УДК 621.396

А.А. Кузнецов[1], А.А. Смирнов[2], Д.А. Даниленко[2]

[1] *Харьковский национальный университет радиоэлектроники, Харьков*
[2] *Кировоградский национальный технический университет, Кировоград*

ДИСПЕРСИОННЫЙ АНАЛИЗ СЕТЕВОГО ТРАФИКА ДЛЯ ОБНАРУЖЕНИЯ И ПРЕДОТВРАЩЕНИЯ ВТОРЖЕНИЙ В ТЕЛЕКОММУНИКАЦИОННЫХ СИСТЕМАХ И СЕТЯХ

Рассматриваются системы обнаружения и предотвращения вторжений в современных телекоммуникационных системах и сетях. Исследуются методы мониторинга событий, состоящие в анализе сетевой активности отдельных служб и информационных сервисов телекоммуникационных систем и сетей. Предлагается использовать математический аппарат дисперсионного анализа для обработки результатов моделирования телекоммуникационных систем и исследования статистических свойств сетевого трафика при определении значимости расхождения или совпадения характеристик. Предлагаемый подход состоит в использовании статистического критерия Фишера, основанного на оценке отношения выборочных дисперсий, что позволяет с заданным уровнем значимости проверять гипотезу об однородности статистических свойств сетевого трафика относительно показателя рассеивания (дисперсии). Полученные результаты экспериментальных исследований рекомендуется использовать для совершенствования механизмов мониторинга сетевой активности отдельных служб и информационных сервисов, в том числе и для обнаружения и предотвращения вторжений в телекоммуникационных системах и сетях.

Ключевые слова: *телекоммуникационные системы и сети, система обнаружения и предотвращения вторжений, дисперсионный анализ.*

Введение

Постановка проблемы исследования. Современное развитие телекоммуникационных систем и сетей и применяемых компьютерных технологий привело к появлению качественно новых услуг и сервисов в информационной сфере, внедрению передовых технологий обработки и передачи данных и их доступности широкой пользовательской аудитории [1]. В то же время интенсивное развитие современных компьютерных технологий привело к появлению новых угроз безопасности информации, возникновению новых форм и способов несанкционированного доступа к вычислительным ресурсам телекоммуникационных систем и сетей [1 – 4]. В частности, наибольшую уязвимость представляют применяемые методы сетевого управления, технологии доступа к предоставляемым сервисам и услугам, процессы мониторинга состояния телекоммуникационных систем и сетей. Под воздействием вредоносного программного обеспечения отдельные коммуникационные и вычислительные компоненты могут быть переведены в несанкционированные режимы функционирования, приводящие к сбоям, различным нарушениям установленного порядка их использования, уничтожению, искажению, блокированию, несанкционированной утечки обрабатываемой и передаваемой информации, а также к нарушению работы методов и алгоритмов маршрутизации между узлами телекоммуникационной системы [2 – 4]. Следовательно, разработка и исследование методов мониторинга сетевой активности, технологий обнаружения вредоносного программного обеспечения и предотвращения его воздействия на защищаемые инфокоммуникационные ресурсы является актуальной научно-прикладной проблемой, ее решение непосредственно связано с обеспечением безопасности современных телекоммуникационных систем и сетей и применяемых компьютерных технологий.

В работе предлагается использовать математический аппарат дисперсионного анализа для исследования свойств сетевого трафика различных служб и информационных сервисов при определении значимости расхождения или совпадения их характеристик. Полученные результаты могут быть использованы для совершенствования механизмов мониторинга сетевой активности, в том числе и для обнаружения и предотвращения вторжений в телекоммуникационных системах и сетях.

Анализ последних исследований и публикаций. Для обеспечения безопасности в современных телекоммуникационных системах и сетях применяются различные организационно-технические мероприятия, наиболее эффективные из которых состоят в построении т.н. систем обнаружения (Intrusion Detection System – IDS) и предотвращения (Intrusion Prevention System – IPS) вторжений [2 – 13]. В основе функционирования IDS и IPS лежит сбор, анализ и обработка информации о событиях, связанных с

безопасностью защищаемой телекоммуникационной системы, накопление полученных данных и, на основе результатов проведенного анализа (мониторинга) сетевой активности отдельных служб и сервисов, принятие решения о состоянии защищаемой системы с выявлением и возможным противодействием несанкционированному использованию инфокоммуникационных ресурсов [2 – 6].

Под системой обнаружения вторжений (СОВ) понимают программное или аппаратное средство, предназначенное для выявления фактов неавторизованного доступа в компьютерную систему или сеть либо несанкционированного управления [2 – 4]. СОВ обеспечивают дополнительный уровень защиты компьютерных систем за счет обнаружения некоторых типов вредоносной активности, которая может нарушить безопасность компьютерной системы. К такой активности относятся сетевые атаки на наиболее уязвимые сервисы, атаки, направленные на повышение привилегий, неавторизованный доступ к важным ресурсам, а также действия вредоносного программного обеспечения (компьютерных вирусов, троянов и червей) [2].

Под системой предотвращения вторжений (СПВ) понимают программную или аппаратную систему сетевой и компьютерной безопасности, обнаруживающую вторжения или нарушения безопасности, а также реализующую автоматическую защиту от выявленных нарушений [2 – 4].

Системы IPS следует рассматривать как расширение систем IDS. В тоже время СПВ отличаются необходимостью отслеживания сетевой активности в реальном времени с быстрым реагированием посредством реализации соответствующих действия по предотвращению выявленных атак. Возможные меры предотвращения атак состоят в блокировке потоков трафика в телекоммуникационной сети, сбросе соединений, выдачи сигналов оператору и т.д. Также IPS могут выполнять дефрагментацию пакетов, переупорядочивание пакетов TCP для защиты от пакетов с измененными SEQ (номерами очереди) и ACK (номерами подтверждения) [2].

