МЕХАНИЗМЫ И СРЕДСТВА ЗАЩИТЫ СЕТЕЙ СВЯЗИ

УДК 004.056.57

А.А. КУЗНЕЦОВ, А.А. СМИРНОВ, Д.А. ДАНИЛЕНКО, А. БЕРЕЗОВСКИЙ

СТАТИСТИЧЕСКИЙ АНАЛИЗ СЕТЕВОГО ТРАФИКА ДЛЯ СИСТЕМ ОБНАРУЖЕНИЯ И ПРЕДОТВРАЩЕНИЯ ВТОРЖЕНИЙ

Введение

Для обеспечения безопасности современных телекоммуникационных сетей применяются так называемые системы обнаружения (IntrusionDetectionSystem – IDS) и предотвращения (IntrusionPreventionSystem – IPS) вторжений [1 – 14]. В основе их функционирования лежат сбор, анализ и обработка информации о событиях, связанных с безопасностью защищаемой телекоммуникационной сети, накопление полученных данных, мониторинг сетевой активности отдельных служб и сервисов, принятие решения о состоянии защищаемой системы с выявлением и возможным противодействием несанкционированному использованию инфокоммуникационных ресурсов [2].

Большинство современных систем обнаружений и предотвращения вторжений (СОПВ) применяет сигнатурный метод [1 – 14], в основе которого лежит использование словаря характерных признаков (сигнатур), описывающих различные атаки, вторжения или вирусы. Сигнатуры могут иметь различную форму и определять конкретные параметры от номера порта в пакете до последовательности байт в серии пакетов [1]. При этом эффективность СОПВ при известной сигнатуре вторжения очень высока, однако для обнаружения новых, еще неизвестных типов вторжений (вирусов, атак и пр.) требуется их подробный анализ и детальное исследование с разработкой новых сигнатур, что и является основным недостатком такого подхода. Другими словами, эффективность СОПВ с сигнатурным методом обнаружения непосредственно зависит от изученности атаки и наличия соответствующей сигнатуры, система защиты должна постоянно обновляться по мере появления новых вирусов, атак и пр. [1 – 14].

Другим, не менее важным, направлением в совершенствовании СОПВ является исследование аномалий (Anomaly-BasedIntrusionDetectionandPreventionSystems) телекоммуникационных систем, в основу которого может быть положен статистический анализ сетевого трафика [2]. При таком подходе СОПВ определяет «нормальную» сетевую активность отдельных служб и информационных сервисов телекоммуникационной системы, после чего весь трафик, не подпадающий под определение «нормального», помечается как «аномальный».

Функционирование статистических методов в СОПВ организовано таким образом:

1. Проводится первичный мониторинг сетевой активности телекоммуникационной системы. На основе результатов наблюдений за активностью сетевых служб и информационных сервисов в течение некоторого периода времени оцениваются показатели, характеризующие штатное функционирование телекоммуникационной системы, при этом выявляется «нормальный» сетевой трафик;

2. Вырабатываются статистические правила (критерии), по которым принимается решение о переходе телекоммуникационной системы в неустановленный (нештатный) режим функционирования. Правило принятия решений может базироваться на статистической проверке гипотезы об однородности наблюдаемого и выявленного ранее «нормального» сетевых

трафиков. Разница между «нормальным» и «аномальным» событием определяется пороговой величиной (в случае проверки гипотез – критической областью);

3. Анализируется сетевой трафик с поиском аномалии в установившейся картине «нормального» сетевого трафика. Всем пакетам дается оценка «аномальности» (включающая в себя степень отклонения отдельных показателей для специфического события), и если эта оценка выше определенного предела (лежит в критической области), СОПВ генерирует сигнал тревоги либо блокирует соответствующий процесс.

Главным преимуществом статистических методов в СОПВ является возможность изучать (мониторить) сетевой трафик и отличать «нормальную» сетевую активность от «аномальной». Кроме того, существует возможность самообучения, самонастраивания, т.е. первичный мониторинг сетевой активности телекоммуникационной системы может осуществляться периодически (при отсутствии вторжений) с корректировкой пороговых величин и самих критериев принятия решений о переходе системы в неустановленный («аномальный») режим функционирования. Все это в совокупности делает статистическую СОПВ гораздо гибче сигнатурной, дает ей возможность без известных сигнатур вторжений обнаруживать и предотвращать новые, еще неизвестные атаки и сетевые вирусы.

Таким образом, построение СОПВ со статистическими методами обнаружения вторжений предполагает разработку надежных инструментов мониторинга сетевой активности, позволяющих с высокой точностью и достоверностью детектировать «нормальный» сетевой трафик и выявлять «аномальные» события.

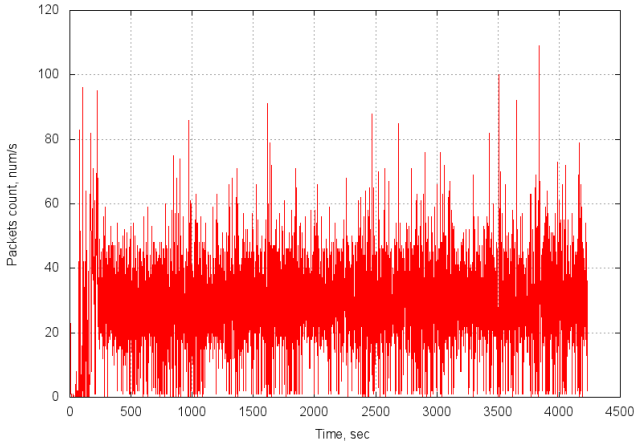
В данной работе проводится статистический анализ сетевого трафика различных служб и информационных сервисов современных телекоммуникационных систем и сетей, обосновываются предложения по мониторингу сетевой активности для обнаружения вторжений и предотвращения их воздействия на защищаемые инфокоммуникационные ресурсы. При проведении статистических исследований использован критерий принадлежности двух выборок одной и той же генеральной совокупности (критерий Вилькоксона) и статистический критерий на основе отношения выборочных дисперсий (критерий Фишера).

Исходные данные проводимых экспериментальных исследований

Для проведения экспериментальных исследований свойств сетевого трафика использованы эмпирические данные, полученные в результате работы программного анализатора (снифера) «Wireshark». Оценивался объем данных, передаваемых через компьютерную сеть за определенный период времени. Измерения трафика, т.е. объема информации, передаваемого в единицу времени, проводились как по числу пакетов, так и по числу бит данных. Эмпирические данные получены и обобщены не менее чем по 100000 временным отсчетам.

