**УДК 681.3.06**

**Л.А. Терейковская**

*Киевский национальный университет строительства и архитектуры, Киев*

# РАЗРАБОТКА СТАТИСТИЧЕСКОЙ МОДЕЛИ РАСЧЕТА ПЕРИОДИЧЕСКИХ СОСТАВЛЯЮЩИХ ДИНАМИКИ ФУНКЦИОНАЛЬНЫХ ПАРАМЕТРОВ INTERNET-СЕРВЕРОВ

*Обоснована необходимость использования статистической модели расчета частотно-временных характеристик функциональных параметров Internet-серверов. С использованием теории вейвлет-преобразований разработано математическое обеспечение такой модели. Указаны пути усовершенствования модели.*

***Ключевые слова: Internet-сервер, техническое состояние, вейвлет-анализ, вейвлет-преобразования***

## Постановка проблемы

В процессе решения задач диагностирования и защиты компьютерных сетей одним из центральных вопросов является определение и прогнозирование технического состояния Internet-серверов, работоспособность которых во многих случаях имеет решающее значение. Наличие достоверного прогноза позволяет на ранних этапах выявить предотказные состояния Internet-серверов, своевременно устранить их причину, а также позволяет предотвратить возможные катастрофические последствия. В настоящее время для определения и прогнозирования технического состояния используется довольно много специализированных систем. Так, для диагностирования технического состояния Internet-серверов применяются средства систем управления, анализаторы сетевых протоколов, системы нагрузочного тестирования, системы сетевого мониторинга. Проблемы защиты информационных ресурсов Internet-серверов решаются с помощью систем обнаружения уязвимостей, систем защиты от спама, сетевых экранов, антивирусов, систем обнаружения атак, систем контроля целостности, криптографических средств защиты. Характерными особенностями использования и систем диагностирования, и систем защиты является либо их периодическое и кратковременное применение для решения определенной проблемы, либо постоянное использование со статическими настройками. Например, методы анализа, используемые в современных системах обнаружения атак, направлены на обнаружение известных и точно описанных типов воздействий, но зачастую оказываются не в состоянии обнаружить их модификации или новые типы атак. В ряде случаев это делает их использование малоэффективным. По этой причине не вызывает сомнения необходимость разработки эффективных методов определения и прогнозирования недопустимых значений функциональных параметров Internet-серверов, как вследствие технических сбоев, так и вследствие несанкционированных воздействий. Основным требованием к этим методам является возможность обнаружения произвольных типов аномального функционирования, а также несанкционированных воздействий, распределенных во времени. Общий подход, лежащий в основе подобных методов, заключается в поиске отклонений функциональных параметров Internet-серверов от своего обычного состояния. Отклонения могут являться результатами сбоев в работе аппаратного и программного обеспечения, а также следствиями нештатных условий эксплуатации или сетевых атак. Эффективность подхода непосредственно зависит от методики прогнозирования значений функциональных параметров Internet-серверов, что и является общей проблемой настоящей статьи.

## Анализ последних достижений и публикаций

Очевидно, что методика прогнозирования технического состояния Internet-серверов в значительной мере зависит от вида и параметров функциональных параметров, определяющих это состояние. Большинство современных методик прогнозирования базируется на использовании эксплуатационных значений параметров сетевого трафика Internet-серверов, передаваемого на основе стека протоколов TCP/IP. Для того чтобы определить и прогнозировать параметры трафика вводят модель каждого из пяти ключевых объектов стека: пакет $K_1$, сеть $K_2$, хост $K_3$, порт $K_4$ и сессия

 

($K_5$). Каждый ключевой объект характеризуют при помощи набора характерных для него параметров $K_i(k_i^1, k_i^2 ...), i \in \{1,5\}$. Каждая из моделей имеет два типа параметров:

– простые параметры;

– составные параметры.

Простые параметры описывают одну из характеристик модели, например, объем трафика в час или нормальную длину пакета. Простые параметры не могут указывать на ключевые объекты. Составные параметры – указатели на логически структурированные множества ключевых объектов нижнего уровня, например, список портов или дерево хостов. В современных системах обнаружения атак и системах диагностирования для реализации статистического анализа параметров TCP/IP-трафика широко используются такие модели, как операционная, среднего значения и среднеквадратичного отклонения, многовариационная, марковского процесса, временных серий. Рассмотрим основные характеристики этих моделей.

