УДК 004.94

Т.В. ГЛАДКИХ, Т. ГНОТ, В. СОЛЬСКИЙ
SoftServe, Украина

ОБНАРУЖЕНИЕ АНОМАЛИЙ – ОБУЧЕНИЕ БЕЗ УЧИТЕЛЯ

Предложено комплексное решение, позволяющее выявлять ситуации, в которых сетевая активность пользователей или сетевого оборудования отличается от ожидаемого шаблона, что может служить индикатором нарушения информационной безопасности. Получен ансамбль моделей, который дает возможность детектировать как типичные проявления аномальной сетевой активности, так и новые элементы сетевых аномалий. Модели, включенные в ансамбль, обладают способностью самонастройки, в связи с чем, могут легко адаптироваться к изменениям в наблюдаемых процессах.

Ключевые слова: обнаружение аномалий, ассоциативные правила, кластеризация, временной ряд, сетевая активность.

T.V. GLADKIKH, T. HNOT, V. SOLSKIY
SoftServe, Ukraine

ANOMALY DETECTION – UNSUPERVISED APPROACH

Abstract – This article proposing the complex solution that allows to detect activity of user or network equipment, which differ from expected behavior pattern, and can be considered as an indicator of information security violation. Final solution is represented as an models ensemble, which can be used for discovering both typical network activity anomalies and new elements of anomaly. Models, which are the parts of ensemble, have a self-turning capability, so they can be easily adapted to changes in analyzed processes.

Keywords: anomaly detection, association rules, clustering, time series, network activity.

Введение

Актуальность решения проблемы обнаружения аномалий заключается в том, что любое отклонение от некой общей картины, отображающей реальное состояние системы, может нести важную информацию о проблеме. Игнорирование отклонений может иметь весьма плачевный исход [1, 2]. Так, к примеру, нехарактерное для среднестатистического рентгеновского снимка затемнение может служить фактом злокачественного новообразования.

Особое место среди широкого диапазона предметных областей, в которых фигурируют вопросы обнаружения изменения шаблона, занимает Информационная Безопасность [3]. Мы живем в эпоху так называемого информационного взрыва, что характеризуется непрерывным развитием и усложнением средств и методов автоматизации процессов обработки информации. Это, в свою очередь, приводит к повышению зависимости общества от степени безопасности используемых им информационных технологий.

Если рассмотреть характер информационной среды организации, то применение традиционных мероприятий по обеспечению безопасности, которые носят, скорее, запретительный характер, зачастую оказывается неэффективным [4]. В виду невероятно большого многообразия возможных сценариев работы пользователей в системе безопасности появляется масса исключений из основных правил, что снижает возможности превентивной защиты и усложняет стандартный анализ выявления инсайдерских угроз. Обнаружение внешних атак также становится более проблематичным, в связи с тем, что «атакующий» осведомлен о стандартных применяемых средствах детектирования вторжений и может применять «маскирующие» средства для атаки. Процессы, связанные с нарушением сетевой безопасности, могут проявить себя, например, в повышении активности по определенным портам, появлении сервисов, нехарактерных для пользователя, изменение характера работы пользователя с сетевыми ресурсами и т.д.

Одним из возможных путей решения этой проблемы – разработка систем, позволяющих выявлять случаи нехарактерного поведения пользователей и сетевого оборудования на основании анализа логов сетевой активности. Применение методов интеллектуального анализа данных, которые лежат в основе подобного рода систем, позволяют выявить характерные шаблоны поведения как отдельных, так и групп пользователей, и сделать выводы относительно поведения, которое отличается от «общепринятого». Такие системы могут обладать свойством самоадаптации, что позволяет минимизировать участие человека в настройке параметров системы. В виду нетривиальности сформулированной задачи в общем виде, без учета специфики работы той или иной организации, она представляет большой интерес для специалистов в области машинного обучения и интеллектуального анализа данных.

Постановка задачи – обнаружение аномалий

Согласно определению, ***аномалия (греч)*** *– это отступление или уклонение от правила, поэтому аномальным называют все отступающее или уклоняющееся от правильного или нормального*.

Таким образом любое нарушение стандартного поведения, которое прослеживаются на исторических данных, может быть интерпретировано как аномалия. При этом факт нарушения некоторого шаблона может быть как известен заранее, так и быть установлен в результате анализа. В этом контексте любая задача, связанная с обнаружением нестандартного поведения, сводится к поиску некоего базового состояния (baseline) и классификации каждого события, зафиксированного в системе как соответствующего

или противоречащего найденному прототипу. Если говорить об информационной безопасности предприятия, то решение этой задачи может быть выполнено путем анализа сетевой активности пользователей и оборудования с целью обнаружения нехарактерного поведения во внешнем и внутреннем сетевом трафике (анализ логов Netflow) [5, 6]. Это нехарактерное поведение и может служить сигналом как об инсайдерской деятельности, так и о попытке внешней атаки на ресурс.

В этом контексте мы можем интерпретировать ***аномалию*** как *любое событие, которое может быть оценено как статистически невозможное, в соответствии с результатами анализа протокола сетевой активности*. Типы выявляемых аномалий:

1. Существенное отклонение наблюдаемого значения от некоторой ожидаемой величины
2. Нарушение в процессе, который отражает изменение измеряемого параметра в пределах области наблюдения
3. Нехарактерная для пользователя или группы пользователей совокупность наблюдаемых значений измеряемых параметров

Обзор существующих методов

В большинстве современных системах обнаружения вторжений [7] используют ***сигнатурные методы*** анализа, которые направлены на обнаружение известных или точно описанных атак, что крайне затрудняет их применение для случая ранее неизвестных атак, либо модификации существующих.

Поэтому в последнее время, все больше внимания уделяется разработке новых методов обнаружения сетевых аномалий, являющихся следствием технических сбоев или несанкционированного воздействия. Существующие исследования в этой области связаны с применением методов мультифрактального [8] и кратномасштабного вейвлет-анализа [9 – 11]

В основе этих методов лежит оценка фрактальной размерности временного ряда с целью определения условий нарушения его самоподобия. Это делается через оценку значимости отклонения фрактальной размерности в соседних отсчетах, что позволяет оценить близость соседних фрагментов временного ряда и отреагировать в случае сильных расхождений. При этом открытым остается вопрос оценки подобных ситуаций при их повторении. Более того, как показывают эксперименты, показатель Херста временного ряда может сильно меняться при смене вида деятельности пользователя в широких пределах, что не позволяет сделать заключение о наличии или отсутствии самоподобия и делает невозможным обнаруживать аномалии, применяя для анализа только этот метод.