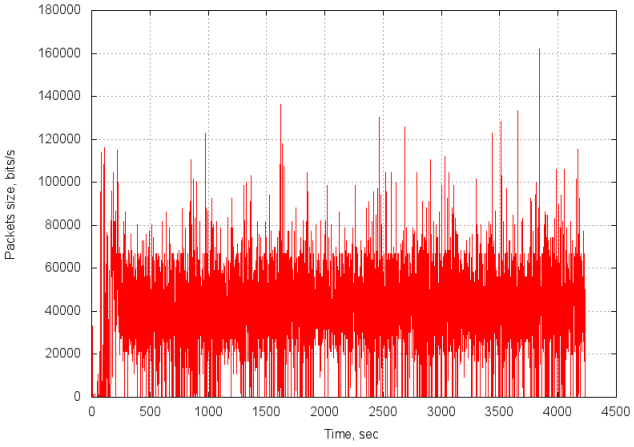
В качестве исходных данных использованы различные телекоммуникационные службы и информационные сервисы, а именно: FTP (FileTransferProtocol) – стандартный протокол, предназначенный для передачи файлов по TCP-сетям; HTTP (HyperTextTransferProtocol) – протокол прикладного уровня передачи данных; электронная почта (e-mail) – технология и предоставляемые ею услуги по пересылке и получению электронных сообщений по распределенной (в том числе глобальной) компьютерной сети; Skype – бесплатное программное обеспечение, обеспечивающее текстовую, голосовую связь и видеосвязь через Интернет; YouTube – сервис, предоставляющий услуги видеохостинга, т.е. доступа к сайтам, позволяющим загружать и просматривать видео. Примеры полученных гистограмм приведены на рис. 1 – 6 (*а* – представление трафика в виде числа переданных пакетов в единицу времени – «*пакет/с*», *б* – представление в виде «*бит/с*»). При проведении дальнейших статистических исследований использованы эмпирические данные по 100 временным отсчетам случайно

выбранных отрезков сетевого трафика, соответствующих различным телекоммуникационным службам и информационным сервисам.



пакет/с

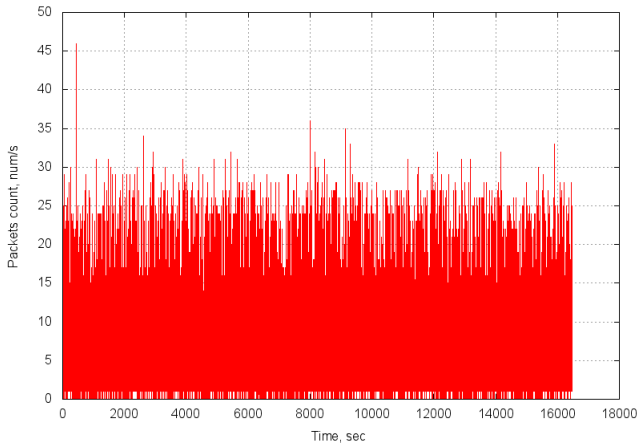
а



бит/с

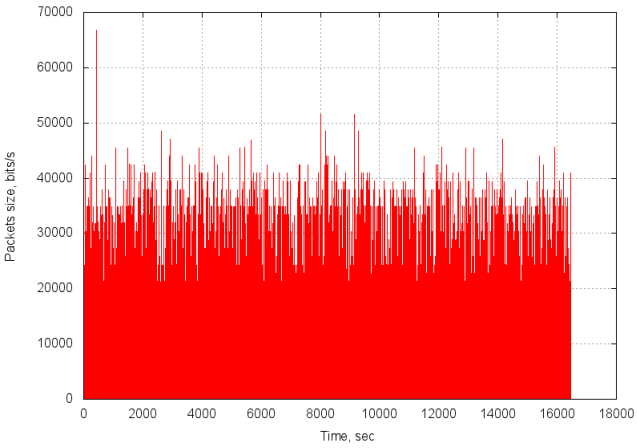
б

Рис. 1. Фрагмент гистограммы сетевого трафика при загрузке данных с сервиса YouTube – 720p



пакет/с

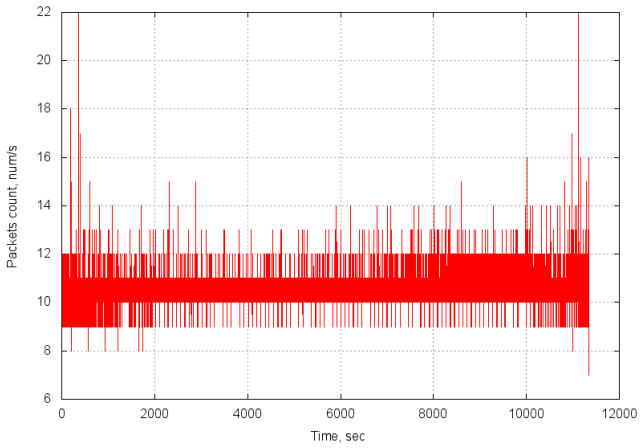
а



бит/с

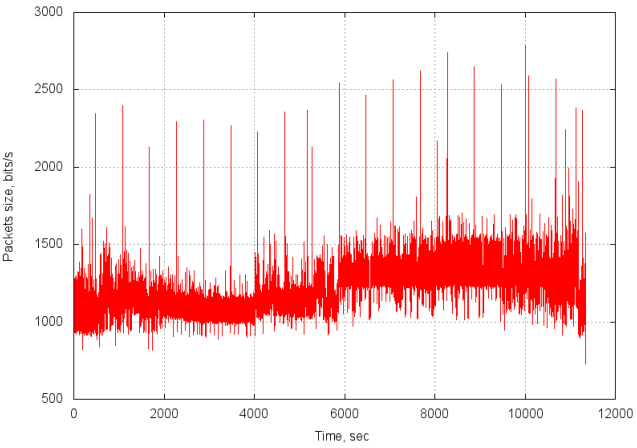
б

Рис. 2. Фрагмент гистограммы сетевого трафика при загрузке данных с сервиса YouTube – 360p



пакет/с

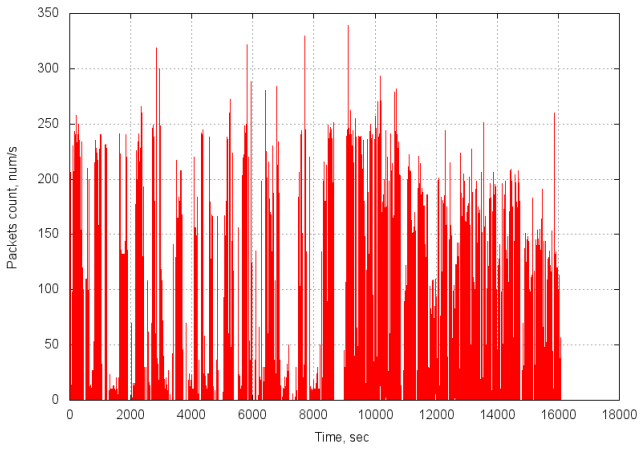
а



бит/с

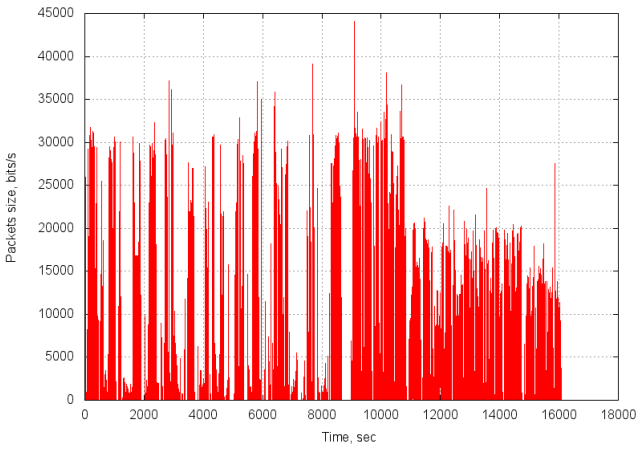
б

Рис. 3. Фрагмент гистограммы сетевого трафика при обмене данными с использованием Skype (voice)



