УДК 681.322:519.7

[1]**Гузий Н. Н.**, канд. техн. наук,
[1]**Минаев Ю. Н.**, д-р техн. наук,
[2]**Филимонова О. Ю.**, канд. техн. наук

К ВОПРОСУ ВЫБОРА ОБОБЩЕННОЙ МОДЕЛИ ТРАФИКА КОМПЬЮТЕРНОЙ СЕТИ

[1]Институт компьютерных технологий Национального авиационного университета
[2]Киевский национальный университет строительства и архитектуры

Рассматриваются вопросы построения обобщенной модели сетевого трафика на основе представления множества параметров стандартного трафика в виде вектора, компонентами которого являются функции инвариантов тензора второго ранга. Показана возможность идентификации атаки на компьютерную сеть.

Постановка задачи в общем виде. Проблема обнаружения аномалий трафика в компьютерных сетях в последнее время привлекает серьезное внимание со стороны специалистов, что объясняется как сложностью проблемы, так и ее прикладным значением.

Проблема обнаружения вторжений в сетях *TCP/IP* сводится к задачам распознавания и практически разрешима во всех случаях при помощи анализа:

– структурных признаков (сигнатур) известных типов атак;

– инвариантных признаков структуры корректных вычислительных процессов;

– корреляционных признаков нормального функционирования распределенных вычислительных систем.

Исследование возможности применения тензорной методологии для анализа трафика компьютерной сети, в частности обнаружения аномальных режимов, представляет значительный научный и практический интерес.

Анализ результатов исследований

В работе [1] приводится анализ инженерных принципов, методик и алгоритмов обнаружения аномалий, которые могут быть полезны для построения перспективных систем защиты. Распознавание аномалий предполагает решение следующего комплекса задач:

– формирование множества эталонов инвариантов нормального (семантически корректного) развития вычислительных процессов в условиях априорной неопределенности воздействий внешней и внутренней среды;

– выбор шкал измерения признаков эталонов или инвариантов;

– обоснование необходимого и достаточного числа информативных признаков, характеризующих инварианты;

– формирование системы правил типа «*если, то*» для распознавания аномалий.

Предложенные т. н. *инварианты подобия*, представляют собой некоторые соотношения, позволяющие однозначно определить эталон "правильного" функционирования некоторой распределенной вычислительной системы. При этом вычислительный процесс семантически корректен, если соответствующая система уравнений размерностей имеет среди множества векторов-решений хотя бы один такой вектор, состоящий из всех ненулевых компонент.

В результате становится возможным предложить универсальную методику обнаружения аномалий на основе инвариантов подобия, основная идея которой заключается в распознавании аномалий вычислительных процессов по результатам сравнения значений инвариантов подобия реальных вычислительных процессов с эталонными значениями инвариантов. Данный подход позволил явно исключить шаг выделения независимых параметров вычислительного процесса на основе эвристических алгоритмов и разработать детерминированный алгоритм обнаружения аномалий.

Авторами в [1] предложены варианты разработки систем обнаружения аномалий на основе инвариантов подобия, которые базируются на том, что датчики (или сенсоры) систем обнаружения вторжений размещаются в выделенных сегментах вычислительной системы (*network-based IDS*) или на рабочих станциях сегментов (*host-based IDS).*

В последнее время для исследования телекоммуникационных сетей получили достаточно широкое применение методы интеллектуального анализа данных (*Data mining* – извлечение знаний) [2], в частности из параметров трафика, извлекать знания (в виде системы правил), позволяющие достаточно объективно идентифицировать аномальный режим.

В работе [3] предложена интеллектуальная технология идентификации атак на компьютерные сети (КС), базирующаяся на тензорном представлении трафика КС (т.н.тензор-трафик (ТТ)). Инварианты ТТ позволяют получить новое знание - достаточно объективно констатировать является ли трафик потенциально опасным с точки зрения наличия в нем атаки на сеть.

Цель работы - исследование возможности применения тензорной методологии для анализа трафика и идентификации атак на компьютерную сеть.