На рис. 1 представлена типичная модель подключения и работы сетевой СОВ (Network-based IDS, NIDS) и сетевой СПВ (NIPS) [2 – 6].

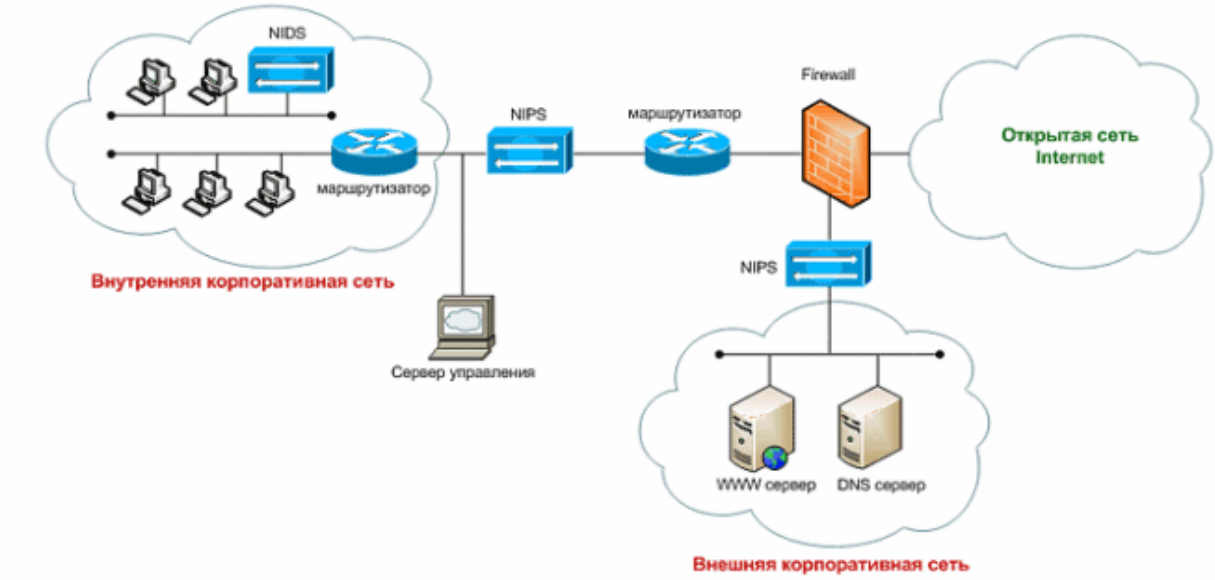

Рис. 1. Типичное размещение сетевых IDS и IPS систем

Архитектура СОВ включает (рис. 2) [2, 3, 6]:

– сенсорную подсистему, предназначенную для сбора событий, связанных с безопасностью защищаемой системы;

– подсистему анализа, предназначенную для выявления атак и подозрительных действий на основе данных сенсоров;

– хранилище, обеспечивающее накопление первичных событий и результатов анализа;

– консоль управления, позволяющую конфигурировать СОВ, наблюдать за состоянием защищаемой системы и СОВ, просматривать выявленные подсистемой анализа инциденты.

В сетевой СОВ сенсоры расположены на важных для наблюдения точках сети, часто в демилитаризованной зоне, или на границе сети (см. рис. 1) [2, 3, 6, 9]. Сенсор перехватывает весь сетевой трафик и анализирует содержимое каждого пакета на наличие вредоносных компонентов. NIDS получает доступ к сетевому трафику, подключаясь к коммутаторам сети и отслеживает вторжения, проверяя сетевой трафик и ведя наблюдение за несколькими узлами (хостами).

Соответствующее программное обеспечение используется для предотвращения проникновения, блокирования возможных атак.

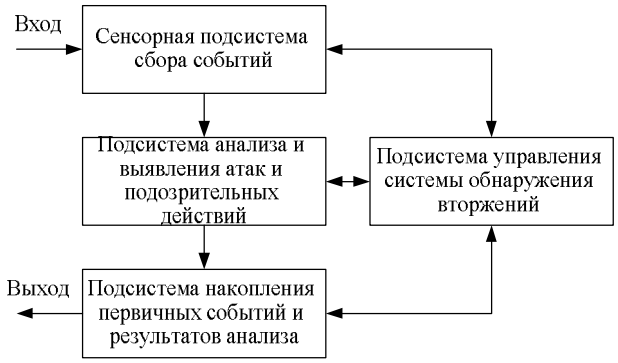

Рис. 2. Структурная схема системы обнаружения вторжений

Протокольные СОВ (Protocol-based IDS, PIDS) используются для отслеживания трафика, нарушающего правила определенных протоколов либо синтаксис языка (например, SQL) [2, 3]. Такая СОВ представляет собой систему (либо агента), которая отслеживает и анализирует коммуникационные протоколы со связанными системами или пользователями. Для веб-сервера подобная СОВ обычно ведет наблюдение за HTTP и HTTPS протоколами. При использовании HTTPS СОВ должна располагаться на таком интерфейсе, чтобы просматривать HTTPS пакеты еще до их шифрования и отправки в сеть.

Основанная на прикладных протоколах СОВ (Application Protocol-based IDS, APIDS) – это система (или агент), которая ведет наблюдение и анализ данных, передаваемых с использованием специфичных для определенных приложений протоколов [2, 3]. Например, на веб-сервере с SQL базой данных соответствующая СОВ будет отслеживать содержимое SQL команд, передаваемых на сервер.

В узловых (*хостовых*) *СОВ* (Host-based IDS, HIDS) сенсор обычно является программным агентом, который ведет наблюдение за активностью узла сети, на который он установлен[2, 3, 13]. Для отслеживания вторжений используется анализ системных вызовов, приложений, модификаций файлов (исполняемых, файлов паролей, системных баз данных), состояния узла сети и прочих источников.