пакет/с

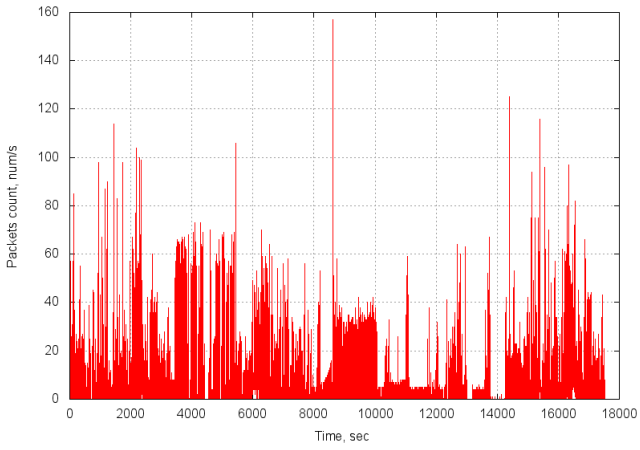
а



бит/с

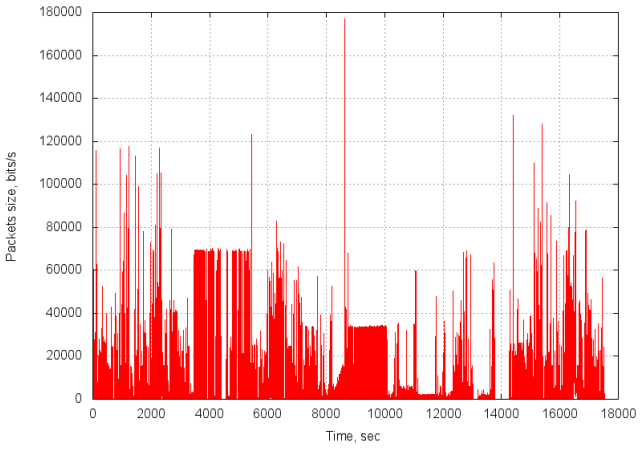
б

Рис. 4. Фрагмент гистограммы сетевого трафика при передаче электронной почты



пакет/с

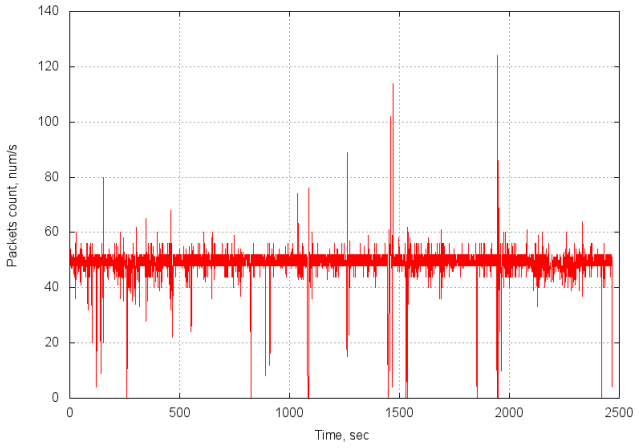
а



бит/с

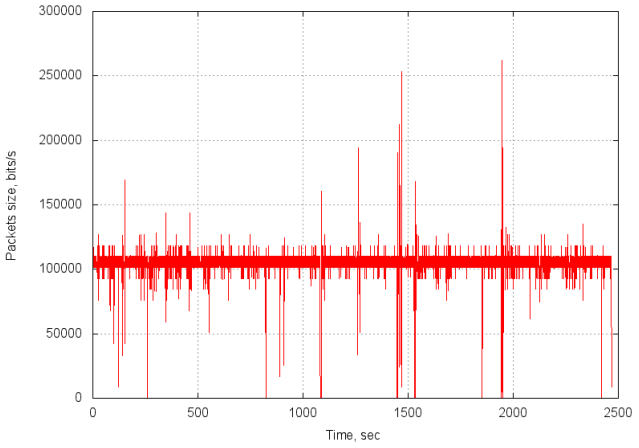
б

Рис. 5. Фрагмент гистограммы сетевого трафика при передаче данных с использованием протоколов HTTP



пакет/с

а



бит/с

б

Рис. 6. Фрагмент гистограммы сетевого трафика при передаче данных с использованием протоколов FTP

В соответствии с основными положениями центральной предельной теоремы теории вероятностей сумма достаточно большого количества слабо зависимых случайных величин, имеющих примерно одинаковые масштабы (ни одно из слагаемых не доминирует, не вносит в сумму определяющего вклада), имеет распределение, близкое к нормальному [15, 16]. Так как объем данных, передаваемых через компьютерную сеть за определенный период времени, является случайной величиной, формируемой под влиянием большого числа слабо зависимых случайных факторов, будем считать распределение этой случайной величины нормальным. При этом должно соблюдаться условие, что ни один из факторов не является доминирующим при формировании сетевого трафика[1].

Исследование однородности сетевого трафика по критерию Вилькоксона

При исследовании сложных технических систем часто возникает задача оценки значимости расхождений двух и более выборок (серий) независимых наблюдений, т.е. требуется установить (с заданной точностью и достоверностью) их принадлежность одной и той же генеральной совокупности.

Пусть имеются две выборки

$$x_1, x_2, \ldots, x_{N1} \text{ и } y_1, y_2, \ldots, y_{N2} \quad (1)$$

случайных величин X и Y, имеющих распределения $P_X(t)$ и $P_Y(t)$ соответственно.

Предположим, что наблюдаемые x_iи y_i дают различные значения выборочных средних

$$x^* = (x_1 + x_2 + \ldots + x_{N_1})/N_1,\ y^* = (y_1 + y_2 + \ldots + y_{N_2})/N_2,\ x^* \neq y^* \quad (2)$$

и/или выборочных рассеиваний (дисперсий)

$$\sigma_X^2 = \frac{1}{N_1}\sum_{i=1}^{N_1}(x_i - x^*)^2,\ \sigma_Y^2 = \frac{1}{N_2}\sum_{i=1}^{N_2}(y_i - y^*)^2,\ \sigma_X^2 \neq \sigma_Y^2. \quad (3)$$

Решение задачи оценки значимости расхождений наблюдаемых значений x_iи y_i сводится к проверке нулевой гипотезы H_0, состоящей в том, что функции распределения $P_X(t)$ и $P_Y(t)$ тождественны для любых t. Альтернативная гипотеза формулируется в виде неравенства $P_X(t)<P_Y(t)$. Критерий принадлежности двух выборок одной и той же генеральной совокупности (критерий Вилькоксона) основан на подсчете *числа инверсий* [15]. Для этого наблюдаемые значения x_iи y_i располагаются в общую последовательность в порядке возрастания их значений. Полученная неубывающая последовательность содержит $N_1 + N_2$ значений и если гипотеза $P_X(t) = P_Y(t)$ верна, то, очевидно, значения из обеих последовательностей $x_1, x_2, \ldots, x_{N1}$ и $y_1, y_2, \ldots, y_{N2}$ хорошо перемешиваются. Степень перемешивания определяется числом инверсий членов первой последовательности относительно второй. Если в упорядоченной общей последовательности некоторому x предшествует одно значение y, это означает, что имеет место одна инверсия. Если некоторому x предшествуют k значений y, то это значение x имеет k инверсий [15].