Решение задачи. Тензорный анализ приобрел чрезвычайно большое распространение в механике деформированной среды. Как показали исследования авторов, достижения тензорного исчисления могут быть эффективно использованы для анализа аномалий трафика и идентификации атак на компьютерные сети. Тензорное исчисление дает возможность представит тензор 2-го ранга (с матрицей 3×3) в виде ортогонального вектора, если компоненты этого вектора определены как некоторые функции собственных значений тензора. Известно, что тензор допускает несколько декомпозиций, в частности декомпозицию „собственные значения-собственные векторы", которая базируется на формуле $Tx = \lambda x$, где λ - собственные значения. Последнее выражение запишем как $(T - \lambda I)x = 0$. Пренебрегая нулевым вектором получаем характеристическое уравнение $|T - \lambda I| = 0$. Для трехмерного тензора

$$\begin{bmatrix} t_{11} - \lambda & t_{12} & t_{13} \\ t_{21} & t_{22} - \lambda & t_{23} \\ t_{31} & t_{32} & t_{33} - \lambda \end{bmatrix} = 0 .$$

Используя правило Крамера, имеем

$$\lambda^3 - I_1\lambda^2 + I_2\lambda - I_3 = 0 ,$$

где коэффициенты I_1, I_2, I_3 равны:

$$I_1 = t_{11} + t_{22} + t_{33},\ I_2 = \sum_{j=1}^{3} Co_j ,$$

$I_3 = \det(T)$, где Co_j – кофакторы:

$$Co_1 = \begin{bmatrix} t_{11} & t_{12} \\ t_{21} & t_{22} \end{bmatrix}, Co_2 = \begin{bmatrix} t_{11} & t_{13} \\ t_{31} & t_{33} \end{bmatrix}, Co_3 = \begin{bmatrix} t_{22} & t_{23} \\ t_{32} & t_{33} \end{bmatrix}.$$

Учитывая большую роль, которую играет эта декомпозиция, остановимся на ней болем детально. Условия неопределенности могут быть *разными* в разных базисах, т.е., в определенном базисе система может быть частично (или полностью) определенной. Напомним, что в контексте работы базис – это состояние исследуемого объекту (КС).

Изменение базиса. Из тензорного анализа известно, что при изменении координат (перенос, вращение и др.) компоненты тензора изменяются, хотя основные свойства тензора остаются инвариантными. Для реального симметричного тензора существует возможность такого поворота координат, что элементы тензора сокращаются к диагональной форме. Очевидно, что диагональная форма значительно более простая.

Рис.1. Главные оси тензорного эллипса

В этой связи укажем, что для основного тензора, который моделирует трафик, и присоединенных тензоров, соответствующие эллипсы различны. В частности, для тензор-переменной (основного тензора) геометрия представляет собою деформированный эллипс, т.к. существует только одно

собственное значение. Собственные векторы известны как главные оси эллипса, собственные значения отвечают длине геометрических осей. Диагонализация тензора отвечает обращению координатных осей так, что они совпадают с главными осями.

В соответствии с основными положениями тензорного анализа, коэффициентами характеристического уравнения являются *инварианты* - константы, значения которых сохраняются при изменении системы координат. Поскольку рассматривается декомпозиция типа „собственные значения - собственные векторы", укажем, что величины инвариантов могут быть определены через собственные значения следующим образом:

$$I_1 = \lambda_1 + \lambda_2 + \lambda_3,$$
$$I_2 = \lambda_1\lambda_2 + \lambda_3\lambda_2 + \lambda_3\lambda_1,$$
$$I_3 = \lambda_1\lambda_2\lambda_3.$$

Инварианты имеют название главных; в общем случае существует 8 независимых инвариантов. Отметим, что к числу главных инвариантов также относится магнитуда тензора, которая не связана с собственными значениями и определяется только на основании элементов тензора.

Анализ трафика, представляемого в виде тензора, позволяет получить дополнительные знания, прежде всего, с учетом характера собственных значений - собственных векторов. В тензорном поле, подобно критическим точкам векторных полей, существуют деформированные точки, где одно или больше собственных значений равны. Эти точки разделяют тензорное поле на области, которые базируются на топологической структуре, и определяются полем собственных векторов. Подобно критическим точкам, собственные векторы тензорного поля могут только начинаться, заканчиваться или пересекаться в деформированных точках. Анализ локального поведения собственного вектора возле этих точек определяет топологию тензорного поля.