Отдельное внимание уделяется применению иммунных [12, 13] и нейросетевых алгоритмов [14]. Первая категория алгоритмов позволяет детектировать ранее неизвестные аномалии и нетребовательны к вычислительным ресурсам, однако, в качестве недостатков выделяют высокий процент ложных срабатываний [15]. Сложность применения алгоритмов второй категории заключается в необходимости разработки нейронных сетей, имеющих способность к дообучению в режиме реального времени [16]. Наиболее перспективным направлением исследования в данной области является разработка адаптивных и робастных систем, которые позволяют идентифицировать любые случаи несоответствия наблюдаемого процесса ранее установленным закономерностям. В работе [17] приводится описание двух алгоритмов, которые позволяют идентифицировать аномалии, при отсутствии априорной информации о нормальном процессе. Метод LERAD позволяет установить значимые ассоциативные связи между отдельными сущностями, которые описывают анализируемый процесс, второй метод (CLAD) основан на кластеризации результатов наблюдений над процессом, выраженных в виде характеристического вектора, с целью обнаружения наблюдений, которые можно интерпретировать как выбросы в n-мерном пространстве. Описанные методы позволяют идентифицировать только второй тип аномалий, что не позволяет говорить о комплексном решении описанной проблемы.

В данной работе мы предлагаем к рассмотрению систему обнаружения аномалий, позволяющую идентифицировать аномалии, связанные с нехарактерными изменениями временного ряда, который описывает поведение измеряемой величины, так и с нехарактерными совокупностями наблюдений нескольких измеряемых параметров. Такого рода решение дает возможность абстрагироваться от специфики предметной области и служит основой для своевременное реагирование в тех случаях, когда шаблоны поведения еще недостаточно изучены.

Обнаружение аномалий через ансамбль моделей

Интерпретация сетевого протокола при использовании стандартных Netflow логов является недоступной. Это значит, что мы не располагаем информацией о том, какие именно зафиксированные события могут быть описаны как норма, а какие, соответственно, как отклонения от нее. Таким образом, ***система комплексного анализа протокола сетевой активности*** для идентификации аномалий должна обладать такими возможностям как сбор и хранение данных о результатах сетевой активности; представление сетевой активности в виде ряда числовых характеристик, дополненных нечисловыми атрибутами (уточняющие факторы), включая маркеры времени; выявление скрытых закономерностей в данных, дающих основу для формирования шаблона поведения; оценка новых наблюдений на предмет соответствия шаблону. В общем виде структура системы может быть представлена на рис. 1. Ядром предложенной системы является ансамбль моделей, каждая из которых позволяет оценить среднестатистическую активность пользователя (или группы пользователей) и классифицировать наблюдения, зафиксированные в системе, как соответствующие норме или аномальные. В качестве

источника данных для построения моделей использовались следующие метрики и категории измерения (табл. 1).

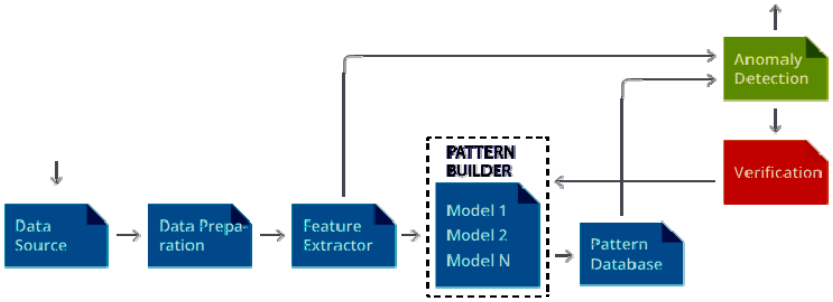


Рис. 1 Обнаружение аномалий – инфраструктура

В соответствии с описанными выше типами аномалий, мы попытались применить совместно три модели. С одной стороны, оценить характер изменения метрики в пределах той или иной категории, рассматривая накопленные данные как процесс, протяженный во времени.

Табл. 1

Метрика	Категория
Суммарный объем переданных данных за единицу времени Суммарное число запросов за единицу времени	Тип протокола Порт источника Порт приемника Группа IP адресов

С другой стороны, совместно анализировать значение метрик в различных категориях для изолированного периода времени.

1. ***Dynamic Threshold Model*** - модель позволяет установить допустимый уровень нагрузки (измеряемой величины) для определенных временных интервалов (речь может идти о времени суток, днях недели, месяцев и т.д.). В последствии каждое наблюдаемое значение оценивается на соответствие установленному порогу, что позволяет идентифицировать случаи появления аномалий. Решение, которое может быть использовано для обнаружения аномалий первого типа.

2. ***Association Rules Based Approach*** – модель позволяет описать активность в сети как совокупность связанных между собой событий, и, как следствие, представить ее в виде стохастического процесса. Любое событие, характеризующееся малой вероятностью, может быть интерпретировано как аномалия.

3. ***Time Series Clustering*** – модель позволяет найти общие закономерности в структуре временного ряда и тем самым фиксировать любые отклонения от установленного шаблона – аномалии второго типа.

Модель на основе динамических пороговых значений

В основе этого подхода лежит представление активности пользователя, выраженное в виде совокупности значений измеряемого параметра (метрики) для определенной категории наблюдений, через временной ряд (Time Series). Такое представление активности пользователей позволяет установить факт наличия или отсутствия общих тенденций, которые могут быть выражены в виде некоего общего тренда или же иметь сезонный (периодический) характер [18].

Так, например, можно рассмотреть временной ряд, который отражает недельную сетевую активность группы пользователей через суммарный объем переданных данных за 1 минуту (рис. 2). Нетрудно видеть, что в пределах этого временного участка прослеживается некая общая тенденция – в течении недели средняя активность пользователей уменьшается. В то же время процесс нельзя назвать равномерным – присутствует некоторая периодическая последовательность, связанная с тем, что максимальная и минимальная активность приходится на середину рабочего дня сотрудников и ночное время, соответственно. Это дает возможность, утверждать, что активность в течении каждого дня может быть описана некоторым общим шаблоном с небольшими поправками на основной тренд. Таким образом, мы можем описать активность пользователя в виде дискретного временного ряда (сигнал квантуется как по времени итак и по амплитуде), оценив при этом возможную ошибку дискретизации (допустимый порог отклонения). В том случае, если ошибка замены непрерывного сигнала дискретным превысит допустимый порог отклонения, ситуация может быть расценена как некая аномалия. В виду всего вышеизложенного, задача сводится к тому, чтобы установить максимальный порог сетевой нагрузки для каждого участка временного ряда, на основании среднестатистической активности пользователя с учетом изменения характера нагрузки, отражаемого в трендовой и сезонной компонентах ряда.

Поскольку сезонная декомпозиция временного ряда должна позволить определить наиболее характерные особенности рассматриваемого процесса, ее следует выполнять над рядом, который уже очищен от шума, и как следствие, избавлен от каких-либо проявлений аномального поведения. Для этого мы, прежде всего, выполняем реконструкцию временного ряда через сингулярное спектральное разложение

траекторной матрицы временного ряда с последующим получением его тренда и значимых компонентов. Сезонная декомпозиция восстановленного временного ряда выполняется методом на основе теста Chi2 Пирсона [19]. В результате сезонной декомпозиции появляется возможность разбить весь временной ряд на фрагменты, характеризующиеся общими паттернами изменения сигнала. При этом мы можем ожидать, что найденные закономерности будут повторяться при последующих наблюдениях. Упомянутые выше паттерны могут быть описаны через обобщение изменения сигнала для каждого выделенного временного участка.

