n_0$, т. е. установлением новых правил для принятия или отклонения гипотезы H_0 на n_0-м испытании, если последовательный процесс так и не привел к окончательному решению при $n \le n_0$. Простым и разумным правилом для усечения на n_0-м испытании, по-видимому, должно быть следующее: если последовательный процесс не привел к окончательному решению при $n \le n_0$, то после n_0 испытаний принимается гипотеза H_0, если

$$\ln \eta_0 < \sum_{i=1}^{n_0} z_i \le 0\,, \tag{14}$$

и отклоняется H_0, если

$$0 < \sum_{i=1}^{n_0} z_i \le \ln \eta_1 , \tag{15}$$

где для удобства вычислений сделан переход к логарифму отношения вероятностей:

$$z_i = \ln \frac{P(X_i \mid H_1)}{P(X_i \mid H_0)} .$$

В рассматриваемом случае распределений (1) несложно показать, что

$$z_i = \begin{cases} \ln \dfrac{\theta_1}{\theta_0} & \text{при} \quad X_i = 0; \\ \ln \dfrac{1-\theta_1}{1-\theta_0} & \text{при} \quad X_i = 1. \end{cases} \tag{16}$$

Следовательно,

$$K(X) \equiv \sum_{i=1}^{n_0} z_i = n_0^* \ln \frac{\theta_1}{\theta_0} + (n_0 - n_0^*) \ln \frac{1-\theta_1}{1-\theta_0} , \tag{17}$$

где n_0^* – количество успешных соединений среди всех попыток.

Однако необходимо помнить, что усечение последовательного процесса на n_0-м наблюдении изменяет ошибки первого и второго рода, тем самым увеличивая как вероятность необнаружения злоумышленника, так и вероятность ложного срабатывания. Мера влияния усечения будет, конечно, зависеть от значения n_0: чем больше n_0, тем меньше влияние усечения.

3. Алгоритм обнаружения аномальной сетевой активности

Проведенный формальный анализ позволяет представить алгоритм обнаружения аномальной сетевой активности, пригодный для обнаружения в реальном масштабе времени программных сетевых червей, а также атак рекогносцировки (рис. 2).

Процесс обнаружения аномальной сетевой активности начинается с захвата сетевого трафика в смешанном режиме [1], позволяющем сетевой карте обрабатывать все пакеты, передаваемые в данном сетевом сегменте. Далее происходит декодирование трафика в соответствии с набором процедур для декомпозиции пакетов согласно уровням сетевого стека, т. е. принятый кадр последовательно преобразуется в пакет, сегмент и блок данных. Для каждого УК, пытающегося соединиться с одним из серверов защищаемой сети, необходимо вести учет успешных и неудачных попыток соединения до тех пор, пока не будет решено, кем является УК: законным пользователем или нарушителем. После получения очередного запроса на соединение от УК необходимо установить, по какому критерию следует принимать решение. Если верхняя граница числа наблюдений не превышена, то решение следует принимать согласно обычному последовательному критерию, а в случае превышения границы – по критерию с усечением. Затем в соответствии с выражением (17) вычисляется $K(X)$ и в зависимости от действующего критерия либо принимается решение насчет законности производимых попыток соединения, либо продолжаются наблюдения.

Представленный алгоритм реализован программно на языке Си в качестве модуля обнаружения аномальной активности (МОАА) и интегрирован в макетный образец комплекса средств динамической защиты от несанкционированного доступа, разрабатываемого сотрудниками лаборатории защиты информации Объединенного института проблем информатики Национальной академии наук Беларуси.

