УДК 004.94: 004.7.056 DOI: 10.26565/2519-2310-2019-1-06

ДИСПЕРСИОННЫЙ АНАЛИЗ СЕТЕВОГО ТРАФИКА ДЛЯ ОБНАРУЖЕНИЯ ВТОРЖЕНИЙ В SMART GRIDS

Александр Кузнецов, Влада Григоренко, Андрей Дьяченко, Михаил Багмут

Харьковский национальный университет имени В.Н. Каразина, Харьков, 61022, Украина
kuznetsov@karazin.ua, mrsgriggy@gmail.com, andrey.090220@gmail.com, sapsanmiha@gmail.com

Рецензент: Владимир Максимович, д.т.н, проф., Институт компьютерных технологий, автоматики и метрологии Национального университета «Львовская политехника», Львов, 79013, Украина.
vmax@polynet.lviv.ua

Поступила: Апрель 2019.

Аннотация: *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. 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. Это 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* перспективных *Smart Grids.*

Ключевые слова: *телекоммуникационные системы и сети; системы защиты и предотвращений вторжений; дисперсионный анализ; кибербезопасность; Smart Grid.*

1 Вступ

Развитие информационных технологий содействует устойчивому росту количества пользователей различных электронных услуг и объемов обрабатываемых данных, повышению требований к скорости и надежности функционирования информационных, и компьютерных систем, а также появлению новых форм и способов реализации различных киберугроз. Особенно остро, в этом смысле, стоит вопрос обеспечения кибербезопасности современных систем энергоснабжения. Так, например, в результате лишь одной кибератаки на энергетическую систему Украины с использованием вредоносного кода типа BlackEnergy вечером 23 декабря 2015 года от сети электроснабжения было отключено 27 подстанций и 103 населенных пунктов. Другой пример нарушения функциональной безопасности связан с действием компьютерного вируса Petya.A (разновидность вируса WannaCry) летом 2017 года, в результате чего пострадали информационные системы национального банка Украины, «Укрпошти», «Укрзализниці», нескольких государственных и коммерческих банков, облэнерго, аэропортов, промышленных объектов и телерадиокомпаний. Таким образом, в условиях постоянной угрозы реализации различных кибератак, вопросы обеспечения кибербезопасности в современных информационно-коммуникационных системах чрезвычайно важны и тесно связанны с построением надежных и безопасных Smart Grids.

Проведенный анализ показал, что наибольшую уязвимость представляют методы сетевого управления, технологии доступа к электронным сервисам и услугам, а также процессы мониторинга состояния телекоммуникационных систем и сетей [1-13]. Под воздействием вредоносного программного кода отдельные коммуникационные и вычислительные компоненты могут быть несанкционированно переведены в нештатные режимы функционирования, приводящие к сбоям и нарушению установленного порядка их использования, уничтожению, искажению, блокированию, несанкционированной утечки обрабатываемой и передаваемой информации, а также к нарушению работы алгоритмов маршрутизации между узлами телекоммуникационных систем [2-4]. Очевидно, что разработка и исследование методов монито-

© Кузнецов А., Григоренко В., Дьяченко А., Багмут М., 2019

ринга сетевой активности, и совершенствование технологий обнаружения вредоносного программного кода в целях предотвращения несанкционированного воздействия на защищаемые инфокоммуникационные ресурсы, является актуальной научно-прикладной задачей. Качество решения этой задачи непосредственно влияет на обеспечиваемый уровень безопасности современных телекоммуникационных систем и применяемых информационных технологий.

В данной работе изучается возможность использования математического аппарата дисперсионного анализа для целей исследования свойств сетевого трафика различных служб и информационных сервисов при определении значимости расхождения или совпадения их характеристик. Представленные результаты могут быть использованы для совершенствования механизмов мониторинга сетевой активности, в том числе для целей обнаружения и предотвращения вторжений в телекоммуникационных системах и сетях перспективных Smart Grids.