Обозначим символом u_i число инверсий для значения x_i относительно предшествующих ему значений y. Тогда общее число инверсий (для всех значений из последовательности $x_1, x_2, \ldots, x_{N1}$ относительно значений из последовательности $y_1, y_2, \ldots, y_{N2}$) будет определяться суммой

$$u = u_1 + u_2 + \ldots + u_{N_1}.$$

[1]Это предположение, в определенных случаях, может быть ошибочным, т.к. для некоторых служб и информационных сервисов телекоммуникационной сети существуют отдельные факторы, являющиеся доминирующими при формировании сетевого трафика, т.е. они вносят основной, определяющий вклад в формирование объемов данных, передаваемых в единицу времени

Нулевая гипотеза H_0 отвергается, если число u превосходит выбранную в соответствии с уровнем значимости границу, определяемую из того расчета, что при объемах выборок $N_1 > 10$ и $N_2 > 10$ число инверсий u распределено приблизительно нормально [15] с центром и дисперсией, соответственно:

$$M_u = \frac{N_1 N_2}{2},\ D_u = \frac{N_1 N_2}{12}(N_1 + N_2 + 1). \quad (4)$$

При уровне значимости q и нормальности распределения числа инверсий вероятность непопадания значения u в критическую область (доверительная вероятность) [15]:

$$P_\partial = P(|M_u - u| \le \varepsilon) = 1 - q = 2\Phi\left(\frac{\varepsilon}{\sigma_u}\right), \quad (5)$$

где ε задает величину максимального отклонения полученной оценки от истинного значения.

Зафиксируем значение доверительной вероятности P_∂, значения левой и правой критической границы будут соответственно

$$u_1 = M_u - t_\alpha \sigma_u, u_2 = M_u + t_\alpha \sigma_u, \quad (6)$$

где $\sigma_u = \sqrt{D_u}$ – среднеквадратическое отклонение числа инверсий, t_α – корень уравнения $2\Phi(t_\alpha) = P_\partial$, $t_\alpha = \Phi^{-1}\left(\frac{1-q}{2}\right)$.

Проведем экспериментальные исследования принадлежности двух выборок сетевого трафика одной и той же генеральной совокупности. Для этого сформируем выборки (1) по 100 временным отсчетам случайно выбранных отрезков сетевого трафика, соответствующих различным телекоммуникационным службам и информационным сервисам (YouTube (720p), YouTube (360p), Skype (voice), Skype (video), E-mail, HTTP, FTP, см. рис. 1 – 6), т.е. $N_1 = N_2 = 100$. Оценим выборочные средние (2) и дисперсии (3), а также математическое ожидание и дисперсию (4) числа инверсий:

$$M_u = \frac{N_1 N_2}{2} = 5000,\ D_u = \frac{N_1 N_2}{12}(N_1 + N_2 + 1) = 167500,\ \sigma_u \approx 409{,}3.$$

Зададимся уровнем значимости q = 10%. Используя (5) по (6), рассчитаем значения левой и правой критической границы:

$$P_\partial = 1 - q = 0{,}9,\ t_\alpha = \Phi^{-1}(0{,}45) \approx 1{,}65, u_1 = M_u - t_\alpha \sigma_u \approx 4324{,}7\ ,\ u_2 = M_u + t_\alpha \sigma_u \approx 5675{,}3\ .$$

Итоговые результаты проведенных экспериментальных исследований сведены в табл. 1, 2, в которых приведены значения числа инверсий и результаты проверки гипотезы об однородности сетевых трафиков для различной формы представления («*пакет/с*» и «*бит/с*» соответственно). В таблицах здесь и далее применяются следующие обозначения:

«+» – гипотеза об однородности сетевых трафиков не отвергается,

«–» – гипотеза об однородности сетевых трафиков отвергается.

Анализ данных табл. 1 и 2 показывает, что при исследовании выборок сетевых трафиков различных телекоммуникационных служб и сервисов в большинстве случаев гипотеза об их однородности отвергается, т.е наблюдается правильное принятие решения о принадлежности выборок различным процессам. Это положение может быть положено в основу работы одного из элементов системы мониторинга сетевой активности, т.е. на основе использования критерия Вилькоксона может производиться первичное детектирование телекоммуникационной службы или сервиса. Как видно из приведенных данных, результаты проверки гипотез являются симметричными относительно главной диагонали, что подтверждает достоверность полученных результатов в каждом конкретном эксперименте.

Таблица 1

Число инверсий и результаты проверки гипотезы
об однородности сетевых трафиков (пакет/с)

	YouTube (720p)	YouTube (360p)	Skype (voice)	Skype (video)	E-mail	HTTP	FTP
YouTube (720p)	5005 «+»	9555 «–»	9795 «–»	7932 «–»	947 «–»	8314 «–»	559 «–»
YouTube (360p)	542 «–»	5006 «+»	2764 «–»	604 «–»	9016 «–»	4734 «+»	29 «–»
Skype (voice)	302 «–»	7364 «–»	5003 «+»	0 «–»	9774 «–»	6297 «–»	0 «–»
Skype (video)	2140 «–»	9502 «–»	10000 «–»	5009 «+»	999 «–»	7327 «–»	195 «–»
E-mail	54 «–»	1066 «–»	333 «–»	3 «–»	5001 «+»	1990 «–»	2 «–»
HTTP	1732 «–»	5333 «+»	3742 «–»	2699 «–»	8095 «–»	5003 «+»	204 «–»
FTP	9539 «–»	989 «–»	10000 «–»	9901 «–»	10000 «–»	9893 «–»	5007 «+»

Таблица 2

Число инверсий и результаты проверки гипотезы
об однородности сетевых трафиков (*бит/с*)

	YouTube (720p)	YouTube (360p)	Skype (voice)	Skype (video)	E-mail	HTTP	FTP
YouTube (720p)	5006 «+»	9525 «–»	9797 «–»	9733 «–»	961 «–»	9056 «–»	137 «–»
YouTube (360p)	572 «–»	5008 «+»	7171 «–»	2586 «–»	9483 «–»	5288 «+»	7 «–»
Skype (voice)	300 «–»	2900 «–»	5008 «+»	0 «–»	9890 «–»	4323 «–»	0 «–»
Skype (video)	364 «–»	7542 «–»	10000 «–»	5009 «+»	10000 «–»	6923 «–»	0 «–»
E-mail	47 «–»	590 «–»	212 «–»	0 «–»	5001 «+»	1770 «–»	0 «–»
HTTP	1011 «–»	4766 «+»	5684 «–»	3108 «–»	8320 «–»	5009 «+»	4 «–»
FTP	9960 «–»	991 «–»	10000 «–»	10000 «–»	10000 «–»	998 «–»	5007 «+»