Гибридная СОВ совмещает два и более подходов к разработке СОВ. Данные от агентов на хостах комбинируются с сетевой информацией для создания наиболее полного представления о безопасности сети.

Таким образом, в пассивной СОВ при обнаружении нарушения безопасности, информация о нарушении записывается в хранилище данных, а сигналы опасности отправляются на консоль и/или администратору системы по определенному каналу связи. В активной СПВ (IPS) производятся ответные действия на нарушение, сбрасывая соединение или перенастраивая межсетевой экран для блокирования трафика от злоумышленника. Ответные действия могут проводиться автоматически либо по команде оператора.

По выполняемым активным действиям СПВ следует классифицировать следующим образом [1 – 13]:

– *сетевые IPS* (NIPS) отслеживают трафик в компьютерной сети и блокируют подозрительные потоки данных[2 – 9];

– *IPS для беспроводных сетей* (Wireless Intrusion Prevention Systems, WIPS) проверяет активность в беспроводных сетях. В частности, обнаруживают неверно сконфигурированные точки беспроводного доступа к сети, атаки человек посередине, спуфинг MAC-адресов [2 – 10];

– *поведенческий анализ сети* (Network Behavior Analysis, NBA) анализирует сетевой трафик, идентифицирует нетипичные потоки, например DoS и DDoS атаки [2 – 9, 11, 12];

– *система предотвращения вторжений для отдельных узл*ов (Host-based Intrusion Prevention, HIPS) содержит резидентные программы, обнаруживающие подозрительную активность на компьютере [2 – 9, 13].

Проведенный анализ показал, что наиболее эффективными средствами обеспечения безопасности современных телекоммуникационных систем и сетей являются СОВ и СПВ. В основе их функционирования лежит комплексное использование результатов анализа сетевого трафика, содержимого передаваемых пакетов на наличие вредоносных компонентов (в NIDS), проверки корректности выполнения установленных правил работы используемых протоколов (в PIDS), а также анализа активности узлов сети, системных вызовов, приложений, модификаций файлов (в HIDS) [2 – 13]. Этот подход считается наиболее перспективным направлением в развитии средств защиты компьютерных сетей и обеспечения требуемых показателей безопасности [2 – 5]. Проведенный анализ последних исследований и публикаций [2 – 13] показывает, что в основе работы наиболее развитых СОВ и СПВ лежит использование статистических данных о сетевом трафике для выявления несанкционированных режимов функционирования телекоммуникационной системы, нарушения работы методов и алгоритмов маршрутизации, воспрепятствования уничтожению, искажению, блокированию, несанкционированной утечки обрабатываемой и передаваемой информации.

Таким образом, проведение экспериментальных исследований свойств сетевого трафика имеет важное значение как для теоретического обоснования методов обнаружения и предотвращения вторжений, так и для разработки практических рекомендаций по построению программных и аппаратных средств мониторинга сетевой активности отдельных служб и информационных сервисов.

Методика экспериментальных исследований

Современные методы имитационного моделирования предоставляют возможность накапливать результаты статистических испытаний и эффективно проводить различную обработку полученных данных, в частности, выполнять сравнение случайных параметров исследуемого процесса с целью определения значимости расхождения или совпадения их характеристик [14, 15]. Один из наиболее развитых методов такой обработки, основанный на оценке отношений выборочных дисперсий, позволяет подтвердить или опровергнуть статистическую гипотезу об однородности результатов моделирования по показателю рассеивания (дисперсии) [15]. В данной работе предлагается использовать математический аппарат дисперсионного анализа для обработки результатов моделирования телекоммуникационных систем и исследования свойств сетевого трафика различных служб и информационных сервисов при определении значимости расхождения или совпадения их характеристик.

Введем следующие обозначения и определения [14, 15]. Пусть в результате эксперимента с имитационной статистической моделью, состоящего из N наблюдений, получено N значений $x_1, x_2, ..., x_N$ исследуемой случайной величины X. Необходимо по этим полученным данным дать описание случайной величины X, т.е. необходимо определить ее характеристики. В практике моделирования и обработки экспериментальных данных очень часто необходимо решать проблему подтверждения или опровержения гипотезы о принадлежности двух или более выборок одной генеральной совокупности. Признаки, по которым проводится сравнительная оценка, часто не являются детерминированными, обладают рассеиванием. Наиболее распространенной мерой рассеивания, используемой в теории вероятностей и математической статистике, является дисперсия (от лат. dispersio – рассеяние).

В статистическом понимании дисперсия:

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - x^*)^2$$

есть среднее арифметическое квадратов отклонений величин x_i от их среднего арифметического $x^* = (x_1 + x_2 + ... + x_n)/n$, т.е. дисперсия есть мера отклонения от статистического среднего. В сложных технических системах дисперсия характеризует важные конструкторские и технологические показатели. В этом смысле при проведении исследовании различных параметров технических систем наиболее важной характеристикой сравнительных оценок является именно дисперсия, т.к. она обладает наибольшей общностью и позволяет, помимо прочего, проверять гипотезу о равенстве средних значений выборок.

Таким образом, дисперсионный анализ является одним из эффективных механизмов исследования сложных технических систем, как наиболее общий и часто применяемый на практике метод сравнения качеств различных объектов.

Современные приложения дисперсионного анализа охватывают широкий круг задач и трактуются обычно в терминах статистической теории выявления систематических различий между результатами непосредственных измерений, выполненных при тех или иных меняющихся условиях. Если значения неизвестных постоянных $a_1, ..., a_n$ могут быть измерены с помощью различных методов или измерительных средств $M_1, ..., M_m$ и в каждом случае систематическая ошибка может зависеть как от выбранного метода, так и от неизвестного измеряемого значения a_i, то результаты измерений x_{ij} представляют собой суммы вида:

$$x_{ij} = a_i, + b_{ij} + d_{ij},\ i = 1, 2, ..., n;\ j = 1, 2, ..., m,$$

где b_{ij} – систематическая ошибка, возникающая при измерении a_i по методу M_j, d_{ij} – случайная ошибка.