2 Анализ исследований и публикаций

Для обеспечения безопасности в современных телекоммуникационных системах и сетях применяются различные организационно-технические мероприятия, наиболее эффективные из которых состоят в построении интегрированных систем обнаружения (Intrusion Detection System – IDS) и предотвращения (Intrusion Prevention System – IPS) вторжений [2-13]. В основе функционирования современных образцов IDS и IPS находятся процедуры сбора, накопления, анализа и обработки информации о событиях, связанных с безопасностью защищаемой телекоммуникационной системы/комплекса. На основе результатов анализа (мониторинга) сетевой активности отдельных служб и сервисов, осуществляется принятие решения о текущем состоянии ресурсов защищаемой системы с выявлением и противодействием возможному несанкционированному использованию имеющихся инфокоммуникационных ресурсов [2-6].

Под системой обнаружения вторжений (СОВ) следует понимать 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 способна нарушить безопасность компьютерной системы. К такой активности относятся сетевые атаки на наиболее уязвимые сервисы, атаки, направленные на повышение привилегий, неавторизованный доступ к критическим ресурсам, а также действия вредоносного программного обеспечения (компьютерных вирусов, троянов и червей) [2].

Под системой предотвращения вторжений (СПВ) следует понимать программную или аппаратную систему сетевой и компьютерной безопасности, обеспечивающую возможность обнаружения вторжений или нарушения установленных правил информационной безопасности и реализующую автоматическую защиту от выявленных нарушений [2-4].

Системы IPS следует рассматривать как расширение систем IDS с возможностью быстрого реагирования, путем реализации соответствующих действий по предотвращению выявленных атак или несанкционированных действий. Возможные меры предотвращения атак состоят в блокировке потоков трафика в телекоммуникационной сети, прерывании соединений, выдачи сигналов оператору и т.п. Кроме того IPS могут выполнять дефрагментацию и переупорядочивание пакетов TCP для защиты от пакетов с измененными SEQ (*номерами очереди*) и ACK (*номерами подтверждения*) [2].

Как правило, архитектура IDS включает в свой состав следующие элементы [2,3,6]:

- сенсорную подсистему (датчики), предназначенную для сбора событий, связанных с безопасностью защищаемой системы/комплекса;
- подсистему анализа, предназначенную для выявления атак и подозрительных действий на основе данных сенсоров;
- хранилище, обеспечивающее накопление первичных событий и результатов анализа;

- консоль управления, позволяющая конфигурировать IDS, отслеживать текущее состояние защищаемой системы и самой IDS, а также проводить аудит выявленных инцидентов.

В сетевой IDS (NIDS) сенсоры размещаются в наиболее важных точках сети (узлах сегментации), часто в демилитаризованной зоне, и на границе внешнего периметра сети [2,3,6,9]. Ее сенсоры прослушивают сетевой трафик в соответствующих точках и анализируют содержимое каждого пакета на присутствие вредоносных или не декларированных элементов. NIDS имеет доступ к сетевому трафику подключаясь к коммутаторам сети и отслеживает факты вторжения, проверяя сетевой трафик, ведя наблюдение сразу за несколькими узлами (хостами). Для предотвращения возможных инцидентов в состав NIDS входят соответствующее программные модули (блэйды), реализующие различные стратегии и способы парирования неавторизованных проникновений или атак защищаемых сетевых ресурсов (аппаратных и информационных).

Протокольные IDS (Protocol-based IDS, PIDS) используются для отслеживания трафика, нарушающего предусмотренные правила определенных протоколов либо синтаксис языка (например, SQL) [2,3]. PIDS представляет собой систему (либо агента), которая отслеживает и анализирует коммуникационные протоколы со связанными системами или пользователями. Так, для веб-сервера подобная СОВ обычно ведет наблюдение за HTTP и HTTPS протоколами. В случае использования HTTPS PIDS должна быть настроена таким образом, чтобы просматривать HTTPS пакеты еще до их шифрования и отправки в сеть.

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

В узловых/хостовых IDS (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 проводится анализ системных вызовов, приложений, модификаций файлов (исполняемых, файлов паролей, системных баз данных), состояния узла сети и прочих источников.

Гибридная IDS совмещает два и более подходов к реализации СОВ. При этом данные от соответствующих агентов на хостах защищаемой системы комбинируются с сетевой информацией для формирования наиболее полного аудиторского следа.

Таким образом, в пассивной IDS при обнаружении аномалии безопасности, информация о фиксируемом нарушении записывается в соответствующий log-файл (хранилище данных), а уведомления тревоги отправляются на консоль и/или администратору системы по предусмотренному каналу связи.