Операционная модель основывается на том, что каждое зарегистрированное значение параметра должно укладываться в определенные пределы. Если этого не происходит, то фиксируется отклонение от обычного состояния. Допустимые границы определяются на основании анализа предыдущих значений переменной. Данная модель является очень упрощенной и может использоваться только в случаях, когда некоторое значение параметра (ряда параметров) однозначно диагностирует предотказное состояние Internet-сервера. Например, если размер очереди запросов к Internet-серверу больше допустимой.

Модель среднего значения базируется на том, что предотказное состояние (атаку) можно определить в результате сравнения текущего значения функционального параметра $k_i$ с его средним значением $m$ и среднеквадратическим отклонением $s$

$$\exists k_i \notin [m - as, m + as] \Rightarrow O, \qquad (1)$$

где $O$ – отклонение, $a$ – коэффициент $0<a<1$,

$$m = \frac{\sum_{j=1}^{N} k_j^2}{N},\ s = \sqrt{\frac{\sum_{j=1}^{N} k_j^2}{N} - m^2},$$

$N$ – количество зарегистрированных значений параметра.

Модель применима для измерения счетчиков событий, временных интервалов и используемых ресурсов.

Многовариационная модель аналогична модели среднего значения и среднеквадратичного отклонения, но учитывает корреляцию между двумя или большим количеством функциональных параметров.

*Модель временных серий* использует временные периоды вместе со счетчиками событий и измерениями функциональных параметров. В модели учитываются как значения функциональных параметров $k_1$, $k_2$,…, $k_i$, так и временные интервалы между ними. Новое наблюдение является анормальным, если вероятность его появления с учетом времени низка. Преимуществом данной модели является учет временного сдвига между событиями, а недостаток – накладные расходы по вычислению по сравнению с моделью среднего значения и среднеквадратичного отклонения.

*Марковская модель* базируется на представлении динамики функциональных параметров в виде дискретного марковского процесса с ограниченным количеством состояний. Использование марковской модели позволяет прогнозировать динамику функциональных параметров. Аномальное состояние фиксируется в случае большой невязки между прогнозируемым и зарегистрированным значением функционального параметра.

Анализ всех рассмотренных статистических моделей указывает на их недостаточную адекватность реальным статистическим данным, что в свою очередь негативно сказывается на их информативности и достоверности. Неадекватность моделей заключается в игнорировании нестационарности поведения многих функциональных параметров Internet-серверов. Учесть нестационарность возможно за счет использования нескольких марковских процессов, каждый из которых будет моделировать динамику функциональных параметров на условно-стационарном участке. Проблемой реализации такого подхода является определение параметров стационарных участков. Проблема усложняется тем, что динамика многих параметров многопериодична. В качестве примера многопериодичности на рис. 1 показан график объема трафика SNMP-сервера на протяжении 2000 с.

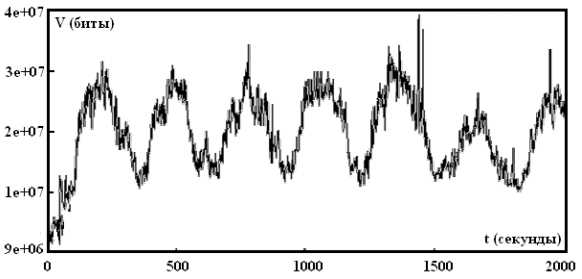


Рис. 1. Трафик SNMP-сервера

Визуальный анализ рис. 1 показывает наличие одного четко выраженного периода длительностью примерно 250-350 с и нескольких более коротких и

менее выраженных периодов. Также можно отметить, что некоторые периодические составляющие возникают не в начальный момент времени, а после некоторого периода эксплуатации. Это указывает на необходимость решения задачи расчета частотно-временных характеристик функциональных параметров Internet-серверов. Отметим, что в настоящее время подобные задачи решаются за счет разработки и использования специальных моделей, которые базируются на теории вейвлет-преобразований.

**Цель статьи.** Используя теорию вейвлет-преобразований разработать математическое обеспечение модели расчета частотно-временных характеристик функциональных параметров Internet-серверов.