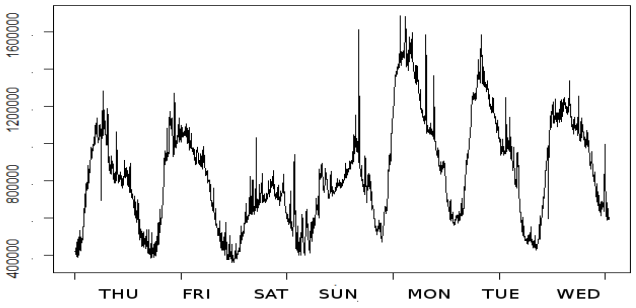


Рис. 2. Пример временного ряда

Пусть в результате сезонной декомпозиции был установлен некоторый период колебаний p^*. Тогда $\left\lfloor \frac{T}{p^*} \right\rfloor$ – число интервалов; $x_{t_i} = x_t, t = \overline{\left\lfloor (1+(i-1)\cdot p^*), i\cdot p^* \right\rfloor, i = 1, \left\lfloor \frac{T}{p^*} \right\rfloor}$ – i-я компонента временного ряда, определяемая периодом колебаний p^*; $x_{reconst.t_i} = x_{reconst.t}, t = \overline{\left\lfloor (1+(i-1)\cdot p^*), i\cdot p^* \right\rfloor}$ – i-я компонента реконструированного временного ряда.

На каждом участке длины p^* выполняется поиск абсцисс локальных экстремумов восстановленного временного ряда $\left\{Ext_q^{\min}, Ext_w^{\max} \mid q = \overline{1, n_{extr}^-}; w = \overline{1, n_{extr}^+}\right\}$:

$$Ext_{qi}^{\min} = \left\{ t, \frac{\partial x_{reconst.t_i}}{\partial t} = 0; \frac{\partial^2 x_{reconst.t_i}}{\partial t^2} > 0 \right\}; \; Ext_{wi}^{\max} = \left\{ t, \frac{\partial x_{reconst.t_i}}{\partial t} = 0; \frac{\partial^2 x_{reconst.t_i}}{\partial t^2} < 0 \right\};$$

где $Ext_q^{\min}$ – q-й локальный минимум; $Ext_w^{\max}$ – w-й локальный максимум

Найденные абсциссы экстремумов формируют множество отсчетов для разбиения фрагмента на участки, в пределах каждого из которых, изменениями ожидаемой нагрузки можно пренебречь. Искомые точки отсчетов представляют собой абсциссы локальных минимумов $Ext_{qi}^{\min}$ и срединные точки участков, ограниченных последовательный парой локальных экстремумов $\left\{\left(Ext_{qi}^{\min} + Ext_w^{\max}\right)/2, \left(Ext_{(q+1)i}^{\min} + Ext_w^{\max}\right)/2\right\}$:

Для каждого временного участка выполняется оценка допустимого порогового значения:

$$Threshold_{ij} = \max\left(x_{reconst.t\,ij} + z(\alpha_{pr})\cdot RMSE_{ij}\right),$$

где $RMSE_{ij} = \frac{\sum_{t=1}^{T_{ij}} \left(x_{t\,ij} - x_{reconst.t\,ij}\right)^2}{T_{ij}}$ – среднеквадратическая ошибка восстановления временного ряда; T_{ij} – число элементов j-го временного интервала; $z(\alpha_{pr})$ – z-статистика при α_{pr} уровне значимости;

Таким образом, пороговое значение каждого интервала определяется максимальным значением интервала прогноза. В результате, для каждого минимального периода p^* получаем следующую функцию:

$$Threshold_{it}, i = \overline{1, \left\lfloor \frac{T}{p^*} \right\rfloor}, t = \overline{1, p^*}$$

Ее визуализация для фрагмента временного ряда показана на рис. 3, *а*. Красными точками обозначены значения временного ряда, которые выходят за установленные пределы и, как следствие, могут расцениваться как аномальные наблюдения. На следующем шаге осуществляется попытка получить обобщенное представление об ожидаемой активности пользователя, так называемую «среднюю активность». Она пользователя как результат интеграции пороговых значений для каждого временного участка (рис. 3, *б*):

$$Agg\_Threshold_{ts} = \bigcup_{t \in ts} Agg\_Threshold_t; \; Agg\_Threshold_t = quantile_{significance\,level}\left\{ Threshold_{it}, i = \overline{1, \left\lfloor \frac{T}{p^*} \right\rfloor} \right\}$$

Как видно из рис. 3, *б*, найденные паттерны отражают общую специфику изменения активности пользователя в пределах установленного периода, но при этом игнорируется тот факт, что при сохранении общего шаблона поведения, средняя нагрузка на участках может не быть постоянной, а в некоторых случаях может существенно отличаться.

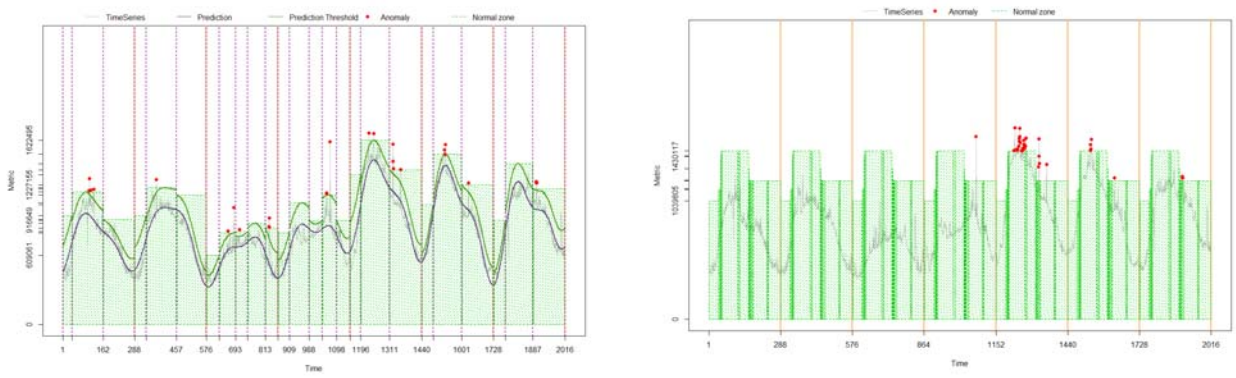
$$Level_i = \frac{Load\{x_{ti}\}}{Load\{x_t\}}, i = \overline{1, \left\lfloor \frac{T}{\|P\|} \right\rfloor}$$

Полученные коэффициенты используются для коррекции пороговых значений.

$$Threshold_i = Agg\_Threshold_{ts} \cdot Level_i, ts = \overline{1, N_{timeslots}}, i = 1, \left\lfloor \frac{T}{p^*} \right\rfloor$$

Таким образом, получаем обобщенное представление анализируемого процесса с учетом средней нагрузки для каждого характерного участка временной шкалы. Нетрудно видеть, что откорректированные пороговые значения позволяют подстроиться под особенности временного ряда, и, тем самым маркировать только те наблюдения которые существенно выбиваются из общей картины.