Исследование однородности сетевого трафика по критерию Фишера

Суть дисперсионного анализа состоит в проверке гипотезы о тождественности выборочных дисперсий одной и той же генеральной совокупности [15, 16]. Пусть имеются две выборки $x_1, x_2, ..., x_{N1}$ и $y_1\ y_2, ..., y_{N2}$ случайных величин X и Y, имеющих нормальное распределение. Дисперсия случайной величины, являясь суммой квадратов ошибок, имеет распределение χ^2 (распределение Пирсона). Задача сравнения дисперсий случайных величин X и Y сводится к проверке исходной гипотезы (нулевой гипотезы H_0) о принадлежности двух выборок одной и той же генеральной совокупности [15, 16]. Для проверки гипотезы о равенстве

дисперсий используют независимую функцию, вычислимую по данным эксперимента. Такой функцией является *функция Фишера* (распределение Фишера, *F*-распределение), ее значение определяется как [15]: $F = \frac{U/k_1}{V/k_2}$, где U и V – случайные величины, имеющие распределение χ^2 ;k_1 и k_2 соответствующие степени свободы случайных величин U и V соответственно, $k_1 = N_1 - 1$, $k_2 = N_2 - 1$;N_1 и N_2 – количество испытаний (объемы выборок).

Таким образом, случайная величина $F = \sigma_1^2 / \sigma_2^2$ имеет распределение Фишера (*F*-распределение), где: σ_1^2 и σ_2^2 – несмещенные оценки дисперсий, а x^* и y^* – несмещенные оценки математических ожиданий. Граничные точки допустимых значений F определяются точками F_1 и F_2, соответствующих вероятностям $q/2$. Если вычисленное по данным эксперимента значение F попадает в область между граничными точками F_1 и F_2, т.е. не попадает в так называемую критическую область, принятая гипотеза не опровергается.

Воспользуемся аппаратом дисперсионного анализа для проверки статистической гипотезы об однородности сетевых трафиков рассматриваемых служб и информационных сервисов телекоммуникационной системы. Для этого выполним следующие этапы статистической проверки гипотез.

1. Сформулируем основную гипотезу H_0: сетевые трафики однотипны по характеристике рассеивания, т.е. их выборочные дисперсии тождественны одной и той же генеральной дисперсии. Сформулируем также конкурирующую гипотезу H_1: сетевые трафики не однотипны по характеристике рассеивания, т.е. их выборочные дисперсии не тождественны одной и той же генеральной дисперсии.

2. Зададим уровень значимости q= 10%, на котором в дальнейшем и будет сделан вывод о справедливости гипотезы. Численно он равен вероятности допустить ошибку первого рода (вероятности ложной тревоги), т.е. вероятности отклонить гипотезу H_0, когда на самом деле она верна.

3. Произведем расчет статистики теста так, чтобы: ее величина зависела от исходной выборки; по ее значению можно было бы сделать вывод об истинности гипотезы H_0; полученная статистика подчинялась бы известному и рассмотренному выше закону распределения Фишера.

4. Построим критическую область, т.е. зададим граничные точки F_1 и F_2 допустимых значений F и из области значений статистики теста выделим подмножество значений (критическую область) $F < F_1$и $F > F_2$, по которым будем судить о существенных расхождениях с предположением. Размер этой области определим из условия выполнения равенства $P(F < F_1 \vee F > F_2) = q = 0{,}1$.

5. Сделаем вывод об истинности гипотезы H_0. Для этого по наблюдаемым значениям выборки рассчитаем статистику теста и по попаданию (или непопаданию) в критическую область ($F < F_1 \vee F > F_2$) вынесем решение об отвержении (или неотвержении) выдвинутой гипотезы H_0.

Расчет статистики теста (этап 3) основывается на подсчете отношения выборочных дисперсий (сумм квадратов, деленных на «степени свободы»), эта статистика имеет распределение Фишера. Построим это распределение для заданных степеней свободы $k_1 = k_2 = N_2 - 1 = N_1 - 1 = 99$. Графики плотности вероятностей $f_x(x)$ распределения Фишера и соответствующего интегрального распределения вероятностей $F_x(x)$ для значений $k_1 = k_2 = 99$ приведены на рис. 7 – 8[2].

[2]При построении графиков функций использован пакет Math Cad15

Используя уровень значимости $q = 0.1$ (рис. 7, 8), найдем значения левой и правой граничных точек: $F_1 = 0{,}717, F_2 = 1{,}394$, вероятность попадания значения F в критическую область соответственно

$$P(F < F_1 = 0{,}717 \vee F > F_2 = 1{,}394) = 0{,}1.$$

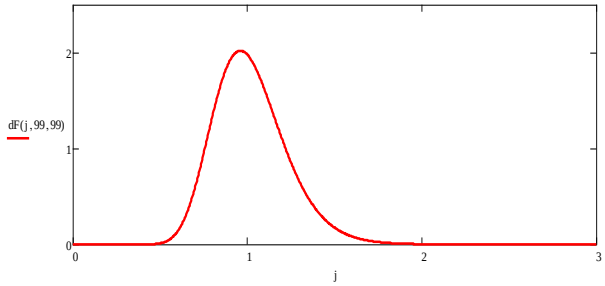


Рис. 7. График плотности вероятности распределения Фишера для числа степеней свободы $k_1 = k_2 = 99$

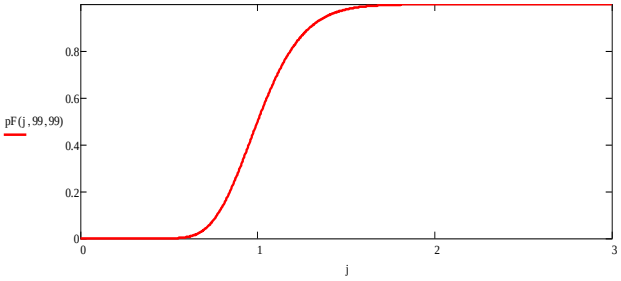


Рис. 8. График интегрального распределения вероятностей Фишера для числа степеней свободы $k_1 = k_2 = 99$

Если значение рассчитанной на третьем этапе статистики попадает в критическую область, т.е. лежит ниже левой или выше правой граничной точки, тогда гипотеза H_0 об однотипности исследуемых сетевых трафиков по характеристике их рассеивания отвергается, т.е. их выборочные дисперсии не тождественны одной и той же генеральной дисперсии. Если это значение не попадает в критическую область, т.е. лежит выше левой и ниже правой граничной точки, тогда гипотеза H_0не отвергается, т.е. полагаем, что исследуемые сетевые трафики однотипны, их выборочные дисперсии тождественны одной и той же генеральной дисперсии. Полученные результаты оценки выборочных данных для сетевых трафиков различных служб и сервисов приведены в табл. 3.