Такую модель принято называть двухфакторной схемой дисперсионного анализа [14, 15], где первый фактор – измеряемая величина, а второй – метод измерения.

Дисперсии эмпирических распределений, соответствующих множествам случайных величин x_{ij}, $x_{ij} - x_{i*} - x_{*j} + x_{**}$, x_{i*} и x_{*j}, где [14, 15]:

$$x_{i*} = \frac{1}{m}\sum_j x_{ij},\ x_{j*} = \frac{1}{n}\sum_i x_{ij},$$

$$x_{**} = \frac{1}{n}\sum_i x_{i*} = \frac{1}{m}\sum_j x_{j*}$$

выражаются формулами:

$$s^2 = \frac{1}{mn}\sum_i\sum_j (x_{ij} - x_{**})^2,$$

$$s_0^2 = \frac{1}{mn}\sum_i\sum_j (x_{ij} - x_{i*} - x_{j*} + x_{**})^2,$$

$$s_1^2 = \frac{1}{n}\sum_i (x_{i*} - x_{**})^2,\ s_2^2 = \frac{1}{m}\sum_j (x_{j*} - x_{**})^2.$$

Эти дисперсии удовлетворяют тождеству [14, 15]:

$$s^2 = s^2_0 + s^2_1 + s^2_2,$$

которое и объясняет происхождение названия дисперсионного анализа. Если величины систематических ошибок не зависят от метода измерений (т. е. между методами измерений нет систематических расхождений), то отношение s^2_2/s^2_0 близко к единице. Это свойство лежит в основе *критерия для статистического выявления систематических расхождений*: если s^2_2/s^2_0 значимо отличается от единицы, то гипотеза об отсутствии систематических расхождений отвергается.

Значимость отличия определяется в согласии с законом распределения вероятностей случайных ошибок измерений. В частности, если все измерения равноточные и случайные ошибки подчиняются нормальному распределению, то критические значения для отношения s^2_2/s^2_0 определяются с помощью таблиц так называемого F-распределения (распределения дисперсионного отношения или распределения Фишера) [14, 15].

Изложенная схема позволяет лишь обнаружить наличие систематических расхождений и, вообще говоря, непригодна для их численной оценки с последующим исключением из результатов наблюдений. Эта цель может быть достигнута только при многократных измерениях (при повторных реализациях указанной схемы).

Таким образом, суть дисперсионного анализа состоит в проверке гипотезы о тождественности выборочных дисперсий одной и той же генеральной совокупности [14, 15].

Пусть имеются две выборки $x_1, x_2, ..., x_{N1}$ и $y_1\ y_2, ..., y_{N2}$ объемом N_1 и N_2, соответственно, 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 [14, 15].

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 [14]:

$$F = \frac{U/k_1}{V/k_2},$$

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).

Другими словами, 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, полученных из независимых выборок, взятых из нормальных совокупностей:

$$\sigma_1^2 = \frac{1}{N_1}\sum_{i=1}^{N_1}(x_i - x^*)^2,\ \sigma_2^2 = \frac{1}{N_2}\sum_{i=1}^{N_2}(y_i - y^*)^2, \quad (1)$$

$$x^* = (x_1 + x_2 + ... + x_{N_1})/N_1,$$

$$y^* = (y_1 + y_2 + ... + y_{N_2})/N_2. \quad (2)$$

Для подтверждения или опровержения гипотезы об однородности исследуемых выборок необходимо выбрать уровень значимости q, численно равный вероятности *неприемлемых* отклонений от принятой гипотезы.

Вид функции плотности распределения Фишера приведен на рис. 3, где также обозначены области неприемлемых *значений F*.

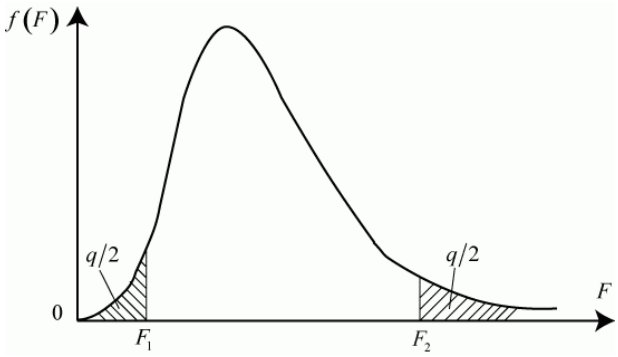

Рис. 3 . График плотности F-распределения

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 т.н. критическую область, принятая гипотеза не опровергается. *Чем меньше* уровень значимости *q*, *тем меньше* вероятность забраковать проверяемую гипотезу, когда она верна, т. е. совершить *ошибку первого рода*. Но *с уменьшением уровня значимости* (увеличения F_2) расширяется область допустимых ошибок, что приводит *к увеличению вероятности принятия неверного решения*, т. е. совершения *ошибки второго рода*. Следовательно, суждение о подтверждении или отклонении выдвинутой гипотезы высказывается с определенной степенью достоверности.

Задачу проводимых экспериментальных исследований сформулируем как задачу проверки гипотезы об однородности наблюдаемых трафиков различных телекоммуникационных служб и информационных сервисов по выборочным дисперсиям. Поставленную задачу проверки гипотезы будем решать следующим образом:

1. Для различных телекоммуникационных служб и информационных сервисов по результатам N наблюдений сетевого трафика сформируем выборку из N значений $x_1, x_2, ..., x_N$ исследуемой случайной величины X.

2. Для каждой выборки по выражениям (1–2) рассчитаем значения выборочных средних (x^* и y^*) и дисперсий (σ_1^2 и σ_2^2).