Разработанная модель позволяет описать активность пользователя в сети через ограниченный набор констант, который может быть получен на первоначальном этапе настройки системы, и не требует для своего функционирования анализа всего набора исторических данных «на лету».



а **б**

Рис. 3 Пороговые значения для каждого фрагмента временного ряда:
***а*) Первоначальный анализ *б*) Средняя нагрузка**

Кроме того, модель может легко адаптироваться под изменяющийся процесс – при увеличении частоты появления случаев существенных отклонений дискретного временного ряда от непрерывного, достаточно откорректировать пороговые значения и/или коэффициенты относительной нагрузки.

Модель на основе ассоциативных правил

Как было сказано выше, процесс, над которым ведется наблюдение можно представить в виде совокупности связанных между собой событий. Если говорить об активности пользователя и/или группы пользователей в сети, то в качестве такого события можно рассматривать одномоментное значение отдельной метрики (табл. 1), вычисленной для категории и/или комбинации значений нескольких категориальных переменных. Так, например, для шага дискретизации по времени 1 минута, можно оценить *суммарное число пакетов, переданных отделом маркетинга за 1 минуту по протоколу TCP/IP в первой половине рабочего дня* и одновременно с этим *оценить общий объем пакетов по UDP протоколу для той же группы пользователей*.

При этом нас интересует не столько абсолютное значение по каждой из указанных категорий, сколько относительное число случаев, когда такая активность была зафиксирована.

Возьмем, к примеру, некоторый временной ряд, который отражает изменение интересующего наблюдателя параметра во времени (рис. 4). В большинстве случаев такого измерения будут представлены некоторой непрерывной величиной. Это означает, что она принимает бесконечное число значений – каждое будет уникальным в пределах шкалы квантования, что не дает возможность оценить вклад каждого отдельного наблюдения в формирование общего шаблона поведения.

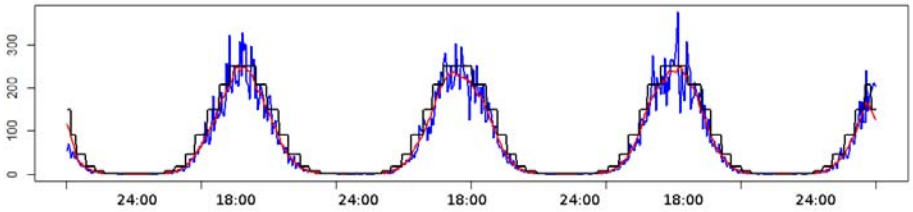


Рис. 4 Пример временного ряда, отражающего изменение измеряемого параметра во времени

Однако, нетрудно видеть, что измеренная величина так или иначе находится в пределах некоторых дискретных диапазонов, последовательность которых прослеживается во времени. Это позволяет через квантование перейти от непрерывного представления к дискретному – ограничить число значений измеряемой величины некоторым конечным множеством, разбив весь диапазон значений на участки равной плотности:

Получаем дискретное представление временного ряда:

$$M_1(t) = quantile\left(Sig, (k+1)/n\right)_{Sig(t) \in [quantile(Sig,\, k/n),\, quantile(Sig,\, (k+1)/n)]} \quad (3)$$

где Sig – исходный непрерывный сигнал; M_1 – дискретное представление исходного сигнала

Фактически мы представляем временной ряд в виде случайной величины, которая имеет дискретное равномерное распределение с энтропией $H = \log_2(1/q)$, где q – число участков равномерной плотности (для $q = 5$, $E = 2.32$). Такое описание процесса само по себе не может служить источником информации для обнаружения аномалий, поскольку ни одно из значений анализируемой метрики не может быть признано «практически невозможным», то есть каждое значение так или иначе укладывается в некоторый паттерн. Уменьшить энтропию системы можно, если принять во внимание временной фактор (t). Так, нетрудно видеть, что начало и конец рабочего дня характеризуются значениями дискретной с.в. из множества {6, 18, 46}, в то время как диапазон с 12:00 до 18:00 характеризуется большими значениями измеряемой величины (табл. 2)

Видно, что энтропия системы уменьшается – переменная t уточняет значение M_1. Эта зависимость отражается в виде взаимной информации:

$$I(M_1,T) = \frac{\log_2(1/q) - \sum_{j=1}^{m} P(t_j) \times H(M_1 | T = t_j)}{\min(H(M_1), H(T))}; 0 \le I(M_1,T) \le 1 \quad (4)$$

Табл. 2

	0.2	0.5	2	6	18	46	92	150	208	250
0-6	0,42	0,17	0,17	0,17	0,07					
6-9					0,3	0,35	0,35			
9-12							0,14	0,43	0,38	0,05
12-18							0,1	0,2	0,2	0,5
18-21					0,35	0,45	0,2			
21-24	0,14	0,43	0,38	0,05						

Чем выше взаимная информация, тем более детерминированным будет анализируемый процесс и тем более однозначно можно описать некоторый шаблон поведения пользователя. В этом случае, значения дискретной случайной величины (анализируемого параметра), которые характеризуются малой возможностью появления в пределах того или иного участка временной оси (тайм-слота), можно расценивать как выбивающиеся из общего шаблона, и как следствие – аномалии. Таким образом, мы можем ожидать, что в среднем поведение системы будет подчинено описанному выше стохастическому закону – для любого наблюдаемого состояния системы должна быть оценена возможность его появления и в том случае, если оно относится к категории маловероятных, то можно говорить об отклонении от ожидаемого поведения и, как следствие, об аномалии.

Такого рода интерпретация процесса позволяет анализировать одновременно значения измеряемого параметра для различных категорий. Если на основании опытных данных будет установлено, что значение одной из категорий может уточнить значение метрики по другой, то есть существуют часто встречающиеся комбинации значений по обеим категориям, то эта информация позволит усилить значимость определенного паттерна – чем больше категорий объединяет часто встречающаяся комбинация – тем выше вероятность обнаружения найденного шаблона в наблюдаемом процессе и тем более значимым будет отклонение от него.

Пусть, к примеру, по результатам наблюдения в течении одного месяца была зафиксирована следующая активность:

Каждая квантованная переменная подчинена дискретному равномерному распределению, но при рассмотрении системы зависимых случайных величин (M_1, M_2) получаем совокупность разреженных матриц, которые дают представление о наиболее ожидаемых комбинациях значений, которые можно интерпретировать как паттерн (табл. 3).