## Изложение основного материала исследований

С точки зрения теории вейвлет-преобразований обработка исходных статистических данных, касающихся функциональных параметров Internet-серверов, с целью определения периодических составляющих представляют собой построения модели соответствующего сигнала. При этом под понятием сигнал понимают физический процесс, протекающий во времени и несущий информацию о некоторых событиях или состоянии объекта наблюдения. Как правило сигнал представляют в виде функции

$$s = s(t, \Theta), \qquad (2)$$

где $s$ – сигнал; $t$ – время; $\Theta$ – вектор случайных параметров.

По причине того, что регистрация статистических данных Internet-серверов производится через определенные интервалы времени на конечном участке, их относят к финитным дискретным сигналам. С точки зрения вейвлет-анализа такой сигнал характеризуется энергией и моментами.

Энергия сигнала на заданном интервале времени $=[t_1, t_2]$ рассчитывается так

$$E = \int_{t_1}^{t_2} s(t)^2 dt. \qquad (3)$$

Начальным моментом $v$-го порядка функции $s(t)$ называется выражение вида

$$m_v = m_0^{-1} \int_{-\infty}^{\infty} t^v s(t) dt, \qquad (4)$$

где $m_0$ – нулевой момент, который рассчитывается так

$$m_0 = \int_{-\infty}^{\infty} s(t) dt. \qquad (5)$$

Центральным моментом $v$-го порядка функции $s(t)$ называется выражение вида

$$\overset{0}{m}_v = m_0^{-1} \int_{-\infty}^{\infty} (t - m_1)^v s(t) dt, \qquad (6)$$

где $m_1$ – первый начальный момент.

Как правило, сигнал можно представить в виде суммы простых колебаний, совокупность интенсивностей которых называется спектром. Другими словами спектр это набор чисел, определяющий долю каждого колебания в сигнале. Теория вейвлет-преобразований позволяет не только определить спектр сигнала, но и локализовать частотные характеристики спектра во времени. Формально интегральное вейвлет-преобразование функции $f(t) \in L^2(R)$ записывается так

$$W(a,b) = |a|^{-0,5} \int_{-\infty}^{\infty} f(t) \psi^* \left( \frac{t-b}{a} \right) dt, \qquad (7)$$

где $\psi$ – базовый вейвлет (базисная функция), * – процедура комплексного сопряжения; $a$ – масштаб вейвлета; $b$ – сдвиг вейвлета, $a,b \in R$, $a \neq 0$.

Базисная функция вейвлета должна отвечать общим требованиям.

Площадь $s$, ограниченная этой функцией равна 0, т.е. должен быть равным 0 нулевой момент функции, рассчитанный соответственно (5).

Энергия функции (3) должна быть конечной, концентрироваться внутри некоторого финитного интервала $\Delta t=[t_1, t_2]$ и быстро убывать до нуля вне этого интервала

Кроме этого, для анализа рядов с полиноминальным трендом в базисных вейвлетах должны равняться нулю центральные моменты $v$-го порядка. В задачах инженерного плана большое распространение получили вейвлет-функции, построенные на основе производных функций Гаусса вида

$$\psi(t) = \frac{(-1)^k}{\sqrt{2\pi}} \frac{\partial^k}{\partial t^k} e^{-0,5t^2}, \qquad (8)$$

где $k$ – целое число, $k \geq 1$.

Увеличение $k$ увеличивает количество центральных моментов, которые равняются 0, что позволяет извлечь из сигнала информацию об особенностях более высокого порядка. С позиций обработки статистики Internet-сервера это предоставляет возможность более детально анализировать высокочастотные составляющие. Заметим, что повышенная детализация может отрицательно повлиять на обобщающую сторону анализа, а также значительно повысить объем вычислительных операций. При этом для общего случая необходимая степень детализации теоретически не обоснована. Следовательно, определение вида и параметров базисных вейвлет-функций, применяемых для анализа статистики

Internet-сервера требует дальнейших исследований.

Результатом вейвлет-преобразования одномерного ряда динамики, которому соответствует статистика Internet-сервера является множество значений коэффициентов *W*, определенное в пространстве *a* (масштаб вейвлета) и *b* (сдвиг вейвлета). Указанное множество называют вейвлет-спектром или амплитудой вейвлет-функции. Вейвлет-спектр визуально интерпретируют как поверхность в трехмерном пространстве *W, a, b*. Для выделения наиболее значимых частот рассчитывают скалограмму, представляющую собой линии локальных экстремумов *W( a, b)* на каждом масштабе *a*

$$S(a_i, b_i) = |W(a_i, b_i)|^2 \qquad (9)$$

Скалограмма позволяет описать распределение энергии функции вейвлет-спектра по различным масштабам. Выделение на скалограмме локальных максимумов по переменным *a* и *b* производится на основании выражения вида

$$C_{i,j} = \begin{cases} S_{i,j} \; \exists (S_{i,j} > S_{i-1,j}) \wedge (S_{i,j} > S_{i+1,j}), \\ S_{i,j} \; \exists (S_{i,j} > S_{i,j-1}) \wedge (S_{i,j} > S_{i,j+1}), \\ 0 - \textit{в противном случае.} \end{cases} \qquad (10)$$

где $S_{i,j}=S(a_i,b_j)$.