3. Выберем уровень значимости q, численно равный вероятности неприемлемых отклонений от принятой гипотезы и рассчитаем соответствующие граничные точки F_1 и F_2 допустимых значений F.

4. Рассчитаем статистику теста F и проверим условие $F_1 \le F \le F_2$.

5. В случае попадания значения F в критическую область гипотеза отвергается, в случае непопадания – принимается.

Полученные результаты проводимых исследований позволят экспериментально подтвердить или опровергнуть гипотезу об однородности свойств сетевого трафика различных телекоммуникационных служб и информационных сервисов методом дисперсионного анализа, обосновать практические рекомендации по построению программных и аппаратных средств мониторинга сетевой активности, обнаружения вредоносного программного обеспечения и предотвращения его воздействия на защищаемые инфокоммуникационные ресурсы и обеспечения безопасности современных телекоммуникационных систем и сетей.

Результаты экспериментальных исследований

Для проведения экспериментальных исследований свойств сетевого трафика были использованы эмпирические данные, полученные в результате работы программного анализатора (снифера) «Wireshark». Выбор этого программного сетевого анализатора связан с возможностью перехвата трафика сетевого интерфейса в режиме реального времени. В ходе практических замеров с использованием снифера «Wireshark» оценивался объем данных, передаваемых через компьютерную сеть за определённый период времени. Замеры трафика, т.е. объема информации, передаваемого в единицу времени, проводились как по числу пакетов, так и по числу бит данных. При этом эмпирические данные были получены и обобщены не менее чем по 100 000 временным отсчетам.

В качестве исходных данных при проведении экспериментальных исследований были использованы различные телекоммуникационные службы и информационные сервисы, а именно: FTP (File Transfer Protocol) – стандартный протокол, предназначенный для передачи файлов по TCP-сетям; HTTP (HyperText Transfer Protocol) – протокол прикладного уровня передачи данных; электронная почта (e-mail) – технология и предоставляемые ею услуги по пересылке и получению электронных сообщений по распределённой (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 сетевого трафика при загрузке данных с сервиса YouTube (720p), при использовании сервиса Skype в случае голосовой связи (voice) и видеосвязи (video), а также услуг электронной почты (E-mail), протоколов HTTP и FTP приведены на рис. 4 – 9[1].

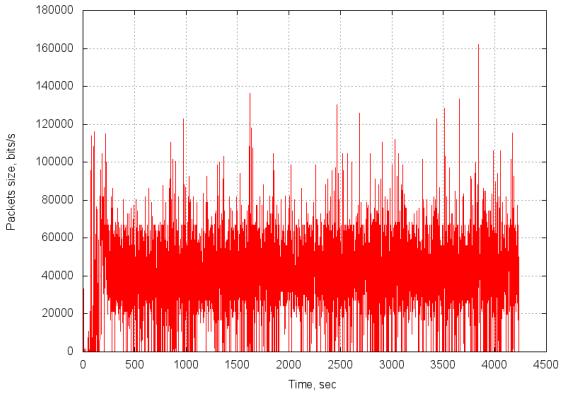

Рис. 4. Фрагмент гистограммы сетевого трафика при загрузке данных с сервиса YouTube (720p, бит/с)

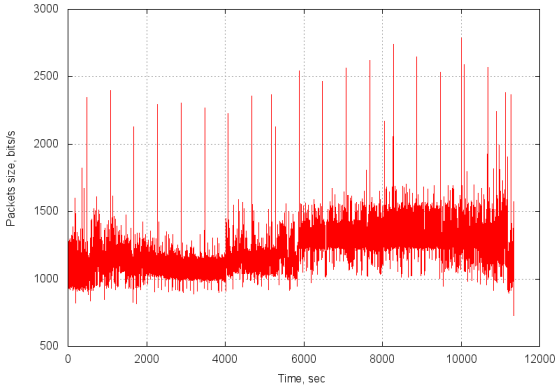

Рис. 5. Фрагмент гистограммы сетевого трафика при обмене данными с использованием Skype (voice, бит/с)

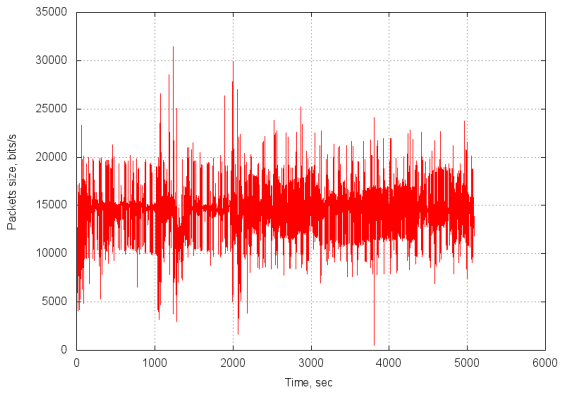

Рис. 6. Фрагмент гистограммы сетевого трафика при обмене данными с использованием Skype (video, бит/с)

[1] Здесь и далее сетевой трафик представлен в виде числа бит данных, переданных в единицу времени (в секунду)

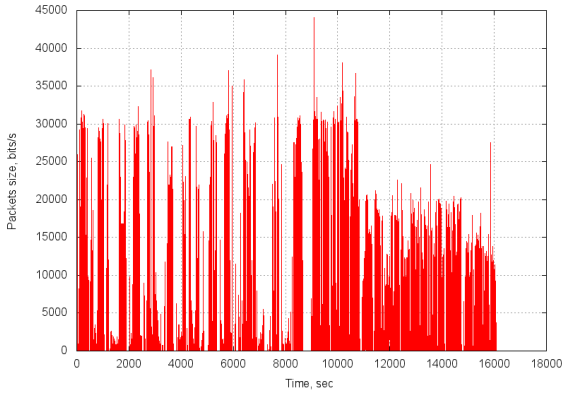

Рис. 7. Фрагмент гистограммы сетевого трафика при передаче электронной почты (бит/с)