Таблица 3

Оценки выборочных данных для сетевых трафиков различных служб и сервисов

Вид трафика (служба,сервис)	Оценка выборочной дисперсии	Оценка выборочного среднего
YouTube (720p)	$2{,}4\times10^8$	41372,8
Skype (voice)	14560,9	1154,9
Skype (video)	$7{,}6\times10^6$	14738,7
E-mail	116079	122,5
HTTP	$2{,}2\times10^8$	11567,8
FTP	$2{,}4\times10^8$	104970

Итоговые результаты дисперсионного анализа сведены в табл. 4, 5. В них приведены значения статистики теста и соответствующие решения по проверке гипотезы об однородности различных трафиков по отношению выборочных дисперсий. Данные табл. 4 соответствуют представлению трафика в форме «*пакет/с*», в табл. 5 – «*бит/с*».

Рассмотрим свойство «симметричности» F – распределения, т.е. свойство, состоящее в том, что случайная величина $F^* = \dfrac{1}{F} = \dfrac{\sigma_2^2}{\sigma_1^2}$ также имеет F – распределение.

Тогда полученные результаты проверки статистической гипотезы об однородности сетевого трафика не должны зависеть от выбора формы отношения выборочных дисперсий. Практически это означает симметричность относительно главной диагонали табл. 4 и 5 по результатам проверки статистических гипотез.

Анализ данных табл. 4 и 5 свидетельствует об их симметричности относительно главной диагонали по результатов статистической проверки гипотезы, что подтверждает достоверность полученных экспериментальных данных.

Таблица 4

Значение статистики теста и результаты проверки гипотезы
об однородности сетевых трафиков (*пакет/с*) по отношению выборочных дисперсий

	YouTube(720p)	YouTube(360p)	Skype (voice)	Skype (video)	E-mail	HTTP	FTP
YouTube(720p)	1 «+»	0,45 «–»	$3{,}60\times10^{-3}$ «–»	0,05 «–»	0,11 «–»	1,55 «–»	0,51 «–»
YouTube(360p)	2,25 «–»	1 «+»	$8{,}10\times10^{-3}$ «–»	0,11 «–»	0,25 «–»	3,49 «–»	1,15 «+»
Skype (voice)	277,50 «–»	123,5 «–»	1 «+»	13,75 «–»	31 «–»	431,5 «–»	141,75 «–»
Skype (video)	20,18 «–»	8,98 «–»	$7{,}27\times10^{-2}$ «–»	1 «+»	2,25 «–»	31,38 «–»	10,31 «–»
E-mail	8,95 «–»	3,98 «–»	$3{,}23\times10^{-2}$ «–»	0,44 «–»	1 «+»	13,92 «–»	4,57 «–»
HTTP	0,64 «–»	0,29 «–»	$2{,}32\times10^{-3}$ «–»	0,03 «–»	$7{,}18\times10^{-2}$ «–»	1 «+»	0,33 «–»
FTP	1,96 «–»	0,87 «+»	$7{,}05\times10^{-3}$ «–»	0,10 «–»	0,22 «–»	3,04 «–»	1 «+»

Таблица 5

Значение статистики теста и результаты проверки гипотезы
об однородности сетевых трафиков (*бит/с*) по отношению выборочных дисперсий

	YouTube(720p)	YouTube(360p)	Skype (voice)	Skype (video)	E-mail	HTTP	FTP
YouTube(720p)	1 «+»	0,5 «–»	$6{,}07\times10^{-5}$ «–»	$3{,}17\times10^{-2}$ «–»	$4{,}84\times10^{-4}$ «–»	0,92 «+»	1,0 «+»
YouTube(360p)	2,0 «–»	1 «+»	$1{,}21\times10^{-4}$ «–»	$6{,}33\times10^{-2}$ «–»	$9{,}67\times10^{-4}$ «–»	1,83 «–»	2,0 «–»
Skype (voice)	$1{,}65\times10^{4}$ «–»	$8{,}24\times10^{3}$ «–»	1 «+»	521,95 «–»	7,97 «–»	$1{,}51\times10^{4}$ «–»	$1{,}65\times10^{4}$ «–»
Skype (video)	32,58 «–»	15,79 «–»	$1{,}92\times10^{-3}$ «–»	1 «+»	$1{,}53\times10^{-2}$ «–»	28,95 «–»	32,58 «–»
E-mail	2067,56 «–»	1033,78 «–»	0.13 «–»	65,47 «–»	1 «+»	$1{,}90\times10^{3}$ «–»	2067,56 «–»
HTTP	1,09 «+»	0,54 «–»	$6{,}62\times10^{-5}$ «–»	$3{,}45\times10^{-2}$ «–»	$5{,}29\times10^{-4}$ «–»	1 «+»	1,09 «+»
FTP	1,0 «+»	0,5 «–»	$6{,}07\times10^{-5}$ «–»	$3{,}17\times10^{-2}$ «–»	$4{,}84\times10^{-4}$ «–»	0,92 «+»	1 «+»

Рассмотрим содержательную часть табл. 4, т.е. случай представления сетевого трафика в виде числа пакетов, передаваемых в единицу времени. Анализ показывает, что в этом случае применение дисперсионного анализа позволяет однозначно детектировать к выборочным статистическим данным по критерию Фишера такие сервисы и службы как: YouTube (720p), Skype (voice), Skype (video), E-mail, HTTP. Показатели рассеивания (выборочные дисперсии)

этих трафиков значительно различаются между собой, их отношение лежит в критической области и соответствующие гипотезы об однородности статистических данных по показателям рассеивания отвергаются. В то же время результаты исследований показывают, что значения выборочных дисперсий для трафиков YouTube (360p) и FTP близки, их отношения не лежат в критической области и гипотезы об однородности статистических данных по показателям рассеивания принимаются.

Рассмотрим данные, приведенные в табл. 5, т.е. рассмотрим случай представления сетевого трафика в виде числа бит, передаваемых в единицу времени. Проведенный анализ показывает, что в этом случае применение математического аппарата дисперсионного анализа позволяет правильно детектировать по критерию Фишера такие сервисы и службы как: YouTube (360p), Skype (voice), Skype (video), E-mail. Данные трафиков YouTube (720p), HTTP и FTP статистически однородны по показателям рассеивания и дисперсионный анализ выборочных данных не позволяет различить эти службы и сервисы.

Таким образом, проведенные исследования позволяют обосновать следующие, важные в прикладном отношении, рекомендации.

1. В случае представления сетевого трафика в виде числа пакетов, передаваемых в единицу времени, математический аппарат дисперсионного анализа позволяет правильно детектировать следующие телекоммуникационные сервисы и информационные службы: YouTube (720p), Skype (voice), Skype (video), E-mail, HTTP. Сетевые трафики сервиса YouTube (360p) и протоколов FTP статистически однородны по показателям рассеивания, они статистически неразличимы по критерию Фишера друг от друга, однако аппарат дисперсионного анализа позволяет с высокой вероятностью правильно их отличить от трафиков других служб и сервисов телекоммуникационной системы.

2. В случае представления сетевого трафика в виде числа бит, передаваемых в единицу времени, наблюдается однородность сетевых трафиков YouTube(720p), HTTP и FTP по показателям рассеивания. В то же время по критерию Фишера с высокой вероятностью удается правильно детектировать YouTube(360p), Skype (voice), Skype (video), E-mail.

