УДК 004.93'12

М.Б. Будько, М.Ю. Будько

АНАЛИЗ И ОПТИМИЗАЦИЯ АЛГОРИТМОВ ВЕЙВЛЕТ-РАЗЛОЖЕНИЯ ДЛЯ ПРИМЕНЕНИЯ В РЕЖИМЕ РЕАЛЬНОГО ВРЕМЕНИ В СИСТЕМАХ ОБНАРУЖЕНИЯ АНОМАЛИЙ ТЕЛЕКОММУНИКАЦИОННЫХ ДАННЫХ

Будько М.Б., Будько М.Ю. **Анализ и оптимизация алгоритмов вейвлет-разложения для применения в режиме реального времени в системах обнаружения аномалий телекоммуникационных данных**.

Аннотация. Одним из подходов к обнаружению сетевых аномалий является анализ рядов показателей функционирования сети. Характеристики, рассчитанные по вейвлет-коэффициентам, действительно, более чувствительны к изменениям ряда, чем характеристики, рассчитанные непосредственно по ряду, но при этом требуют большего объема вычислений, поэтому спектрально-временные алгоритмы, безусловно, подлежат оптимизации для применения в системах реального времени. Кроме того, существуют различные подходы к выполнению вейвлет-разложений, каждый из которых занимает свое место по информативности (по количеству уточняющих коэффициентов), области достоверных значений, вычислительной сложности преобразований. В статье предлагается обоснованный подход к выполнению таких алгоритмов для применения в режиме реального времени в системах обнаружения аномалий телекоммуникационных данных.

Ключевые слова: вейвлет, всплеск, спектрально-временной анализ, система обнаружения аномалий, системы реального времени, телекоммуникационные данные, анализ временных рядов, обработка сигналов.

Budko M.B., Budko M.Y. **Analysis and optimization of wavelet decomposition algorithms for use in real-time anomaly detection systems.**

Abstract. One of the approaches to the detection of network anomalies is the analyses of parameters of functioning of a network. Characteristics, calculated on a wavelet coefficients, indeed, are more sensitive to changes in the number, than the characteristics calculated directly in a row, but this requires more calculations, the spectral-time algorithms, of course, subject to optimize for application in real-time systems. In addition, there are different approaches to the implementation of wavelet expansions, each of which has its place on the informative value (the number of qualifying ratios), the authentic values, the computational complexity of the transformations. The article offers a reasonable approach to the implementation of these algorithms for use in real-time anomaly detection systems.

Keywords: wavelet, spectral-temporal analysis, intrusion detection systems, real-time systems, telecommunications data, time series analysis, signal processing.

1. Введение. *Система обнаружения вторжений* (*СОВ*) — программное или аппаратное средство, собирающее информацию из различных точек защищаемой компьютерной системы и анализирующее эту информацию для выявления попыток нарушения и реальных нарушений защиты (вторжений) — *происшествий с безопасностью*. Соответствующий английский и наиболее распространенный тер-

мин — *Intrusion Detection System* (*IDS*). Обнаружение аномалий часто представляется в виде дополнительного сервиса, который рекомендуется включать в любую систему безопасности для возможности улавливать атаки, не описанные сигнатурами.

Среди известных методов анализа данных для определения преданомальных участков и выявления закономерностей сигнала было использовано вейвлет-разложение рядов показателей функционирования сети.

Вейвлет-преобразование разлагает анализируемый процесс на составляющие его волны, компоненты разного масштаба и, кроме того, в отличие от спектральных преобразований, дает «локализованную» во времени информацию о процессе. Горизонтальное сечение картины коэффициентов демонстрирует изменение компоненты выбранного масштаба со временем. Вертикальное сечение картины коэффициентов в некоторый момент времени демонстрирует поведение процесса в окрестности выбранного момента времени, что позволяет определить наличие особенности и набор задействованных масштабов.

2. Основная часть. Характеристики, рассчитанные по вейвлет-коэффициентам, действительно, более чувствительны к изменениям ряда, чем характеристики, рассчитанные непосредственно по ряду, но при этом требуется больший объем вычислений, поэтому спектрально-временные алгоритмы, безусловно, требуют оптимизации для применения в системах реального времени. Кроме того, существуют различные подходы к выполнению вейвлет-разложений, каждый из которых занимает свое место по информативности (по количеству уточняющих коэффициентов), области достоверных значений, вычислительной сложности преобразований.