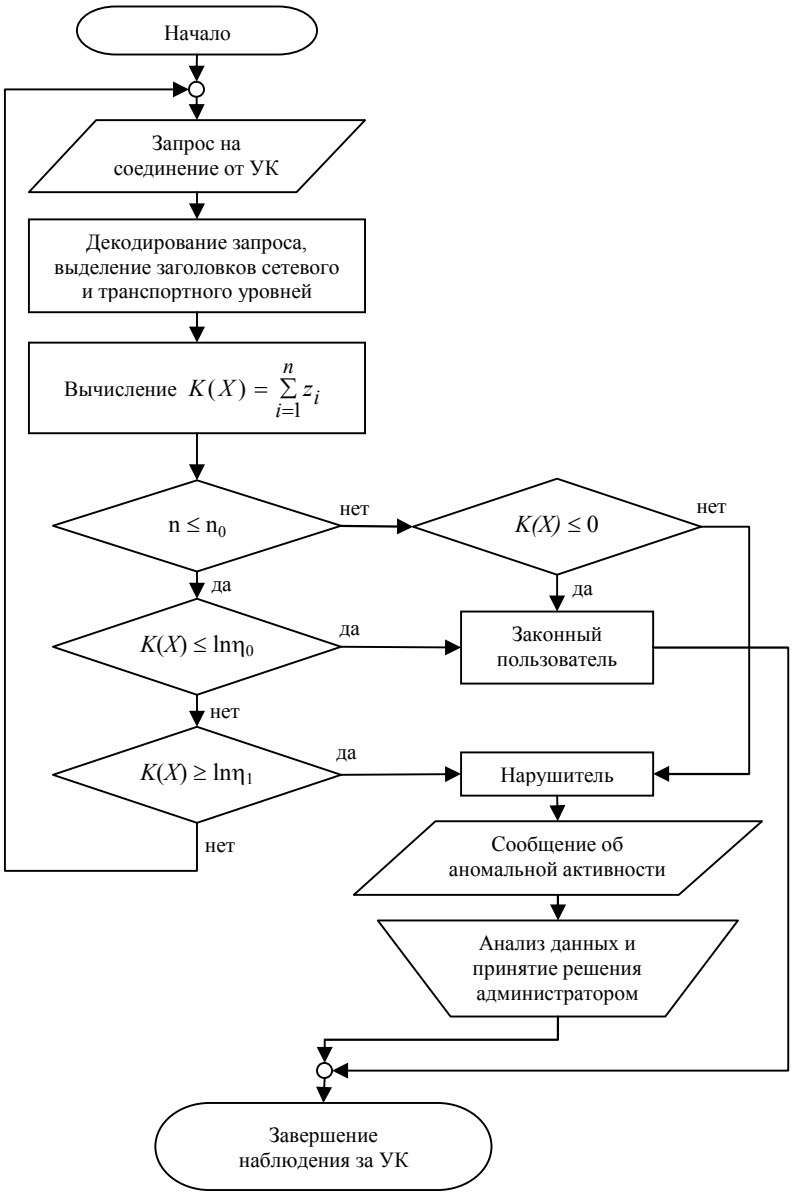

Рис. 2. Блок-схема алгоритма обнаружения аномальной сетевой активности

4. Испытания МОАА

При проведении испытаний предложенного подхода использовалась корпоративная сеть Республиканского центра трансфера технологий (РЦТТ), имеющая следующую конфигурацию:

– почтовый сервер (ОС Windows Server 2003, ЦП Intel PIII 1233 МГц, ОЗУ 512 Мб);

– файловый сервер (ОС Windows Server 2003, ЦП Intel PIII 1233 МГц, ОЗУ 512 Мб);

– веб-сервер (ОС Windows Server 2003, ЦП Intel PIV 3200 МГц, ОЗУ 2048 Мб);

– пять рабочих станций (ОС Windows XP SP2, ЦП Intel PIV 2233 МГц, ОЗУ 256 Мб);

– симулятор атак (ОС Debian Sarge Stable, ЦП Intel PIV 2233 МГц, ОЗУ 256 Мб);

– система обнаружения атак (ОС FreeBSD 4.10, ЦП AMD Sempron 1833 МГц, ОЗУ 1024 Мб).

Все десять систем объединены в сеть стандарта Ethernet 100 Мбит/с с помощью коммутаторов 3Com OfficeConnect. Указанные выше три серверные системы, работающие под управлением Windows Server 2003, представляют собой публичные серверы и круглосуточно доступны из сети Интернет. Рабочие станции использовались как генераторы неаномального трафика в сети. Подобный трафик генерировался как вручную пользователями РЦТТ, так и с помощью

средств автоматизированной загрузки веб-страниц и файлов, а также периодической отправкой почтовых сообщений по проверенным адресам.

Важную роль в испытаниях играл симулятор атак, использующийся для генерации запросов на соединение со службами сети и моделирования воздействий программных сетевых червей [11].

Система обнаружения атак использовала не только описанный метод обнаружения аномальной активности, но и два других общедоступных метода: sfPortscan из программного комплекса обнаружения сетевых атак Snort версии 2.4.0, а также подход, реализованный в утилите Scanlogd версии 2.2.5 [2]. Суть метода sfPortscan состоит в постоянном отслеживании успешных соединений и сообщений о неудачных попытках подключения в течение динамически подстраиваемого «временного окна». Размер «окна» меняется в зависимости от параметров трафика в сети (главным образом в зависимости от энтропии адресов и номеров портов сетевых пакетов: чем больше энтропия, тем больше «окно»). Алгоритм работы утилиты Scanlogd значительно проще и заключается в реагировании на обращение к определенному количеству портов в течение заданного промежутка времени. Фактически анализируется взвешенное значение таких обращений. Особенность утилиты состоит в невозможности тонкой настройки, так как все параметры ее функционирования заданы жестко.

Испытания проводились 10 дней, в течение которых в экспериментальной сети было зарегистрировано всего 29 790 попыток соединения с 1379 уникальных адресов. Проведенный в последствии анализ трафика показал (табл. 1), что не менее чем с 87 адресов осуществлялись атаки рекогносцировки, не менее чем с 34 – вирусные атаки программных червей. Оставшиеся 1258 адресов были признаны принадлежащими лояльным клиентам центра.