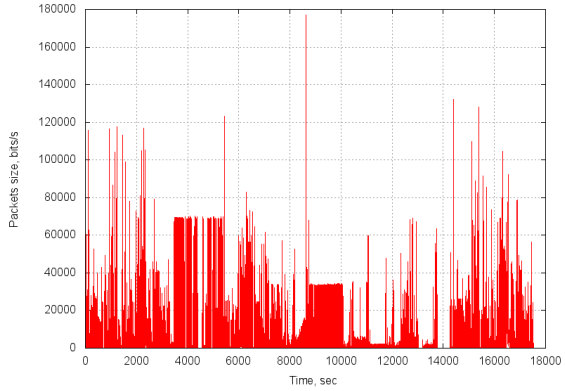

Рис. 8. Фрагмент гистограммы сетевого трафика при передаче данных с использованием HTTP (бит/с)

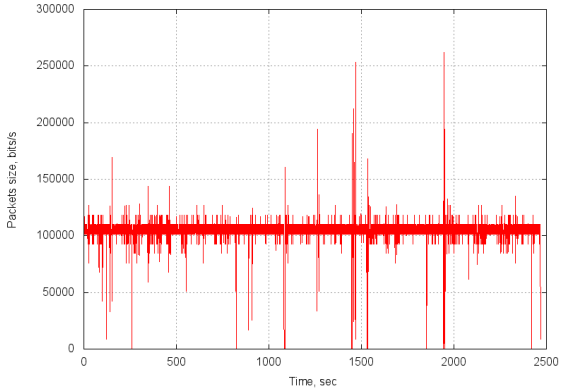

Рис. 9. Фрагмент гистограммы сетевого трафика при передаче данных с использованием FTP (бит/с)

При проведении статистических исследований использованы эмпирические данные по 100 временным отсчетам случайно выбранных отрезков сетевого трафика, соответствующих различным телекоммуникационным службам и информационным сервисам. Другими словами, оценка однородности сетевого трафика проводилась по выборочным данным с использованием основной метрики рассеивания – дисперсии случайной величины.

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 [14]. Так как объем данных, передаваемых через компьютерную сеть за определённый период времени, является случайной величиной, формируемой под влиянием большого числа слабо зависимых случайных факторов, будем считать распределение этой случайной величины нормальным. При этом, естественно, должно соблюдаться условие, что ни один из факторов не является доминирующим при формировании сетевого трафика. (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).

Принимая указанные предположения, воспользуемся аппаратом дисперсионного анализа для проверки статистической гипотезы об однотипности сетевых трафиков рассматриваемых служб и информационных сервисов телекоммуникационной системы. Для этого выполним следующие основные 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, 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. Зададим уровень значимости равным $q = 0{,}1$.

3. Произведем расчет статистики теста так, чтобы: её величина зависела от исходной выборки; по её 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 принятии) 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. 10-11[2].

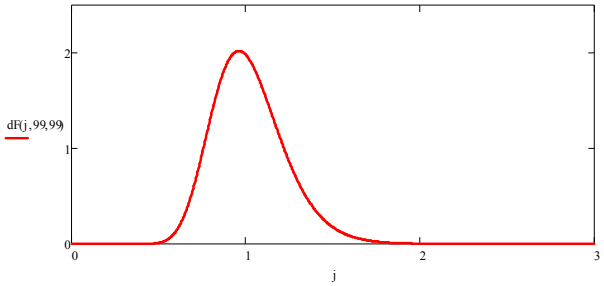

Рис. 10. График плотности вероятности распределения Фишера для числа степеней свободы $k_1 = k_2 = 99$

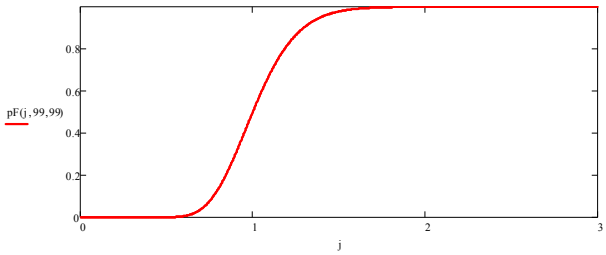

Рис. 11. График интегрального распределения вероятностей Фишера для числа степеней свободы $k_1 = k_2 = 99$

При решении практических задач часто требуется найти значение x, при котором функция распределения $F_x(x)$ случайной величины x принимает заданное значение p, т.е. требуется решить уравнение $F_x(x) = p$. Решения такого уравнения (соответствующие значения x) в теории вероятностей принято называть квантилями [14, 15].

Построим график обратного кумулятивного распределения вероятностей для заданного числа степеней свободы. Этот график описывает поведение квантили интегрального распределения вероятностей, т.е. поведение зависимости $x = F^{-1}_x(p)$. Для рассматриваемого случая, когда в качестве $F_x(x)$ используется интегральное распределение вероятностей Фишера с числом степеней свободы $k_1 = k_2 = 99$ (см. рис. 11), график обратного кумулятивного распределения будет иметь вид, приведенный на рис. 12.

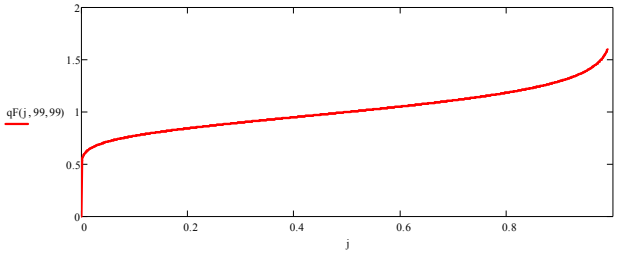

Рис. 12. График обратного кумулятивного распределения вероятностей Фишера

Используя уровень значимости $q = 0.1$ с использованием приведенной на рис. 11 зависимости, найдем такое значение правой граничной точки F_2 функции Фишера F, при котором $1 - F_x(F_2) < q/2 = 0.05$, что эквивалентно нахождению такой квантили $x = F_2$, при которой $x = F^{-1}_x (p = 1 - q/2)$, т.е. правая граничная точка F_2 определяется по правилу $F_2 = F^{-1}_x (0{,}95)$.