Были рассмотрены: 1) дискретное преобразование с изменением масштабов и сдвигов по значению, равному временной локализации вейвлета; 2) дискретный диадный кратномасштабный анализ и 3) дискретизированное разложение. Для дискретного диадного разложения, в том числе при реализации быстрого алгоритма, в схемах был учтен существующий подход, связанный с децимацией исходной последовательности, значительно снижающей вычислительную сложность. Также определены углы влияния и достоверности. Для этого преобразования в область недостоверности попадает значительная часть коэффициентов низкочастотных уровней. Быстрый алгоритм является при этом предпочтительным, т.к. его коэффициенты более точно описывают исходный ряд.

Аналогично были проанализированы углы влияния и достоверности для дискретизированного разложения с использованием 2-х и 4-х элементных «материнских» вейвлетов и децимацией последовательности. Бóльшая область определения вейвлета влечет увеличение области недостоверных значений коэффициентов. Несмотря на схожесть общего вида угла правдоподобия с полученным для дискретного диадного разложения, раствор угла для текущего преобразования будет уже, т.е. площадь области недостоверных значений будет меньше, а достоверных больше, что объясняется меньшими шагами при изменении масштабов и сдвигов, т.е. более детальным представлением коэффициентов разложения.

С позиции информативности по количеству уточняющих коэффициентов лучшим является случай 3), затем 2), затем 1).

С точки зрения увеличения области достоверных значений лучшим является случай 1), затем 3), затем 2).

В порядке возрастания вычислительной сложности преобразования располагаются порядке 1), 2), 3).

Оптимальным кажется метод на основе вейвлет–фреймов, являющийся компромиссом между диадным кратномасштабным анализом с грубоватым представлением результата и дискретизированным разложением с избыточным представлением сетки. Причем именно первое должно быть взято за основу, т.к. имеет хорошо оптимизированные по организации вычислений и используемой памяти быстрые алгоритмы статического вычисления сетки коэффициентов, и расширено для наделения достоинствами дискретизированного разложения. Но как оказалось при тестировании, область достоверных значений играет очень важную роль, поэтому пришлось остановиться на дискретном разложении, имеющем полностью достоверную сетку коэффициентов при правильном выборе длины окна сканирования, с использованием 2-х элементного вейвлета, что делает разложение диадным.

Оптимизация расчетов заключалась в том, что 1) для расчета отдельных коэффициентов более «низкочастотного» уровня использовалось достаточно много одних и тех же уже вычисленных коэффициентов более «высокочастотного» уровня и в том, что 2) при сдвиге последовательности схема вычислений для одних коэффициентов повторяла схему вычислений для других коэффициентов того же уровня, уже вычисленных на предыдущих итерациях.

Применение вейвлет-анализа к сетевому трафику накладывает ограничения и на сами вейвлеты, основное из которых минимальная временная локализация, поскольку она влияет на алгоритмическую

задержку вычислений. Кроме того, бóльшая длина вейвлета негативно влияет на область правдоподобия, расширяя область недостоверности за счет участия в вычислениях несуществующих отсчетов последовательности. Крайние же правые коэффициенты являются наиболее значимыми для детектирования особенностей временной последовательности в режиме реального времени, т.к. именно они отражают последние изменения в сигнале.

Поэтому, было решено в исследовании использовать в качестве основного вейвлет Хаара с областью задания в 2 отсчета и одним нулевым моментом.

Во многих информационных источниках говорится о лучшей идентификации поведения временного ряда при исследовании его разными вейвлетами. Такая позиция оправдана и эмпирически доказана. Поэтому в исследовании должен использоваться дополнительный вейвлет, улавливающий более «тонкие» особенности сигнала. Большая временная локализация обуславливает более позднее формирование достоверного вида аномалии на картине вейвлет-коэффициентов и позиционирует такой вейвлет, как «страхующий».

3. Заключение. Итак, при тестировании использовался 2–элементный вейвлет Хаара с минимальной для вейвлетов областью определения. Применялось дискретное разложение, которое в случае 2–элементного вейвлета равносильно диадному, имеет быстрые алгоритмы вычисления, характеризуется отсутствием области недостоверных коэффициентов и возможностью пересчета только отдельных значений при движении окна сканирования вдоль временного ряда. Количество пересчитываемых коэффициентов достаточно быстро (в зависимости от размера окна сканирования) стабилизируется на значении, равном числу масштабных уровней разложения.

Предложенный подход обнаружения аномалий не раскрывает всю динамику процесса с поиском источников в виде отдельных сервисов, вирусов, функционирующих с ошибками каналов, устройств и т.д., но он позволяет зафиксировать типичные для аномалий и других особенных состояний характеристики, связанные с нетрадиционными частотными картинами и тем самым обнаружить или предопределить опасность нанесения вреда функционированию сети. Сама возможность выявлять преданомальные участки может служить толчком к целенаправленной разработке аналогичных методов во временной области без применения вейвлет-разложений, использование которых требует дополнительных знаний и навыков при обработке данных и увеличивает время вычислений.