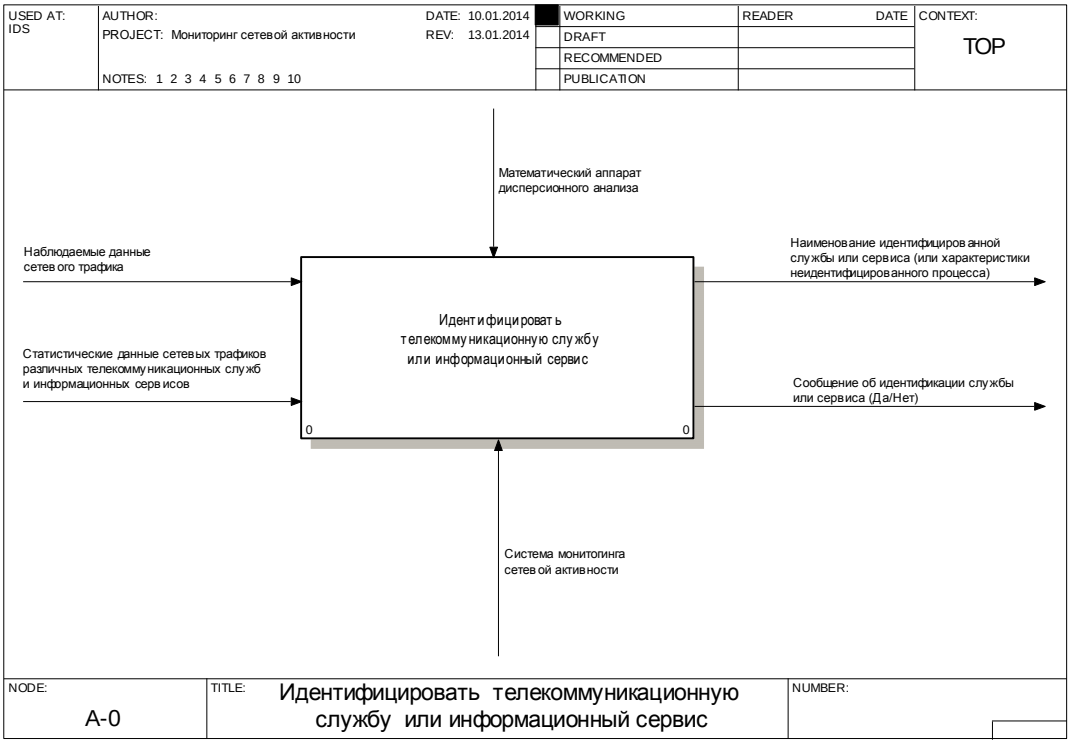
3. Для мониторинга сетевой активности целесообразно использовать различные представления сетевого трафика с расчетом статистики Фишера по отношению выборочных дисперсий. Это позволит правильно детектировать все рассмотренные виды трафиков и реализовать мониторинг сетевой активности различных служб и сервисов телекоммуникационной системы.

4. Окончательное решение о идентификации телекоммуникационных служб и информационных сервисов по наблюдаемому сетевому трафику должно приниматься по мажоритарному правилу с учетом результатов проверки гипотез об однородности наблюдаемых данных сетевого трафика и данных различных служб и сервисов, представленных как в форме «*пакет/с*», так и в форме «*бит/с*». Именно такой подход позволит наиболее полно использовать накопленные статистические данные о свойствах сетевого трафика, характерных различным телекоммуникационным службам и информационным сервисам.

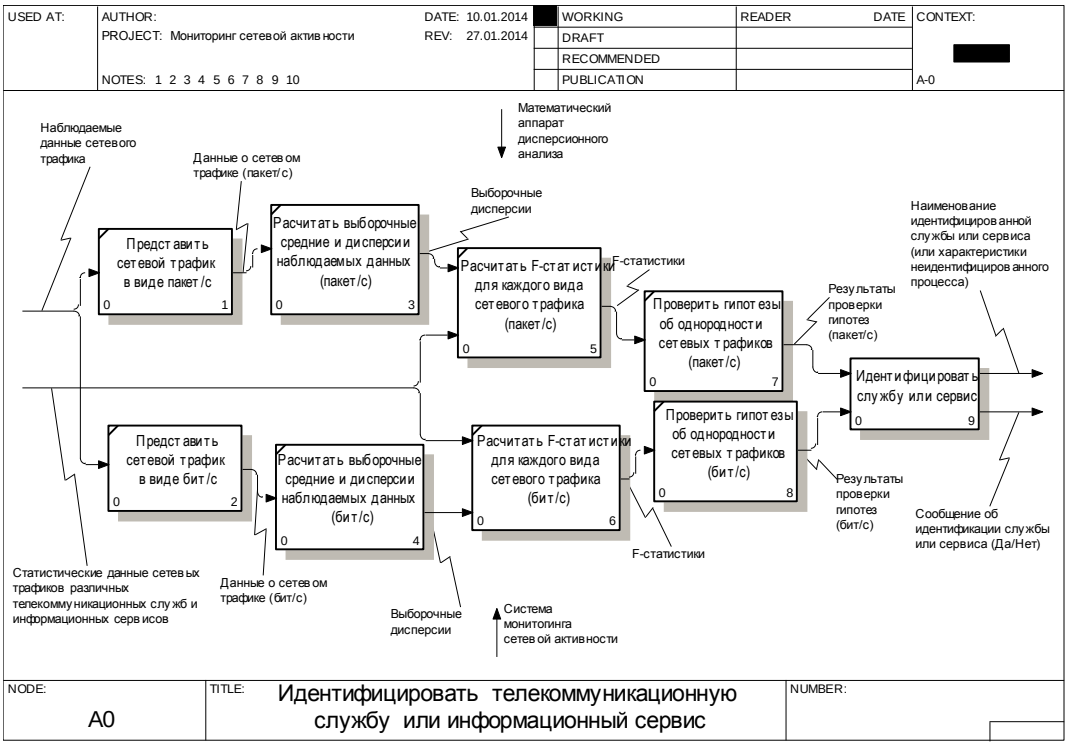
Анализ полученных результатов и обоснование предложений по их использованию

Полученные результаты экспериментальных исследований позволяют обосновать практические рекомендации по идентификации сетевых служб и сервисов для совершенствования методов и средств мониторинга сетевой активности. В частности, мониторинг сетевой активности с расчетом статистики Фишера может быть выполнен по схеме, приведенной на рис. 9[3].

[3]При разработке схемы мониторинга использована методология функционального моделирования IDEF0 [17], и соответствующий программный инструментарий AllFusionProcessModeler r7.



а



б

Рис. 9. Фрагмент системы мониторинга сетевой активности

В качестве входов основного функционального блока А-0 «Идентифицировать телекоммуникационную службу или информационный сервис» выступают (рис. 9, *а*):

– наблюдаемые данные сетевого трафика, получаемые от сенсорной подсистемы сбора событий (см. рис. 10);

– статистические данные сетевых трафиков различных телекоммуникационных служб и информационных сервисов.