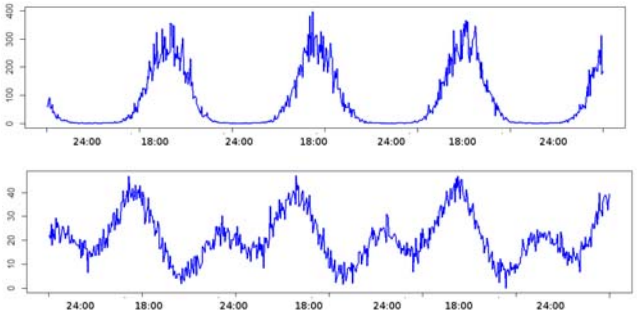


Рис. 5 Пример сетевой активности пользователя, выраженная в двух метриках

Табл.3

	<14	[14, 18)	[18, 22)	[22, 30)	>30
<0.6	0,00	0,62	0,34	0,03	0,00
[0.6, 7)	0,00	0,22	0,56	0,22	0,00
[7, 46)	0,38	0,14	0,00	0,17	0,31
[46, 154)	0,44	0,00	0,00	0,00	0,56
>154	0,32	0,08	0,11	0,24	0,26

Включение параметра времени позволяет получить еще более строгое представление о возможности появления той или иной комбинации параметров. По-другому зависимость (табл. 3) можно представить в виде графа, где вершины представляют собой случайные события (состояния системы), каждое из которых соответствует значению отдельной метрики или комбинации нескольких значений, а веса связей отражают их условные вероятности (рис. 6а). Как видно из графа, поведение системы описывается совокупностью паттернов, которые могут быть обнаружены в каждом конкретном наблюдении (состоянии системы). Граф этот может быть упрощен за счет нивелирования малых весов связей и усечения путей с малым суммарным весом (рис. 6б). В результате мы получаем совокупность, так называемых, правил, которые определяют нормальное поведение системы. Таким образом, задача получения паттернов в сетевой активности пользователя сводится к задаче обнаружения ассоциативных связей между отдельными состояниями описанной выше системы, которые формируют ассоциативные правила [20], описывающие искомые шаблоны. Такой подход дает возможность использовать для воссоздания среднестатистической активности пользователя известные методы поиска ассоциативных правил, в частности, алгоритм Apriory. Не вдаваясь в основы самого алгоритма, покажем его применение в контексте поставленной задачи:

1. Измерения по каждой категории, которые описывают сетевую активность, представляются в виде дискретного сигнала, что дает возможность описывать их в виде системы зависимых дискретных случайных величин, равномерно распределённых в пределах всего временного диапазона (3).

2. Для каждого момента времени t получаем список всех транзакций (отдельных значений каждой случайной величины и систем различных комбинаций случайных величин), которые соответствуют данному наблюдению:

$$\mathbf{T}(t)=\{T_i(t)\}=\{M_k(t),(M_k(t),M_r(t)),\ (M_k(t),M_r(t),M_u(t)),\cdots(M_1(t),M_2(t),\cdots M_N(t))\},$$

где $k,r,u=\overline{1,N},\ k\neq u\neq r$

3. Формируем общий список транзакций для всех обработанных наблюдений:

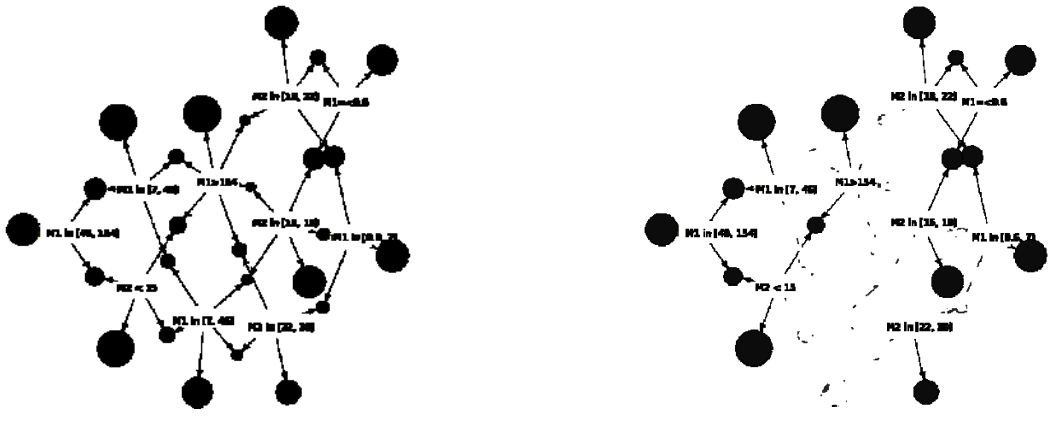
$$\mathbf{T}=\bigcup\mathbf{T}(t)=\{M_k,(M_k,M_r),\ (M_k,M_r,M_u),\cdots(M_1,M_2,\cdots M_M)\}$$

4. Для всех многокомпонентных транзакций получаем список ассоциативных правил, которые удовлетворяют требованиям минимальной поддержки и достоверности:

$$\mathbf{A}=\left\{A_{kr}\middle|\left(\sup(A_{kr})\geq T_{\sup}\right)\cap\left(\mathrm{conf}(A_{kr})\geq T_{\mathrm{conf}}\right)\right\}=\left\{\left(T_{k,k\neq r}\Rightarrow T_r\right)\right\};\quad r,k=\|\mathbf{T}\|$$

$$\sup(A_{kr})=P(T_k)\times P\left(T_r\middle|T_{k,k\neq r}\right)=P\left(\prod_{M_k\in(T_r\cap T_k)}M_k\right);\ \mathrm{conf}(A_{kr})=P\left(T_r\middle|T_{k,k\neq r}\right)=\frac{P(T_r\cap T_k)}{P(T_k)}=\sup(A_j)\Big/P\left(\prod_{M_i\in T_k}M_i\right)$$

Каждое из отобранных правил является некоторой компонентой искомого паттерна поведения пользователя. Теперь каждое наблюдение можно оценить с точки зрения того, под действие какого множества однокомпонентных и многокомпонентных правил оно попадает, что даст возможность установить, насколько рассматриваемое наблюдение соответствует тому, что относится к норме.



а **б**

Рис. 6 Стохастическая модель сетевой активности пользователя:
а) без учета минимальной значимости связи: б) упрощенное представление

Для усиления значимости правила можно использовать такую характеристику как лифт. Она показывает насколько само правило (связка компонентов) является ценнее случайного угадывания, фактически указывает на степень зависимости соответствующих случайных величин. Чем выше лифт – тем более значимым является правило и тем более значим найденный паттерн в описании поведения сетевой активности. Таким образом, можно ввести новую величину – значимость правила:

$$\mathrm{importance}(A_{kr})=\sup(A_{kr})\times\mathrm{lift}(A_{kr})=\frac{(P(T_r\cap T_k))^2}{P(T_r)P(T_k)};\ \mathrm{lift}(A_{kr})=\frac{P(T_r\cap T_k)}{P(T_r)P(T_k)}$$