В активной IPS системе в ответ на фиксируемые нарушения реализуются предусмотренные защитные реакции: - сбрасывается соединение или в соответствующих сегментах сети изменяются правила работы межсетевых экранов (МСЭ) и т.п.. Причем, ответные действия могут проводиться как автоматически, так и по командам администратора системы безопасности.

С учетом специфики размещения и характера выполняемых действий IPS классифицируются как [1-13]:

- *сетевые IPS* (NIPS) - отслеживают трафик в сети и блокируют подозрительные потоки данных [2-9];
- *IPS для беспроводных сетей* (Wireless Intrusion Prevention Systems, WIPS) - анализирует сетевую активность в беспроводных сетях. В частности, обнаруживают неверно сконфигурированные точки доступа к сети, неверно сконфигурированные МСЭ в зонах взаимодействия беспроводного и проводного сетевых сегментов, атаки «человек посередине», спуфинг MAC-адресов и др. [2-10];
- *поведенческий анализ сети* (Network Behavior Analysis, NBA) анализирует сетевой трафик и идентифицирует нетипичные потоки, например DoS и DDoS атаки [2-9, 11-12];

- *IPS для отдельных узл*ов (HIPS) содержит пакет специализированных резидентных программ, обнаруживающих подозрительную активность на узле ее установки [2-9, 13].

Обзор соответствующих источников [2-13] и анализ специфики условий работы соответствующих защитных решений показал, что наиболее эффективными направлением обеспечения высокого уровня информационной безопасности, в современных телекоммуникационных системах, является интегрирование различных решений IDS и IPS. В основе функционирования подобных решений лежит комплексное использование результатов анализа сетевой активности и активная имплементация разнообразных мер противодействия актуальным угрозам [2-13]. Такой подход можно считать наиболее перспективным и сбалансированным с точки зрения обеспечения требуемых показателей безопасности [2-5]. Проведенный анализ позволяет утверждать, что в основе работы наиболее развитых СОВ и СПВ лежит использование статистических данных о циркулирующем сетевом трафике и параметрах работы сетевого оборудования. Исследование этой информации 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.

3 Методика экспериментальных исследований

Современные методы имитационного моделирования предоставляют широкие возможности по накоплению различных результатов статистических испытаний и эффективно проводить соответствующую 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 рамках данной работы предлагается использовать математический аппарат дисперсионного анализа для обработки результатов моделирования работы телекоммуникационных систем и проведения исследований свойств сетевого трафика для 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. По этим данным необходимо дать описание случайной величины X, т.е. необходимо определить ее характеристики. В практике моделирования и обработки экспериментальных данных довольно часто необходимо решать задачу подтверждения или опровержения гипотезы о принадлежности двух или более выборок одной генеральной совокупности. При этом признаки, по которым проводится сравнительная оценка, часто не являются детерминированными и 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. другими словами, дисперсия есть мера отклонения от статистического среднего.

В сложных технических системах дисперсия характеризует важные конструкционные и технологические показатели. В этом смысле при проведении исследований различных параметров технических систем наиболее важной характеристикой получаемых сравнительных оценок, является именно дисперсия, т.к. она обладает наибольшей общностью и позволяет, помимо прочего, проверять гипотезу равенстве средних значений выборок.

Таким образом, дисперсионный анализ является одним из эффективных механизмов исследования сложных технических систем и процессов, 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 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/средством M_j, а d_{ij} – случайная ошибка.

Такую модель принято называть двухфакторной схемой дисперсионного анализа [14,15], где первый фактор – измеряемая величина, а второй – собственно, 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 следующими выражениями:

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

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

Эти дисперсии удовлетворяют тождеству [14,15]: $s^2 = s^2_0 + s^2_1 + s^2_2$, которое и объясняет происхождение названия дисперсионного анализа. Так 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, или *F*-распределение), а 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 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:

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

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

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

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. 1, где также обозначены области неприемлемых значений F.

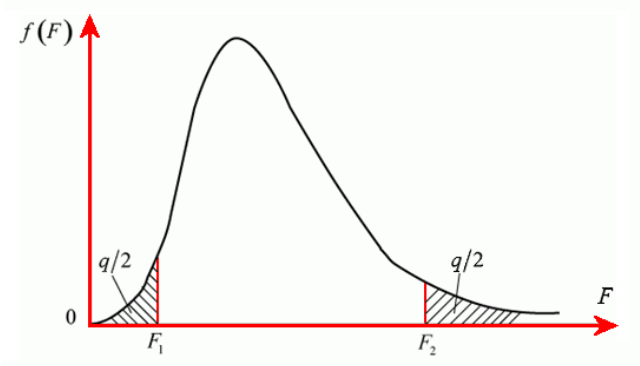

Рис. 1 - Плотность F-распределения

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.