Функцию (10) называют скелетоном. Считается, что использование скелетона позволяет качественно разделить низкочастотные (информативная составляющая) и высокочастотные (шум) составляющие анализируемого сигнала. В качестве примера на рис. 3 показан вейвлет-спектр, а на рис. 4 скалограмма аналитической функции количества запросов к Веб-серверу. График указанной функции показан на рис.2. Отметим, что график рис.2 отвечает трехпериодическому процессу динамики запросов ($l_1$=1, $l_2$=8, $l_3$=24 ч).

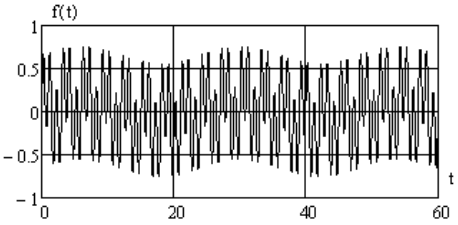


Рис. 2. График количества запросов к Веб-серверу

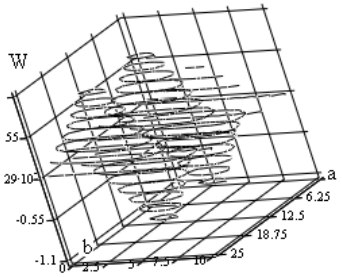


Рис. 3. Вейвлет-спектр запросов к Веб-серверу

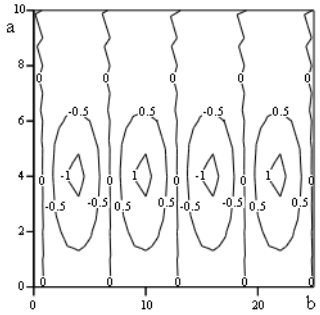


Рис. 4. Скалограмма запросов к Веб-серверу

Анализ показанных на рис. 3,4 вейвлет-спектра и скалограммы указывает на возможность реализации вейвлет-анализа статистики Internet-серверов. Однако, требуется проведение дополнительных исследований для использования полученных результатов в модели прогноза технического состояния. Кроме этого, необходимо адаптировать вейвлет-анализ к дискретному ряду данных, которым отвечает реальная статистика Internet-серверов.

## Выводы

1. С использованием теории вейвлет-преобразований разработано математическое обеспечение (2)-(10) модели расчета частотно-временных характеристик функциональных параметров Internet-серверов.

2. Перспективы дальнейших исследований в данном направлении заключаются в определении оптимальных параметров базисных вейвлетов, разработке методики интерпретации рассчитанных частотно-временных характеристик и адаптации модели к дискретным входным данным.

## Список литературы

*1. Астафьева Н.М. Вейвлет-анализ: Основы теории и примеры применения / Н.М. Астафьева // Успехи физических наук. – 1996. – т.166, № 11 – С. 1145–1170.*

*2. Михальов О.І. Вейвлет-мультифакторний аналіз складних зображень / О.І. Михальов, Ю.О.Водолазький // Вісник Вінницького політехнічного інституту. – 2009. – № 2 – С. 84–87.*

*3. Сыропятов А. А. Метод мониторинга трафика защищенных высокоскоростных коммерческих сетей нового поколения / А. А. Сыропятов // Наукові записки УНІДЗ. – 2009. – № 2(1) – С. 65–73.*

*4. Якубен М. Б. Обнаружение сетевых атак методом поиска аномалий на основе вероятностного и верификационного моделирования / М. Б. Якубен // Штучний інтелект. – 2005. –№3. – С. 679–687.*

*Статья поступила в редколлегию: 10.12.2010*

**Рецензент**: д-р техн. наук, проф. С.В. Цюцюра, Киевский национальный университет строительства и архитектуры, Киев.