Чем выше это значение, тем более характерен найденный паттерн для описания среднестатистической активности пользователя в сети. В виду того, что определенные комбинации метрик по различным категориям не являются несовместными событиями, каждое наблюдение может получить некоторый ранг – уровень доверия к тому, что наблюдаемое состояние не является отклонением от нормы, в зависимости от того под действие какого числа правил оно попадает. Ранжированные наблюдения могут быть разделены на два класса – норма или аномалия, при этом разделение выполняется с учетом некоторого уровня толерантности – степени значимости невысокого ранга наблюдения для того, чтобы оно могло быть отнесено в группу аномалий. В качестве ранга наблюдения возьмем нормированную срезневзвешенную значимость правил, под которые оно попадает:

$$\mathrm{R}(\mathrm{t})=\frac{\mathrm{R}'(\mathrm{t})}{\sum \mathrm{R}'(\mathrm{t})};\ \mathrm{R}(\mathrm{t})=\frac{\frac{1}{\|\mathbf{T}\|}\sum\sum \mathrm{importance}(A_{kr})\times w(A_{kr},t)}{\max\left(\frac{1}{\|\mathbf{T}\|}\sum\sum \mathrm{importance}(A_{kr})\times w(A_{kr},t)\right)};\ R(t)\in[0,1];\ w(A_{kr},t)=\begin{cases}1, T_r\in\mathbf{T}(t),\\ 0, otherwise\end{cases}$$

Чем меньше ранг наблюдения, тем больше оснований полагать, что оно попадает под определение аномалии. Введя понятие индекса толерантности получаем:

$$\mathrm{Anomaly}(\mathrm{t})=\begin{cases}\mathrm{True}, \mathrm{R}(\mathrm{t})<L,\\ \mathrm{False}, \mathrm{otherwise}\end{cases}$$

Пример применения модели для случая обнаружения аномалий в сетевой активности группы пользователей приведен на рис. 7. На диаграмме приведены временные диаграммы, отображающие результаты измерений по 21-й категории, представленные в нормализованном виде. Желтыми вертикальными линиями отображены моменты времени, в которые была зафиксирована аномалия.

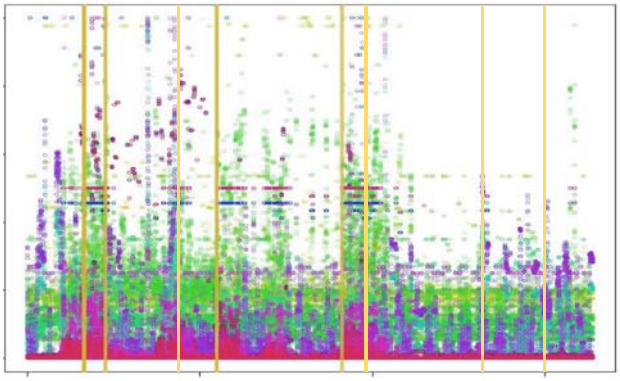


Рис. 7 Пример применения модели для случая обнаружения аномалий

Разработанная модель позволяет учитывать несколько факторов при оценке активности пользователя и позволяет описать ее в виде конечного множества правил, каждое из которых характеризуется некоторым уровнем значимости его в найденной системе паттернов. Она также как и предыдущая модель, может быть применена для анализа процесса онлайн и может быть адаптирована под меняющийся процесс за счет переоценки весов обнаруженных правил и\или добавления новых.

Time Series Clustering

В основе последней модели лежит метод сегментации временного ряда, позволяющий представить ряд в виде совокупности дискретных сегментов, которые формируют отдельные группы (кластеры), характеризующиеся тем, что компоненты временного ряда, принадлежащие одному кластеру, близки друг к другу, в то время как фрагменты из разных кластеров имеют существенные отличия [21]. Для оценки степени близости (подобия) временных рядов, можно использовать алгоритм динамической трансформации временной шкалы (DTW), [22].

Процесс сегментации временного ряда предполагает формирование множества фрагментов этого ряда, путем прохождения его плавающим окном с заданного размера и шагом дискретизации.

Результат попарного сравнивая полученных фрагменты ряда через оценку DW-расстояния, формирует матрицу смежности, которая отражает степень близости отдельных фрагментов ряда. Полученная матрица может быть использована как основа для решения задачи кластеризации фрагментов ряда. В предложенной реализации мы использовали агломеративную иерархическую кластеризацию. Для оценки расстояния между кластерами использовалось расстояние между ближайшими элементами кластеров, поскольку такая оценка расстояния позволяет формировать, так называемые протяженные кластеры, в которых подобие элементов определяется близостью хотя бы одной пары из них. Таким образом, можно ожидать, что компоненты ряда, которые отражают нормальную сетевую активность, будут формировать крупные кластеры, в то времая как аномальные компоненты будут принадлежать кластерам экстремально малого размера. В виду того, что число возможных классов нам заранее не известно – для определения их оптимального числа был использован критерий Дуда-Харта [23].

Результат кластеризации первых 13000 минут (80%) для трех метрик (countInputUDP, countInputTCP and averagePacketSizeOutputUDP) приведен на рис. 8. Фрагменты временного ряда отображены в пространстве двух главных компонент, при этом различными цветами выделены компоненты, принадлежащие различным кластерам. Видно, что в каждом случае формируется один большой кластер, содержащий те фрагменты временного ряда, которые формируют общий паттерн поведения. Кластеры малого размера могут быть интерпретированы как аномалия. На диаграммах (рис. 9) приведена

визуализация результата кластеризации на исходных временных рядах (компоненты одного кластера принадлежат заштрихованной области).

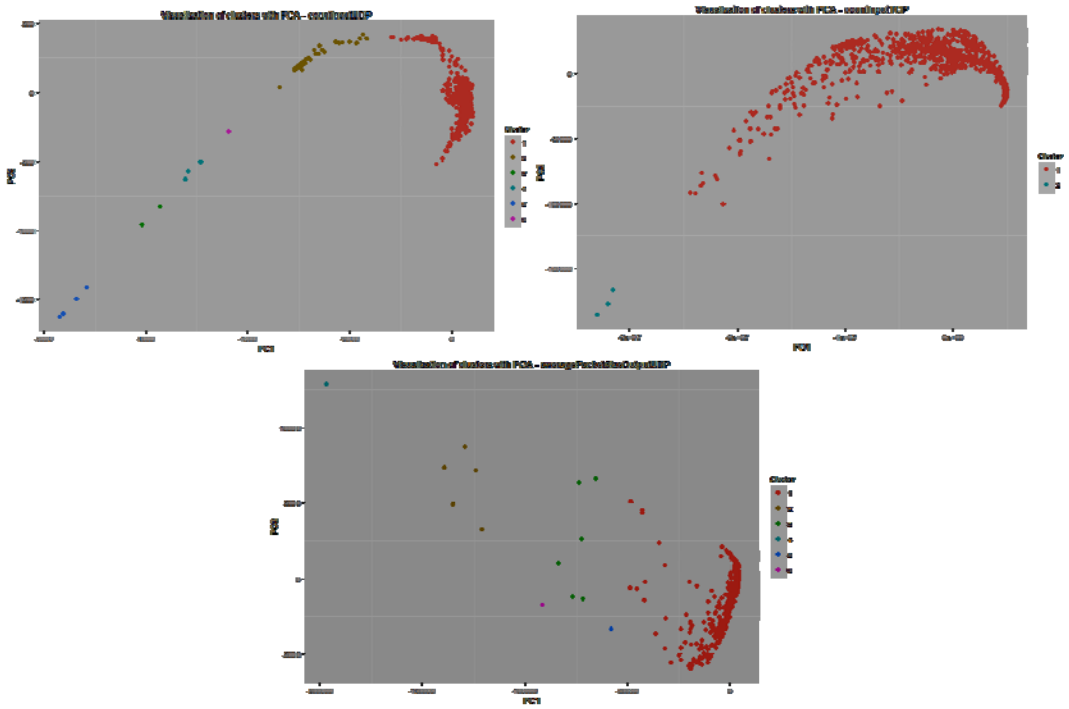
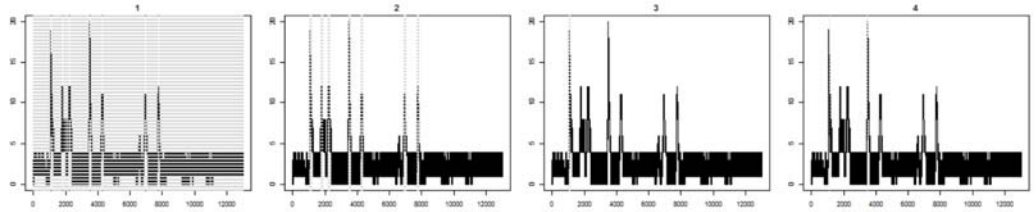