Сформулированную задачу решаем следующим образом:

1. Для различных служб и 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 в соответствии с выражениями (1-2), рассчитываем значения выборочных средних (x^* и y^*) и дисперсий (σ_1^2 и σ_2^2).
3. Выбираем уровень значимости q, численно равный вероятности неприемлемых отклонений от принятой гипотезы и рассчитываем граничные точки F_1 и F_2 допустимых значений F.
4. Рассчитываем статистику теста F и проверяем условие $F_1 \le F \le F_2$.

5. При попадании значения F в критическую область гипотеза отвергается, а в случае непопадания – принимается.

Полученные результаты исследований позволят:

- экспериментально подтвердить или опровергнуть гипотезу об однородности свойств сетевого трафика различных служб и информационных сервисов методом дисперсионного анализа;

- обосновать практические рекомендации по организации программных и аппаратных средств мониторинга сетевой активности, обнаружения вредоносного программного обеспечения и предотвращения его воздействия на защищаемые инфокоммуникационные ресурсы.

4 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 сетевого трафика в режиме реального времени. В ходе замеров с использованием «Wireshark» оценивался объем данных, передаваемых через компьютерную сеть за определённый период времени. Измерения объема трафика 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 использовались различные телекоммуникационные службы и информационные сервисы:

- 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 обмена электронными сообщениями по распределённой (в том числе глобальной) компьютерной сети;

- Skype – программное решение, обеспечивающее текстовую, и аудиовизуальную связь посредством Интернет;

- YouTube – сервис, предоставляющий услуги видеохостинга.

Примеры гистограмм сетевого трафика при загрузке данных с сервиса YouTube (720p, бит/с), при использовании сервиса Skype в случае голосовой связи (voice) и видеосвязи (video), а также услуг электронной почты (E-mail) и протоколов HTTP и FTP приведены на рис. 2 – 7[1].

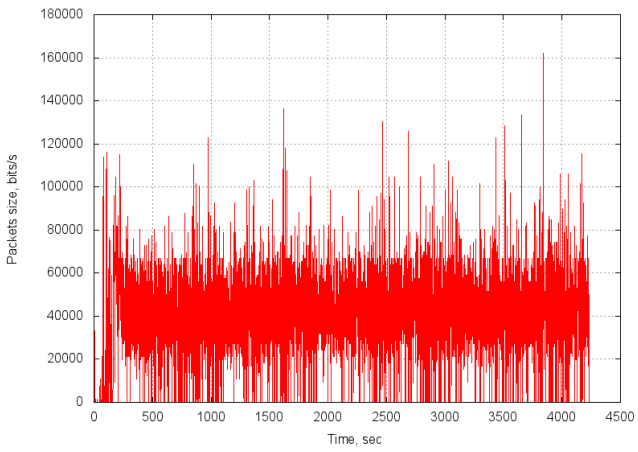

Рис. 2 - Фрагмент гистограммы сетевого трафика для YouTube

При проведении исследований использованы эмпирические данные по 100 временным отсчетам случайным образом выбранных отрезков сетевого трафика, соответствующих различным службам и сервисам. Т.е. 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, имеет распределение, близкое к нормальному [14].

[1] Сетевой трафик представлен в виде числа бит данных переданных за 1 секунду.

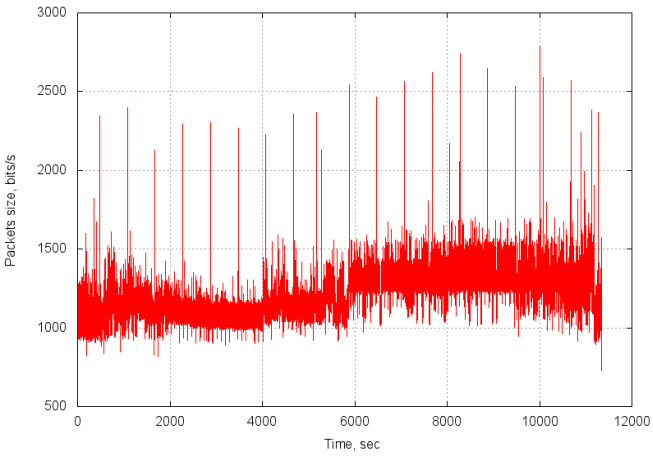

Рис. 3 - Фрагмент гистограммы сетевого трафика для Skype (voice, бит/с)