Найдем это значение, получим $F_2 = 1{,}394$, что наглядно подтверждается на приведенных рис. 10 – 12. Таким образом, вероятность того, что значение F превысит правую граничную точку F_2 равна q/2:

$$P(F > F_2 = 1{,}394) = q/2 = 0{,}05.$$

Аналогичным способом найдем значение левой граничной точки F_1, при котором $1 - F_x(F_1) < 1 - q/2 = 0.95$, что эквивалентно нахождению такой квантили $x = F_1$, при которой $x = F^{-1}_x (p = q/2)$, т.е. левую граничную точку F_1 определим по правилу $F_1 = F^{-1}_x (0{,}05)$. Получим значение $F_1 = 0{,}717$, что также наглядно демонстрируют зависимости на рис. 10 – 12. Очевидно, что вероятность того, что значение F не превысит левую граничную точку, F_1 также равна q/2:

$$P(F < F_1 = 0{,}717) = q/2 = 0{,}05,$$

а вероятность попадания значения F в критическую область будет, соответственно, равна

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

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_0 принимается, т.е. полагаем, что исследуемые сетевые трафики однотипны, их выборочные дисперсии тождественны одной и той же генеральной дисперсии.

Применим рассмотренный метод дисперсионного анализа к проверке гипотезы об однотипности

[2] При построении графиков функций использован пакет символьной математики MathCad15

различных трафиков, присущих различным службам и информационным сервисам телекоммуникационной сети. Для этого для каждого исследуемого трафика сформируем выборку по 100 временным отсчетам данных, проведем обработку выборочных данных, т.е. проведем оценку выборочных средних и выборочных дисперсий по аналитическим выражениям (1-2).

Полученные результаты экспериментальных исследований сведены в табл. 1. Применим аппарат дисперсионного анализа, полученные результаты экспериментальных исследований применительно к проверке гипотезы об однородности свойств сетевого трафика сведены в табл. 2.

Таблица 1

Результаты 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

Таблица 2

Значения статистики теста и результаты проверки гипотезы об однородности сетевых трафиков относительно выборочных дисперсий

	YouTube	Skype (voice)	Skype (video)	E-mail	HTTP	FTP
YouTube	1 Принимается	$4{,}91\times10^{-5}$ Отвергается	0,03 Отвергается	$4{,}78\times10^{-4}$ Отвергается	0,89 Принимается	0,99 Принимается
Skype (voice)	$1{,}67\times10^{4}$ Отвергается	1 Принимается	518,91 Отвергается	7,97 Отвергается	14914,70 Отвергается	16455 Отвергается
Skype (video)	32,12 Отвергается	$1{,}58\times10^{-3}$ Отвергается	1 Принимается	0,02 Отвергается	28,74 Отвергается	31,71 Отвергается
E-mail	2091,02 Отвергается	0,03 Отвергается	65,09 Отвергается	1 Принимается	1870,91 Отвергается	2064,11 Отвергается
HTTP	1,12 Принимается	$5{,}49\times10^{-5}$ Отвергается	0,04 Отвергается	$5{,}35\times10^{4}$ Отвергается	1 Принимается	1,10 Принимается
FTP	1,03 Принимается	$4{,}98\times10^{-5}$ Отвергается	0,03 Отвергается	$4{,}84\times10^{4}$ Отвергается	0,91 Принимается	1 Принимается

Полученные результаты дисперсионного анализа свидетельствуют о том, что статистический критерий на основе отношения выборочных дисперсий дает надежный механизм проверки однородности сетевого трафика. В частности, дисперсионный анализ с критерием значимости $q = 0{,}1$ позволяет правильно идентифицировать используемые сервисы Skype и E-mail по выборочным наблюдениям из 100 временных отсчетов. Значения выборочных дисперсий для этих видов трафика существенно отличаются от значений выборочных дисперсий других сетевых служб и сервисов, в частности, от трафиков сервиса YouTube, протоколов HTTP, FTP.

В тоже время показатели рассеивания статистических данных для трафиков сервиса YouTube и протоколов HTTP и FTP очень близки по своим значениям. Статистика теста, полученная на основе расчета отношения выборочных дисперсий, отличается для этих служб не значительно, что свидетельствует о похожести (однородности) соответствующих данных. Практически это означает, что метод дисперсионного анализа не позволяет правильно различить эмпирические данные трафика YouTube и трафиков HTTP и FTP, они однородны по показателям статистического рассеивания.

Выводы

Проведенные исследования показали, что применение методов дисперсионного анализа позволяет подтвердить или опровергнуть гипотезу об однородности свойств сетевого трафика для различных телекоммуникационных служб и информационных сервисов. В частности, в ходе исследований по отношению выборочных дисперсий наблюдаемого сетевого трафика установлена разнородность соответствующих статистических данных. Это позволяет с высокой вероятностью детектировать сетевую активность отдельных телекоммуникационных служб и информационных сервисов, т.е. успешно проводить мониторинг сетевой активности для обнаружения вредоносного программного обеспечения и предотвращения его воздействия на защищаемые инфокоммуникационные ресурсы и обеспечения безопасности современных телекоммуникационных систем и сетей. Совершенствование соответствующих механизмов мониторинга сетевой активности для

систем обнаружения и предотвращения вторжений является *перспективным направлением дальнейших исследований.*

Список литературы

1. Олифер В.Г. Компьютерные сети. Принципы, технологии, протоколы / В.Г. Олифер, Н.А. Олифер. – СПб.: Питер, 2010. – 944 с.