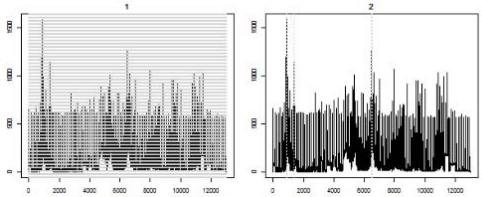
Рис. 8 Результат кластеризации первых 80% наблюдений

Видно, что малокомпонентные кластеры относятся к таким изменениям динамики временного ряда, которое является нехарактерным для большинства случаев. То насколько эти отклонения можно считать аномальными определяется мощностью соответствующих множеств. При этом значимость аномалии в определенный период времени может определяться числом метрик, в результате анализа которых этот период был установлен.

countInputUDP



countInputTCP



averagePacketSizeOutputUDP

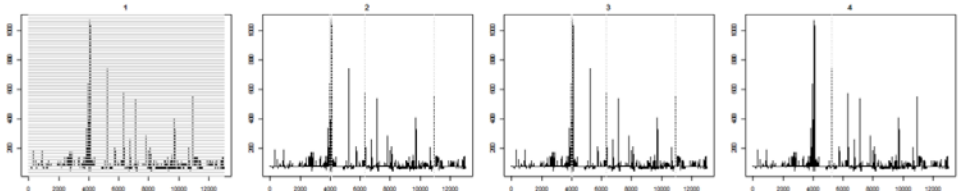


Рис. 9 Визуализация результата кластеризации на исходных временных рядах

Полученная модель позволяет оценить сетевую активность через динамику изменения измеряемого параметра и/или системы измеряемых параметров, что может позволить выявить некоторую нехарактерность даже в тех случаях, когда абсолютные значения находятся в пределах допустимых диапазонов. Единственным ее недостатком является сложность применения «на лету», поскольку реализация требует вычисления степени подобия новых наблюдений с элементами существующих

кластеров.

Ансамбль моделей

В конечной реализации модуля, позволяющего выполнять анализ сетевой активности пользователя для обнаружения случаев нарушения информационной безопасности, описанные выше модели были объединены в ансамбль (рис. 10). Как видно из диаграммы, первые две модели (Dynamic Thresholds и Association Rules) используют один и тот же набор данных, относящийся к результатам измерений по *N* категориям, полученных в режиме реального времени. Роль арбитра выполняет блок «Anomaly Confidence Level», возвращающий степень уверенности в том, что в рассматриваемый период времени наблюдается некоторая аномальная сетевая активность. В то же время Time Series Clustering Model работает со всем пулом исторических данных, что предполагает ее включение последовательное включение в модель.

В виду того, что модели отличаются режимом использования (online и offline), модель на основе сегментации временного ряда используется в не-бизнес время, для уточнения ситуации в случаях, когда аномалия не была обнаружена. По результатам офлайн проверки и последующей верификации принимается решение о необходимости адаптации моделей к новым условиям, что приводит к изменению множества шаблонов нормального поведения и, как следствие, позволяет более точно описывать наблюдаемый процесс.

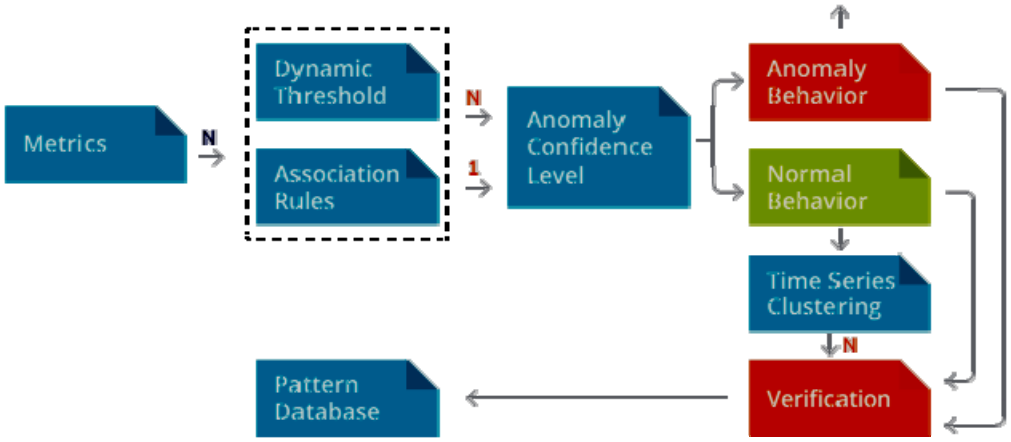


Рис. 10 Ансамбль моделей

Выводы

В рамках проведенной работы, было получено комплексное решение для обнаружения случаев, когда сетевая активность пользователя или группы пользователей отличается от известного шаблона, и, как следствие, может служить индикатором попытки нарушения информационной безопасности. Предложенное решение представляет собой ансамбль из трех моделей, которые позволяют анализировать аномалии трех типов:

1. Значимое отклонение наблюдаемого значения от ожидаемого – ***Dynamic Threshold Model.*** Основными преимуществами модели является простота имплементации и легкость в адаптации модели к меняющимся процессам. К недостаткам можно отнести изоляцию результатов анализа каждой отдельной метрики от наблюдений в остальных категориях, что затрудняет поиск событийных паттернов.

2. Нехарактерная для пользователя или группы пользователей совокупность наблюдаемых значений измеряемых параметров – ***Association Rules Based Model.*** Основным преимуществом модели является возможность описать наблюдаемый процесс как совокупность связанных событий. В качестве недостатка можно отметить нечувствительность к слабым изменениям динамики процесса.