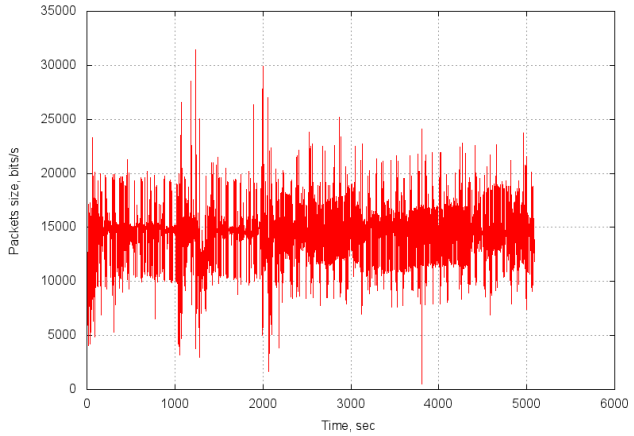

Рис. 4 - Фрагмент гистограммы сетевого трафика для Skype (video, бит/с)

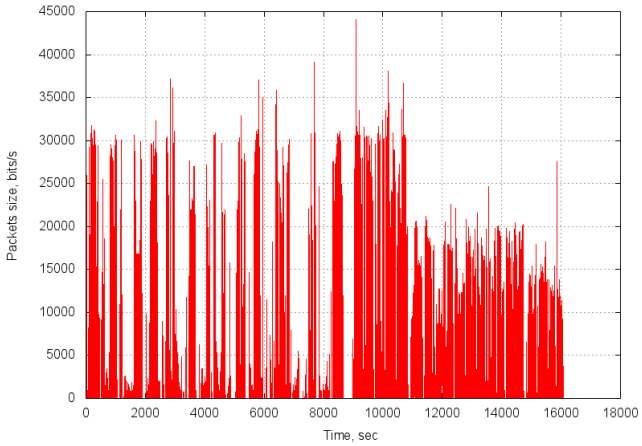

Рис. 5 - Фрагмент гистограммы сетевого трафика для E-mail (бит/с)

Так как объем данных, передаваемых через сетевую инфраструктуру 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 (вносят основной вклад в объемы, 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 моделируемой 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. Также сформулируем 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 в соответствии с которым в дальнейшем 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; полу-

ченная статистика подчинялась бы известному и рассмотренному выше закону распределения Фишера.

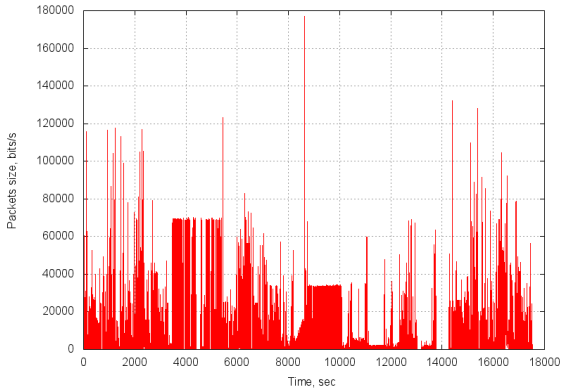

Рис. 6 - Фрагмент гистограммы сетевого трафика для HTTP (бит/с)

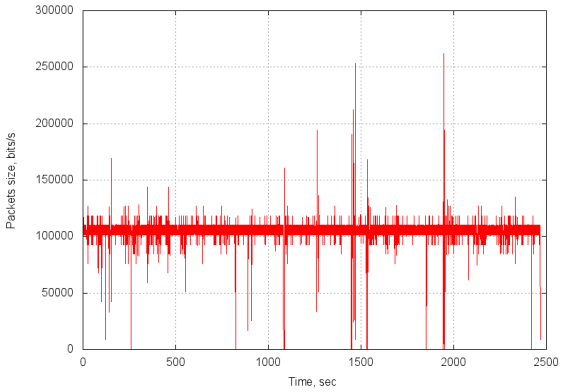

Рис. 7 - Фрагмент гистограммы сетевого трафика для FTP (бит/с)

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 этим значениям 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 отклонении (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 $k_1 = k_2 = N_2 - 1 = N_1 - 1 = 99$.