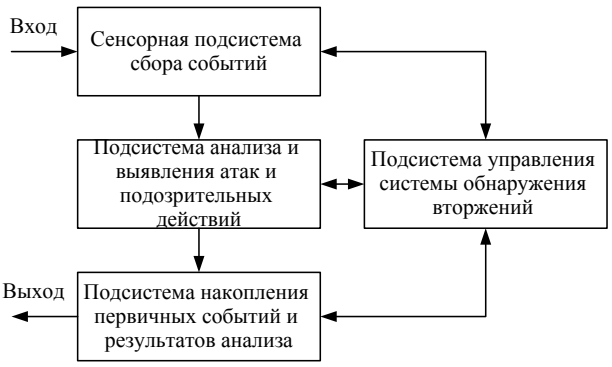


Рис. 10. Структурная схема системы обнаружения вторжений

В качестве управления блока А-0 используется математический аппарат дисперсионного анализа, который определяет условия и необходимые функции для идентификации телекоммуникационных служб или информационных сервисов.

Данные, полученные в результате выполнения функции А-0, содержат:

– наименование идентифицированной службы или сервиса. В случае невозможности идентификации с известной службой (сервисом) выводятся характеристики неидентифицированного процесса;
– сообщение об идентификации службы или сервиса (Да/Нет).

В качестве основного механизма, реализующего функциональный блок А-0, является система мониторинга сетевой активности, которая идентифицирует средства, поддерживающие выполнение соответствующих функций.

Основными функциями, приведенными на рис. 9, *б*, являются (детализация блока А0):

– функции А1, А2, которые представляют наблюдаемые данные сетевого трафика в форме «*пакет/с*» и «*бит/с*»;

– функции А3, А4, которые производят расчет выборочных средних и дисперсий для сетевого трафика в форме «*пакет/с*» и «*бит/с*»;

– функции А5, А6, которые производят расчет значений статистики Фишера для каждой службы или сервиса в форме «*пакет/с*» и «*бит/с*»;

– функции А7, А8, которые реализуют статистическую проверку гипотез об однородности наблюдаемых данных и данных различных телекоммуникационных служб и сервисов, представленных в форме «*пакет/с*» и «*бит/с*»;

– функция А9, реализующая обработку результатов проверки гипотез и формирующая выходные данные в виде наименования идентифицированной службы или сервиса и сообщения об идентификации (Да/Нет).

Таким образом, применение методов статистического анализа позволяет проводить мониторинг сетевой активности для обнаружения вторжений и предотвращения их воздействия на защищаемые информационные ресурсы.

Выводы

Проведенные исследования показали, что статистические методы анализа сетевого трафика являются эффективным средством детектирования аномалий телекоммуникационных систем. СОПВ по соответствующим статистическим критериям определяет «нормальную» сетевую активность отдельных служб и информационных сервисов телекоммуникационной системы, «аномальность» функционирования определяется пороговой величиной (критической областью).

Исследованы статистические критерии проверки однородности сетевого трафика (критерии Вилькоксона и Фишера), полученные результаты показали, что в большинстве случаев гипотеза об однородности трафиков различных служб и сервисов отвергается, т.е наблюдается правильное принятие решения о принадлежности выборок различным процессам. Это положение может быть положено в основу работы одного из элементов системы мониторинга сетевой активности. Обоснованы практические рекомендации по построению фрагмента системы мониторинга сетевой активности, в качестве входов основного функционального

блока которой выступают как наблюдаемые данные сетевого трафика, получаемые от сенсорной подсистемы сбора событий, так и статистические данные сетевых трафиков различных телекоммуникационных служб и информационных сервисов.

Таким образом, применение методов статистического анализа позволяет проводить мониторинг сетевой активности для обнаружения вторжений и предотвращения их воздействия на защищаемые информационные ресурсы

Перспективным направлением дальнейших исследований является разработка программного комплекса, позволяющего реализовать предлагаемую схему мониторинга сетевой активности в составе перспективной СОПВ, экспериментальная проверка адекватности приведенных рассуждений и конструктивности предложенных решений.

Список литературы: 1. *Карпук, Н. М.* Статистический анализ сетевого трафика. Электронная библиотека Белорусского государственного университета. – 2008. – С. 116 – 119. Режим доступа: http://elib.bsu.by/bitstream/123456789/7401/1/6.pdf. *NIST Special* Publication 800-94. Guide to Intrusion Detection and Prevention Systems (IDPS). – Computer Security Division Information Technology Laboratory National Institute of Standards and Technology, Gaithersburg. – 127 pages (February 2007). 3. *Brian Caswell, Jay Beale, Andrew Baker*. Snort Intrusion Detection and Prevention Toolkit. – Syngress Media, U.S. 2006. . http://www.lehmanns.de/ shop/sachbuch-ratgeber/21797174-9780080549279 – snort-intrusion-detection-and-prevention-toolkit#drm1. 4. *Ушаков, Д.В*. Развитие принципов функционирования систем обнаружения сетевых вторжений на основе модели защищенной распределенной системы : дис… канд. техн. наук: 05.13.19. – Москва, 2005. – 175 с. 5. *Запечников, С.В., Милославская, Н.Г., Толстой, А.И., Ушаков, Д.В.* Информационная безопасность открытых систем. Учебник для вузов : в 2-х т. – М., 2008. – Т. II: Средства защиты в сетях. – 558 с. 6. *Comparison* of Firewall, Intrusion Prevention and Antivirus Technologies. http://www.juniper.net/solutions/literature/white_papers /200063.pdf 7. *Intrusion* Prevention Systems (IPS). http://www.securecomputing.com/ pdf/Intru-Preven-WP1-Aug03-vF.pdf 8. *Intrusion* Prevention Systems (IPS). http://hosteddocs.ittoolbox.com/ BW013004.pdf 9. *State* of the Practice of Intrusion Detection Technologies. http://www.sei.cmu.edu/pub/documents/99.reports/pdf/99tr028.pdf 10. *Wireless* Intrusion Detection and Response. http://users.ece.gatech.edu/~owen/Research/ Conference%20Publications/wireless_IAW2003.pdf 11. *Anomaly* Detection in IP Networks. http://users.ece.gatech.edu/~jic/ sig03.pdf 12. *Design* and Implementation of an Anomaly Detection System: an Empirical Approach. http://luca.ntop.org/ADS.pdf 13. *Host*-Based Intrusion Detection Systems. http://staff.science.uva.nl/~delaat/snb-2004-2005/p19/report.pdf 14. *Олифер, В.Г., Олифер, Н.А.* Компьютерные сети. Принципы, технологии, протоколы. – СПб. : Питер, 2010. – 944 с. 15. *Смирнов, Н. В., Дунин,-Барковский, И. В.* Курс теории вероятностей и математической статистики для технических приложений ; изд. 2. – М. : Наука, 1969. – 512 с. 16. *Шефе, Г.* Дисперсионный анализ : пер. с англ. ; 2-е изд. – М. : Наука, 1980. – 512 с. 17. *РД IDEF 0 – 2000*. Методология функционального моделирования IDEF0. Руководящий документ. Изд. официальное. – М. : ГОССТАНДАРТ РОССИИ. – 75 с.

Харьковский национальный
университет радиоэлектроники

Поступила в редколлегию 18.01.2014