Литература

1. *Астафьева Н.* Вейвлет анализ: основы теории и примеры применения // Успехи Физических Наук. 1996. Т. 166, № 11. С. 1145–1170.
2. *Бендат Дж., Пирсол А.* Применения корреляционного и спектрального анализа: Пер. с англ. М. : Мир, 1983. 312 с., ил.
3. *Будько М.Б.* Особенности применения вейвлетов к анализу данных мониторинга сети // III Всероссийская научно-техническая конференция ИКВО НИУ ИТМО «Проблема комплексного обеспечения информационной безопасности и совершенствование образовательных технологий подготовки специалистов силовых структур» : сб. трудов. СПб: Изд-во НИУ ИТМО, 2012. С. 8–14.
4. *Будько М.Б., Будько М.Ю.* Отслеживание изменений в структуре сети и решение задач повышения безопасности на основе анализа потоков данных // Научно-технический вестник Санкт-Петербургского государственного университета информационных технологий, механики и оптики. СПб : Изд-во СПбГУ ИТМО, 2009. № 59. С. 78–82.
5. *Воробьев В., Грибунин В.* Теория и практика вейвлет-преобразования. СПб : Изд-во ВУС, 1999. 208 с.
6. *Добеши И.* Десять лекций по вейвлетам : Пер. с англ. – Ижевск : НИЦ «Регулярная и хаотическая динамика», 2001. 464 с.
7. *Дремин И., Иванов О., Нечитайло В.* Вейвлеты и их использование // Успехи физических наук. 2001. Т. 171, № 5. С. 465–561.
8. *Лазоренко О.В., Лазоренко С.В., Черногор Л.Ф.* Вейвлет-анализ модельных сигналов с особенностями. 1. Непрерывное вейвлет-преобразование // Радиофизика и радиоастрономия. 2007. Т. 12, № 2. С. 182–204.

Будько Марина Борисовна — к.т.н., доцент кафедры мониторинга и прогнозирования информационных угроз Института комплексного военного образования Санкт-Петербургского национального исследовательского университета информационных технологий, механики и оптики (НИУ ИТМО). Область научных интересов: корреляционный, спектральный и спектрально-временной анализ, математическая статистика, дискретная математика, вычислительные сети. Число научных публикаций — 20. budkomb@mail.ru, www.ifmo.ru; НИУ ИТМО, Кронверкский пр., д. 49, Санкт-Петербург, 197101, РФ; р.т. +7(812)595-4132.

Budko Marina Borisovna — Ph.D., lecturer of faculty Monitoring and forecasting of IT threats Institute of complex military education Saint-Petersburg National Research University of Information Technologies, Mechanics and Optics (ITMO). Research interests: correlative analysis, spectral analysis, mathematical statistics, discrete mathematics, computer networks. The number of publications — 20. budkomb@mail.ru, www.ifmo.ru; ITMO, IKVO, Kronverksky pr-t., 49, St. Petersburg, 197101, Russia; office phone. +7(812)595-41-32.

Будько Михаил Юрьевич — к.т.н., доцент кафедры мониторинга и прогнозирования информационных угроз Института комплексного военного образования Санкт-Петербургского национального исследовательского университета информационных технологий, механики и оптики (НИУ ИТМО). Область научных интересов: вычислительные комплексы, системы и сети. Число научных публикаций — 22. bmu@mail.ru,

www.ifmo.ru; НИУ ИТМО, Кронверкский пр., д. 49, Санкт-Петербург, 197101, РФ; р.т. +7(812)595-4132.

Budko Mikhail Yurievich — Ph.D., lecturer of faculty Monitoring and forecasting of IT threats Institute of complex military education Saint-Petersburg National Research University of Information Technologies, Mechanics and Optics (ITMO). Research interests: computing complexes, systems and networks, network monitoring. The number of publications — 22. bmu@mail.ru, www.ifmo.ru; ITMO, IKVO, Kronverksky pr-t., 49, St. Petersburg, 197101, Russia; office phone. +7(812)595-41-32.

Поддержка исследований. В публикации представлены результаты исследований, поддержанные грантом Министерства обороны Российской Федерации на тему «Разработка методов обнаружения и противодействия вторжениям в вычислительных сетях военного назначения», рук. Г.П. Жигулин.