Зависимости 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. 8-9 (*использован MathCad15*).

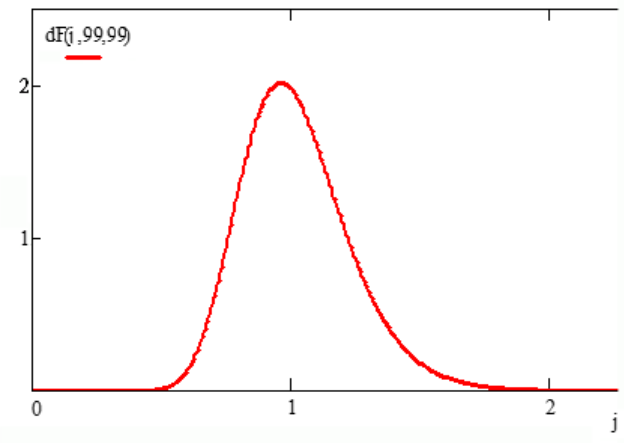

Рис. 8 - Зависимость плотности вероятностей распределения Фишера

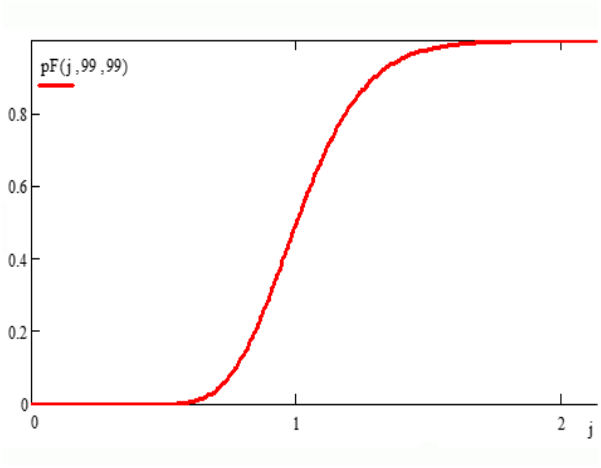

Рис. 9 - Зависимость интегрального распределения вероятностей Фишера

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. 9), график обратного кумулятивного распределения имеет вид, приведенный на рис. 10.

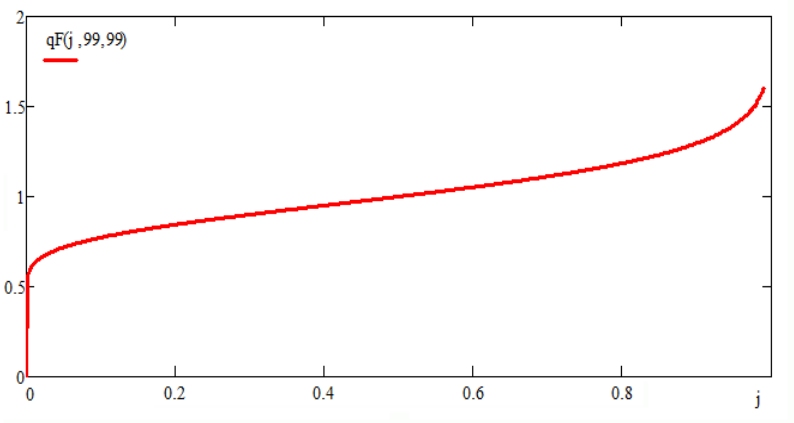

Рис. 10 - Зависимость обратного кумулятивного распределения вероятностей Фишера

Используя уровень значимости $q = 0.1$ с учетом зависимости на рис. 9, 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 и получим $F_2 = 1{,}394$, что наглядно подтверждается рис. 8–10. Таким образом, вероятность того, что значение F превысит правую граничную точку F_2 равна $q/2$:

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

Аналогично 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 наглядно демонстрируют зависимости на рис. 8–10. 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 3-м 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 же 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 различных трафиков, присущих различным службам и информационным сервисам телекоммуникационной сети. Для этого для каждого исследуемого трафика сформируем вы-

борку 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 таблицу 1.