2. 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 p. (February 2007).

3. Brian Caswell. Snort Intrusion Detection and Prevention Toolkit / Brian Caswell, Jay Beale, Andrew Baker. [Электронный ресурс] // Syngress Media, U.S. 2006. – Режим доступа к ресурсу: http://www.lehmanns.de/shop/sachbuch-ratgeber/21797174-9780080549279-snort-intrusion-detection-and-prevention-toolkit#drm1/

4. Запечников С.В. Информационная безопасность открытых систем: учебник для вузов. В 2-х томах / С.В. Запечников, Н.Г. Милославская, А.И. Толстой, Д.В. Ушаков. – М., 2008. – Т. II: Средства защиты в сетях. – 558 с.

5. Ушаков Д.В. Развитие принципов функционирования систем обнаружения сетевых вторжений на основе модели защищенной распределенной системы: Дисс. ... канд. техн. наук: 05.13.19 / Ушаков Д.В. – М., 2005. – 175 с.

6. Comparison of Firewall, Intrusion Prevention and Antivirus Technologies. [Электронный ресурс]. – Режим доступа к ресурсу: http://www.juniper.net/solutions/literature/white_papers /200063.pdf.

7. Intrusion Prevention Systems (IPS). [Электронный ресурс]. – Режим доступа к ресурсу: http://www.securecomputing.com/pdf/Intru-Preven-WP1-Aug03-vF.pdf.

8. Intrusion Prevention Systems (IPS). [Электронный ресурс]. – Режим доступа к ресурсу: http://hosteddocs.ittoolbox.com/ BW013004.pdf.

9. State of the Practice of Intrusion Detection Technologies. [Электронный ресурс]. – Режим доступа к ресурсу: http://www.sei.cmu.edu/pub/documents/99.reports/pdf/99tr028.pdf.

10. Wireless Intrusion Detection and Response. [Электронный ресурс]. – Режим доступа к ресурсу: http://users.ece.gatech.edu/~owen/Research/Conference%20Publications/wireless_IAW2003.pdf.

11. Anomaly Detection in IP Networks. [Электронный ресурс]. – Режим доступа к ресурсу: http://users.ece.gatech.edu/~jic/ sig03.pdf.

12. Design and Implementation of an Anomaly Detection System: an Empirical Approach. [Электронный ресурс]. – Режим доступа к ресурсу: http://luca.ntop.org/ADS.pdf.

13. Host-Based Intrusion Detection Systems. [Электронный ресурс]. – Режим доступа к ресурсу: http://staff.science.uva.nl/~delaat/snb-2004-2005/p19/report.pdf.

14. Смирнов Н.В. Курс теории вероятностей и математической статистики для технических приложений. Изд. 2. / Н.В. Смирнов, И.В. Дунин-Барковский. – М.: Наука, 1969. – 512 с.

15. Шеффе Г. Дисперсионный анализ; изд. 2: пер. с англ. / Г. Шеффе. – М.: Наука, 1980. – 512 с.

Поступила в редколлегию 3.02.2014

Рецензент: д-р техн. наук, проф. Ю.В. Стасев, Харьковский университет Воздушных Сил им. И. Кожедуба, Харьков.

ДИСПЕРСІЙНИЙ АНАЛІЗ МЕРЕЖЕВОГО ТРАФІКУ ДЛЯ ВИЯВЛЕННЯ ТА ЗАПОБІГАННЯ ВТОРГНЕНЬ В ТЕЛЕКОМУНІКАЦІЙНИХ СИСТЕМАХ І МЕРЕЖАХ

О.О. Кузнецов, О.А. Смірнов, Д.О. Даниленко

Розглядаються системи виявлення та запобігання вторгнень в сучасних телекомунікаційних системах і мережах. Досліджуються методи моніторингу подій, що складаються в аналізі мережевої активності окремих служб та інформаційних сервісів телекомунікаційних систем і мереж. Пропонується використовувати математичний апарат дисперсійного аналізу для обробки результатів моделювання телекомунікаційних систем і дослідження статистичних властивостей мережевого трафіку при визначенні значущості розбіжності чи збігу характеристик. Пропонований підхід полягає у використанні статистичного критерію Фішера, заснованого на оцінці співвідношення вибіркових дисперсій, що дозволяє з заданим рівнем значущості перевіряти гіпотезу про однорідність статистичних властивостей мережевого трафіку щодо показника розсіювання (дисперсії). Отримані результати експериментальних досліджень рекомендується використовувати для вдосконалення механізмів моніторингу мережевої активності окремих служб та інформаційних сервісів, в тому числі і для виявлення і запобігання вторгнень в телекомунікаційних системах та мережах.

***Ключові слова**: телекомунікаційні системи та мережі, система виявлення і запобігання вторгнень, дисперсійний аналіз.*

DISPERSION ANALYSIS NETWORK TRAFFIC FOR INTRUSION DETECTION AND PREVENTION IN TELECOMMUNICATION SYSTEMS AND NETWORKS

A.A. Kuznetsov, A.A. Smirnov, D.A. Danilenko

We consider systems of intrusion detection and prevention in modern telecommunications systems and networks. We study methods for monitoring events consisting in the analysis of network activity of individual services and information services of telecommunication systems and networks. Encouraged to use mathematical apparatus ANOVA processing simulation results telecommunication systems and study the statistical properties of network traffic when determining the significance of differences or matching characteristics. The proposed approach is to use a statistical test of Fisher , based on an assessment of the relationship sample variances , which allows a given level of significance test the hypothesis of homogeneity of the statistical properties of network traffic with respect to dispersion (variance). The obtained experimental results should be used for improving the mechanisms for monitoring the network activity of individual services and information services, including for intrusion detection and prevention in the telecommunication systems and networks.

***Keywords**: telecommunication systems and networks, system intrusion detection and prevention, analysis of variance.*