Таблица 1

Результаты испытаний

Вид активности	Количество УК	Принятое решение	МОАА	Snort v. 2.4.0 sfPortscan	Scanlogd v. 2.2.5
Рекогносцировка	87	Обнаружение	85	67	53
Сетевой червь	34	Обнаружение	33	31	23
Легальная	1258	Ложное срабатывание	5	43	93

Во время испытаний параметрам реализованного в МОАА алгоритма экспертным образом были назначены следующие значения: $a = 0,005$; $b = 0,97$; $\theta_0 = 0,85$; $\theta_1 = 0,2$; $n_0 = 50$. Очевидно, что предложенный метод по способности к обнаружению ($P_O \approx 85/87 \approx 0,98$ для атак рекогносцировки и $P_O \approx 33/34 \approx 0,97$ для обнаружения активности сетевых червей) и малому числу ложных срабатываний ($P_{ЛО} \approx 5/1258 \approx 0,004$) значительно превосходит методы, реализованные в системах Snort и Scanlogd. Возможно, тонкая настройка метода sfPortscan позволила бы несколько улучшить результаты его применения, но для общности оценки было решено проводить испытания с настройками по умолчанию.

Не менее важным результатом явилось то, что для принятия одной из гипотез в среднем требовалось всего около семи наблюдений. Также необходимо отметить, что усечение процедуры последовательного анализа произошло всего два раза за все время испытаний, причем оба эти случая были вызваны симулятором атак, работающим так, чтобы значение $K(X)$ случайно блуждало около нуля [9]. Таким образом, усечение процедуры последовательного анализа позволило предотвратить реализацию атаки отказа в обслуживании на систему обнаружения атак.

Заключение

Представлены новые метод и алгоритм обнаружения аномальной сетевой активности в реальном масштабе времени, основанные на статистическом последовательном анализе с возможностью ограничения при необходимости количества наблюдений. Предложенный алгоритм базируется на разработанной математической модели, в основу которой положен эмпирически зарегистрированный факт – относительная частота успешных попыток подключения к сетевым службам законных пользователей значительно превышает относительную частоту успешных

попыток подключения злоумышленников. Алгоритм реализован программно в качестве компонента комплекса обнаружения сетевых атак и испытан в реальных условиях. Проведенные испытания показали, что данный метод во многом превосходит существующие методы обнаружения аномальной активности. Основные особенности разработанного алгоритма заключаются в том, что он позволяет достаточно быстро и одновременно достоверно (на практике за пять-шесть попыток соединения) обнаруживать атаки рекогносцировки и аномальную сетевую активность программных червей, а также в его концептуальной простоте и непритязательности к вычислительным ресурсам.

Список литературы

1. Лебедь, С.В. Межсетевое экранирование / С.В. Лебедь. – М.: МГТУ, 2002. – 304 с.

2. Лукацкий, А.В. Обнаружение атак / А.В. Лукацкий. – СПб.: БХВ-Петербург, 2003. – 608 с.

3. Баранов, П.А. Описание проектирования экспертного анализатора обнаружения сетевых атак / П.А. Баранов // Мат. XII Общероссийской науч.-техн. конф. «Методы и технические средства обеспечения безопасности информации». – Россия, 2004. – С. 85.

4. Leckie, C. A probabilistic approach to detecting network scans / C. Leckie, R. Kotagiri // Proc. of the Eighth IEEE Network Operations and Management Symposium. – Italy, 2002. – P. 359–372.

5. Yegneswaran, V. Internet intrusions: global characteristics and prevalence / V. Yegneswaran, P. Barford, J. Ullrich // Proc. of the 2003 ACM SIGMETRICS. – USA, 2003. – P. 138–147.

6. Jung, J. Fast portscan detection using sequential hypothesis testing / J. Jung, V. Paxson, A. Berger // Proc. of the IEEE Symposium on Security and Privacy. – USA, 2004. – P. 211–225.

7. Блекуэлл, Д. Теория игр и статистических решений / Д. Блекуэлл, М. Гиршик. – М.: ИЛ, 1958. – 376 с.

8. Вальд, А. Последовательный анализ / А. Вальд. – М.: Физматгиз, 1960. – 328 с.

9. Ширяев, А.Н. Статистический последовательный анализ / А.Н. Ширяев. – М.: Наука, 1976. – 272 с.

10. Kharin, A. On Robustifying of the Sequential Probability Ratio Test for a Discrete Model under «Contaminations» / A. Kharin // Austrian Journal of Statistics. – 2002. – Vol. 31. – № 4. – P. 267–277.

11. Анищенко, В.В. Система моделирования удаленных атак на компьютерные сети / В.В. Анищенко, Ю.В. Земцов // Мат. II Белорусско-российской науч.-техн. конф. «Технические средства защиты информации». – Беларусь, 2004. – С. 13–14.

Поступила 22.02.06

Объединенный институт проблем
информатики НАН Беларуси,
Минск, Сурганова, 6
e-mail: xyzemts@mail.ru

Y.U. Ziamtsou

ANOMALY DETECTION ON BASIS OF TRUNCATED PROCEDURE OF SEQUENTIAL ANALYSIS

This paper describes a new real-time anomaly detection technique and developing the corresponding algorithm. Proposed technique is based on theoretically grounded truncated procedure of statistical sequential analysis. Test work was conducted to reveal that the developed algorithm performs faster and also more accurately than other current solutions.