Таблица 1 - Результаты оценки данных для сетевых трафиков различных служб и сервисов

Вид трафика (служба, сервис)	Оценка выборочной дисперсии	Оценка выборочного среднего
YouTube (720p)	2,4×108	41372,8
Skype (voice)	14560,9	1154,9
Skype (video)	7,6×106	14738,7
E-mail	116079	122,5
HTTP	2,2×108	11567,8
FTP	2,4×108	104970

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 верно определять 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 YouTube и протоколов HTTP, FTP.

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

5 Выводы

Проведенные исследования позволяют утверждать, 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. Этот факт позволяет с высокой вероятностью детектировать сетевую активность различных телекоммуникационных служб и информационных сервисов.

Избирательность сетевого детекта создает хорошие исходные условия для обнаружения несанкционированной сетевой активности и действий вредоносного программного кода, что способствует повышению общего уровня защиты инфокоммуникационных ресурсов телекоммуникационных систем и сетей.

Адаптация предложенных механизмов мониторинга сетевой активности с алгоритмами работы IDS и IPS систем является перспективным направлением для расширения возможностей систем сетевой защиты. В частности, полученные результаты могут быть полезны при построении новых механизмов обнаружения и предотвращения вторжений в перспективных Smart Grid системах.

Ссылки

[1] Cybersecurity for Smart Grid Systems URL: https://www.nist.gov/programs-projects/cybersecurity-smart-grid-systems (Last accessed: 25 June 2018)

[2] Dagle J. E. Cyber-physical system security of smart grids. 2012 IEEE PES Innovative Smart Grid Technologies (ISGT). Washington, DC, 2012. P. 1 – 2.
[3] Lightweight Stream Ciphers for Green IT Engineering / Kuznetsov O. and all. Green IT Engineering: Social, Business and Industrial Applications. Studies in Systems, Decision and Control. 2019. Vol. 171. P. 113 – 137.
[4] Christodorescu M., Jha S. Testing. Malware Detectors. Proceedings of the ACM SIGSOFT International Symposium on Software Testing and Analysis (ISSTA'04). Boston: Massachusetts, USA, 2004. 11 p.
[5] Prospective Lightweight Block Cipher for Green IT Engineering /Andrushkevych A. and all. Green IT Engineering: Social, Business and Industrial Applications. Studies in Systems, Decision and Control. 2019. Vol. 171. P. 95 – 112.
[6] Methods of Information Protection in Communications Systems and Methods of Their Cryptoanalysis/ Gorbenko I.D., Dolgov V.I., Rublinetskii V.I., Korovkin K.V. Telecommunications and Radio Engineering. 1998. Vol. 52, Issue 4. P. 89 – 96.
[7] OpenStack-Based Evaluation Framework for Smart Grid Cyber Security / Albarakati A. and all. 2018 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm). Aalborg, 2018. P. 1 – 6.
[8] Jahan S., Habiba R. An analysis of smart grid communication infrastructure & cyber security in smart grid. 2015 International Conference on Advances in Electrical Engineering (ICAEE). Dhaka, 2015. P. 190 – 193.
[9] Smart grid information security - a research on standards/ Wang Y., Zhang B., Lin W., Zhang T. 2011 International Conference on Advanced Power System Automation and Protection. Beijing, 2011. P. 1188 – 1194.
[10] Security and Reliability Perspectives in Cyber-Physical Smart Grids / Lei H., Chen B., Butler-Purry K. L., Singh C. 2018 IEEE Innovative Smart Grid Technologies - Asia (ISGT Asia). Singapore, 2018. P. 42 – 47.
[11] Smart grid information security - a research on standards / Wang Y., Zhang B., Lin W., Zhang T. 2011 International Conference on Advanced Power System Automation and Protection. Beijing, 2011. P. 1188 – 1194.
[12] Impact of cyber-security issues on Smart Grid / Yang Y and all. 2011 2nd IEEE PES International Conference and Exhibition on Innovative Smart Grid Technologies. Manchester, 2011. P. 1 – 7.
[13] Zamula A., Kavun S. Complex systems modeling with intelligent control elements. International Journal of Modeling, Simulation, and Scientific Computing. 2017. Vol. 8, № 1. [19 pages]
[14] Zetter K. The Ukrainian Power Grid Was Hacked Again. Vice Motherboard. URL: https://motherboard.vice.com (Last accessed: January 10, 2017)
[15] Lipovsky R. New wave of cyberattacks against Ukrainian power industry. We Live Security. URL: https://www.welivesecurity.com (Last accessed: 20 February 2016)

Reviewer: Volodymyr Maxymovych, Doctor of Sciences (Eng.), Full Prof., Institute of Computer Technologies, Automation and Metrology, Lviv Polytechnic National University, 79013, Ukraine.
E-mail: vmax@polynet.lviv.ua

Received on Aprilr 2019.