3. Нехарактерная динамика в наблюдаемом процессе – ***Time Series Clustering Model***. Преимущества модели – найденные паттерны отражают внутреннюю динамику наблюдаемого процесса. Недостатки – не позволяет обнаруживать событийные паттерны и ее применение «на лету» крайне затруднено.

Применение ансамбля моделей позволяет нивелировать указанные недостатки и принимать решение относительно целесообразности адаптации моделей к новым, меняющимся условиям. Наше решение позволит выявлять как типичные проявления аномальной сетевой активности так и детектировать нестандартные и новые элементы сетевых аномалий. Способность самонастройки позволит решению адаптироваться к «легальным» изменениям сетевых процессов.

Литература

1. 2014: The Year in Cyberattacks, http://europe.newsweek.com/2014-year-cyber-attacks-295876?rm=eu.

2. Global cyber-attacks up 48% in 2014, http://www.cgma.org/magazine/news/pages/201411089.aspx.

3. 5 Information Security trends, http://5gensure.eu/news/5-information-security-trends-will-dominate-2016.

4. Human Factors in Information Security Management Systems - http://resources.infosecinstitute.com/human.

5. Netflow for incident detection - https://www.first.org/global/practices/Netflow.pdf.

6. Netflow and Security Cisco, Cisco Systems, Inc. 2005.

7. Varun Chandola, Anomaly Detection : A Survey - ACM Computing Surveys, Vol 41 Issue 3, Art15, 2009.

8. Afshin Shaabany Network traffic deviation detection based on fractal dimension, Journal of Computing and Information Technology, Vol 20, No 1, 2012.

9. Chen Shi-wen Self-adaptive Detection Method for DDoS Attack Based on Fractional Fourier Transform and Self-similarity, IPCSIT vol. 58, 2012, pp 42-47.

10. Ruoyu Yan Hurst Parameter for Security Evaluation of LAN Traffic, Information Technology Journal, 11, pp. 269-275., 2012.

11. P. Ably SelfSimilarity and long-range dependence through the wavelet lens, Theory and Applications of Long Range Dependence, 2002, pp 345-379.

12. Forrest, S. An immunological approach to change detection: Algorithms, analysis and implications, IEEE Symposium on Security and Privacy, 1996.

13. Jamie Twycross An Immune Inspired Approach to Anomaly Detection, Handbook of Research on Information Security and Assurance, Chapter X, 2008.

14. Zhang, Z. Hide: a hierarchical network intrusion detection system using statistical preprocessing and neural network classification, IEEE Workshop on Information Assurance and Security, 2001, 85–90.

15. S. Yu Conserved Self Pattern Recognition Algorithm, ICARIS 2008, Vol 5132, 2008, pp. 279–290.

16. Laheeb, Anomaly network intrusion detection system based on distributed time-delay neural network (dtdnn) Journal of Engineering Science and Technology, Vol. 5, No. 4, 2010, pp 457 – 471.

17. Chan, Learning rules and clusters for Anomaly detection, Managing Cyber Threats, Vol 5, pp 81-99.

18. James Douglas Hamilton, Time Series Analysis, Princeton University Press, 1994, p. 820.

19. Castle, Jennifer L A low-dimension portmanteau test for non-linearity, Journal of Econometrics, vol. 158(2), pp 231-245, 2010.

20. Jean-Marc Adamo , Data Mining for Association Rules and Sequential Patterns: Sequential and Parallel Algorithms, Springer-Verlag New York, 2001, p. 254.

21. Ronald S. King Cluster Analysis and Data Mining, Mercury Learning and Information, 2014, p. 300.

22. Meinard Müller, Information Retrieval for Music and Motio, Springer-Verlag New, 2007, p. 313.

23. Richard O. Duda, Peter E. Hart Pattern Classification.- John Wiley & Sons, Inc.Copyright, 2001.

References

1. 2014: The Year in Cyberattacks, http://europe.newsweek.com/2014-year-cyber-attacks-295876?rm=eu.

2. Global cyber-attacks up 48% in 2014, http://www.cgma.org/magazine/news/pages/201411089.aspx.

3. 5 Information Security trends, http://5gensure.eu/news/5-information-security-trends-will-dominate-2016.

4. Human Factors in Information Security Management Systems - http://resources.infosecinstitute.com/human.

5. Netflow for incident detection - https://www.first.org/global/practices/Netflow.pdf.

6. Netflow and Security Cisco, Cisco Systems, Inc. 2005.

7. Varun Chandola, Anomaly Detection : A Survey - ACM Computing Surveys, Vol 41 Issue 3, Art15, 2009.

8. Afshin Shaabany Network traffic deviation detection based on fractal dimension, Journal of Computing and Information Technology, Vol 20, No 1, 2012.

9. Chen Shi-wen Self-adaptive Detection Method for DDoS Attack Based on Fractional Fourier Transform and Self-similarity, IPCSIT vol. 58, 2012, pp 42-47.

10. Ruoyu Yan Hurst Parameter for Security Evaluation of LAN Traffic, Information Technology Journal, 11, pp. 269-275., 2012.

11. P. Ably SelfSimilarity and long-range dependence through the wavelet lens, Theory and Applications of Long Range Dependence, 2002, pp 345-379.

12. Forrest, S. An immunological approach to change detection: Algorithms, analysis and implications, IEEE Symposium on Security and Privacy, 1996.

13. Jamie Twycross An Immune Inspired Approach to Anomaly Detection, Handbook of Research on Information Security and Assurance, Chapter X, 2008.

14. Zhang, Z. Hide: a hierarchical network intrusion detection system using statistical preprocessing and neural network classification, IEEE Workshop on Information Assurance and Security, 2001, 85–90.

15. S. Yu Conserved Self Pattern Recognition Algorithm, ICARIS 2008, Vol 5132, 2008, pp. 279–290.

16. Laheeb, Anomaly network intrusion detection system based on distributed time-delay neural network (dtdnn) Journal of Engineering Science and Technology, Vol. 5, No. 4, 2010, pp 457 – 471.

17. Chan, Learning rules and clusters for Anomaly detection, Managing Cyber Threats, Vol 5, pp 81-99.

18. James Douglas Hamilton, Time Series Analysis, Princeton University Press, 1994, p. 820.

19. Castle, Jennifer L A low-dimension portmanteau test for non-linearity, Journal of Econometrics, vol. 158(2), pp 231-245, 2010.

20. Jean-Marc Adamo , Data Mining for Association Rules and Sequential Patterns: Sequential and Parallel Algorithms, Springer-Verlag New York, 2001, p. 254.

21. Ronald S. King Cluster Analysis and Data Mining, Mercury Learning and Information, 2014, p. 300.

22. Meinard Müller, Information Retrieval for Music and Motio, Springer-Verlag New, 2007, p. 313.

23. Richard O. Duda, Peter E. Hart Pattern Classification.- John Wiley & Sons, Inc.Copyright, 2001.

Рецензія/Peer review : 6.2.2016 р. Надрукована/Printed : 26.3.2016 р.