Рекомендовано лабораторией автоматизации научных исследований СПИИРАН.
Статья поступила в редакцию 12.03.2013.

РЕФЕРАТ

Будько М.Б., Будько М.Ю. **Анализ и оптимизация алгоритмов вейвлет-разложения для применения в режиме реального времени в системах обнаружения аномалий телекоммуникационных данных.**

Система обнаружения вторжений (СОВ) — программное или аппаратное средство, собирающее информацию из различных точек защищаемой компьютерной системы и анализирующее эту информацию для выявления попыток нарушения и реальных нарушений защиты (вторжений) — происшествий с безопасностью. Обнаружение аномалий часто представляется в виде дополнительного сервиса, который рекомендуется включать в любую систему безопасности для возможности улавливать атаки, не описанные сигнатурами.

Среди известных методов анализа данных для определения преданомальных участков и выявления закономерностей сигнала было использовано вейвлет-разложение рядов показателей функционирования сети. Существуют различные подходы к выполнению вейвлет-разложений. Были рассмотрены: дискретное преобразование с изменением масштабов и сдвигов по значению, равному временной локализации вейвлета; дискретный диадный кратномасштабный анализ и дискретизированное разложение. Были выделены следующие критерии: информативность (по количеству уточняющих коэффициентов), область достоверных значений, вычислительная сложность преобразования. Применение вейвлет-анализа к сетевому трафику накладывает ограничения и на сами вейвлеты, основное из которых минимальная временная локализация, поскольку она влияет на алгоритмическую задержку вычислений. Кроме того, бóльшая длина вейвлета негативно влияет на область правдоподобия. Поэтому, было решено в исследовании использовать в качестве основного вейвлет Хаара с областью задания в 2 отсчета и одним нулевым моментом. Во многих информационных источниках говорится о лучшей идентификации поведения временного ряда при исследовании его разными вейвлетами. Поэтому в исследовании должен использоваться дополнительный вейвлет, улавливающий более «тонкие» особенности сигнала. Большая временная локализация обуславливает более позднее формирование достоверного вида аномалии на картине вейвлет-коэффициентов и позиционирует такой вейвлет, как «страхующий». Предложенный подход обнаружения аномалий не раскрывает всю динамику процесса с поиском источников в виде отдельных сервисов, вирусов, функционирующих с ошибками каналов, устройств и т.д., но он позволяет зафиксировать типичные для аномалий и других особенных состояний характеристики, связанные с нетрадиционными частотными картинами и тем самым обнаружить или предопределить опасность нанесения вреда функционированию сети. Сама возможность выявлять преданомальные участки может служить толчком к целенаправленной разработке аналогичных методов во временной области без применения сравнительно трудоемких вейвлет–разложений.

SUMMARY

Budko M.B., Budko M.Y. **Analysis and optimization of wavelet decomposition algorithms for use in real-time anomaly detection systems.**

Intrusion Detection System (IDS) – hardware or software that collects information from different points of the protected computer system and analyze this information to detect attempts to breach and actual violations of protection (intrusion). Anomaly detection is often presented as an additional service, which should include any security system to be able to detect attacks that are not described signatures.

Among the known methods of data analysis to determine anomaly sites and reveal the regularities of the signal was used wavelet decomposition of series of indicators of the performance of the network. There are different approaches to the implementation of wavelet expansions. Considered: discrete transformation with the change of scale and shifts in value equal time localization of the wavelet; discrete dyadic multiresolution analysis and discretized expansion. Identified the following criteria: informativeness (by specifying the number of qualifying ratios), the area of reliable values, the computational complexity of the transformation.

Application of wavelet analysis to the network traffic imposes limitations on wavelets, the main of which is the minimum time of localization, because it influences the algorithmic delay calculations. Moreover, most wavelet length affects the area of likelihood. Therefore, it was decided to use as a primary Haar wavelet with domain of 2 samples and one zero moment. In many information sources referred to better identify the behavior of the time series in the study of its by different wavelets. Therefore, the study should be used additional wavelet, capture more «subtle» features of the signal. Most time localization causes a later generation of reliable type anomalies in the picture of the wavelet coefficients and positions such wavelet, as «hedges».

The proposed approach anomaly detection does not reveal all of the dynamics of the process with the search for sources in the form of individual services, viruses, operating with errors channels, devices, etc., but it allows you to fix the typical for anomalies and other special conditions characteristics associated with non-traditional frequency pattern and thus discover or predetermine the risk of harm to the functioning of the network.The possibility to identify predanomalnye areas can serve the impetus for the purposeful development of similar methods in the time domain without the use of a relatively labor-intensive wavelet expansions.