Authors:

Alexandr Kuznetsov, Doctor of Sciences (Eng.), Full Prof., Academician of the Academy of Applied Radioelectronics Sciences, V.N.Karazin Kharkiv National University, Kharkov, Ukraine.
E-mail: kuznetsov@karazin.ua

Vlada Hryhorenko, computer science student, V.N. Karazin Kharkiv National University, Ukraine.
E-mail: wdpgames@yandex.ru

Andrii Diachenko, computer science student, V.N. Karazin Kharkiv National University, Ukraine.
E-mail: andrey.090220@gmail.com

Mykhaylo Bagmut, postgraduate, V.N. Karazin Kharkiv National University, Ukraine.
E-mail: sapsanmiha@gmail.com

Dispersion analysis of network traffic for intrusion detection in Smart Grids.

Abstract. We consider the systems of detection and prevention of intrusions in modern telecommunication systems and networks. Methods of monitoring events, consisting of analysis of network activity of individual services and information services, are analyzed. It is proposed to use the mathematical apparatus of analysis of variance for processing results of modeling telecommunication systems and studying the statistical properties of network traffic in determining the significance of discrepancies or coincidence of characteristics. The proposed approach is to use the Fisher statistical criterion based on an estimate of the ratio of sample variances. This allows you to test the hypothesis about the homogeneity of statistical properties of network traffic with respect to the variance index (variance) with a given level of significance. The obtained results of experimental studies are recommended to be used to improve mechanisms for monitoring the network activity of individual services and information services, including for detecting and preventing intrusions in telecommunications systems and networks of promising Smart Grids.

Keywords: Telecommunication Systems and Networks; Intrusion Detection and Prevention System; Analysis of Variance; Cyber Security; Smart Grid.

Рецензент: Володимир Максимович, д.т.н., проф., Інститут комп'ютерних технологій, автоматики та метрології Національного університету «Львівська політехніка», Львів, 79013, Україна.
E-mail: vmax@polynet.lviv.ua

Поступила: Апрель 2019.

Автори:

Олександр Кузнецов, д.т.н., проф., академік Академії наук прикладної радіоелектроніки, Харківський національний університет імені В.Н. Каразіна, Харків, Україна.
E-mail: kuznetsov@karazin.ua

Влада Григоренко, студентка факультету комп'ютерних наук, Харківський національний університет ім. В.Н. Каразіна, Харків, Україна.
E-mail: mrsgriggy@gmail.com

Андрій Д'яченко, студент факультету комп'ютерних наук**,** Харківський національний університет ім. В.Н. Каразіна, Харків, Україна
E-mail: andrey.090220@gmail.com

Михайло Багмут, аспірант, Харківський національний університет ім. В.Н. Каразіна, Харків, Україна.
E-mail: sapsanmiha@gmail.com

Дисперсійний аналіз мережевого трафіку для виявлення вторгнень в Smart Grids.

Анотація. Розглядаються системи для виявлення та запобігання вторгнень у сучасних телекомунікаційних системах і мережах. Досліджуються методи моніторингу подій, що базуються на аналізі мережевої активності окремих служб та інформаційних сервісів. Пропонується використовувати математичний апарат дисперсійного аналізу для обробки результатів моделювання телекомунікаційних систем і дослідження статистичних властивостей мережевого трафіку при визначенні значущості розбіжності або збігу характеристик. Пропонований підхід полягає у використанні статистичного критерію Фішера, заснованого на оцінці відношень вибіркових дисперсій. Це 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 для виявлення і запобігання вторгнень у телекомунікаційних системах та мережах перспективних Smart Grids.

Ключові слова: телекомунікаційні системи та мережі; системи для виявлення та запобігіння вторгнень; дисперсійний аналіз; кібербезпека; Smart Grids.