Аналитическое определение собственных значений. Если

$$I_B = trB, II_B = \frac{1}{2}((trB)^2 - trB^2), III_B = \det B$$

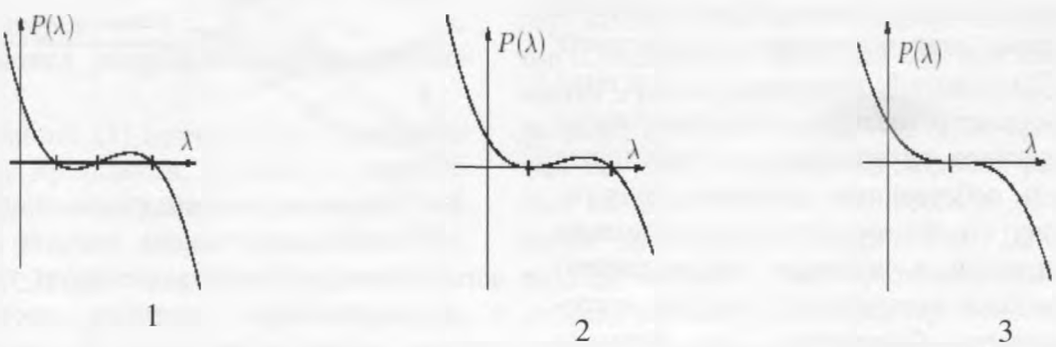
то в случае симметричного тензора B величины λ_u, u=1,2,3 аналитически могут быть определены следующим образом [4]:

$$\lambda_k = \frac{1}{3}(I_B + 2\sqrt{I_B^2 - 3II_B}\cos\frac{\beta + (k-1)2\pi}{3}), k = 1,2,3;$$

$$\beta = \arccos\frac{2I_B^3 - 9I_B II_B + 27III_B}{2\sqrt{(I_B^2 - 3II_B)^3}}.$$

Положительно определенные тензоры имеют только положительные собственные значения и три соответствующих случая приведены на рис.2. Аналитическое вычисление 2-го инварианта девиатора тензора B, равного $B^D = B - \frac{1}{3}(trB)I$ имеет вид

$$J_2(B) = \frac{1}{2}((trB^D) - B^D \cdot B^D) = -\frac{1}{2}B^D \cdot B^D = -\frac{1}{2}(I_B^2 - 3II_B) \le 0$$



1 2 3

Рис.2. Кривые характеристических полиномов : 1- собственные значения $\lambda_1 \neq \lambda_2 \neq \lambda_3$; 2- собственные значения $\lambda_1 = \lambda_2 \neq \lambda_3$; 3- собственные значения $\lambda_1 = \lambda_2 = \lambda_3 = \lambda$

Главный тензор становится сферическим тензором с равными собственными значениями. Точки означают внутреннее произведение двух тензоров второго порядка. В случае двух равных собственных значений третий инвариант девиатора имеет вид

$$J_3(B) = \det B^D = \frac{1}{27}(2 I_B^3 - 9 I_B I_B + 27 III_B$$

или

$$27 J_3 \approx 2(-3 J_2)^{3/2}.$$

Трафик описывается системой 9 независимых величин, которая дает формальную возможность представить его как двухвалентный тензор. Тензоры четных валентностей могут быть применены для анализа и защиты КС в условиях неопределенности. Для двухвалентных тензоров в декартовой системе координат возможно матричное представление, поэтому, если говорим в тензоре в рамках данной статьи, имеем в виде квадратную матрицу 3*3. В дальнейшем изложении основные определения приведены в соответствии с работой [5].

В механике тензор напряжений представляет собой метрический тензор (единичная матрица в нашем случае), с точностью до скалярного множителя. Такой тензор называется шаровым (радиус шара метрического тензора равняется 1/3, в общем случае этот радиус равняется трети величины следа тензора). След тензора *T* (двойная свертка | линейный инвариант | первый инвариант | *Sp T*) - для матричного представления - сумма диагональных элементов, в механике имеет смысл утроенного гидростатического давления, в нашем случае в зависимости от избранной системы учета параметров трафика он может иметь другое содержание, которое можно представить в каждом случае отдельно. В частности, это могут быть численные характеристики: длина передаваемого пакета, количество передаваемых пакетов или др.

Если вычтем из тензора его шаровой тензор, то получим объект, известный как девиатор тензора, т.е. всегда существует разложение тензора на шаровую компоненту и девиаторную компоненту. Укажем, что семантически шаровая часть тензора интерпретируется как его неизменная часть, а девиатор характеризует изменения (вариабельность). Это чрезвычайно важный элемент анализа, т. к. при анализе трафика шаровая часть может характеризовать нормальное состояние, а девиатор присутствие атаки или наличие аномалий.

Из тензорного исчисления известно, что инварианты тензора могут быть представлены как определенные функции вот главных значений (собственных векторов). Приведем некоторые соотношения между разными системами инвариантов, введя такую систему обозначений: *S1, S2, S3* главные значения тензора *T, Dv* – девиатор тензора, используя возможности пакета символьной математики *Mathematica.* Эти соотношения рационально использовать при идентификации атак, сводя задачу к задаче классификации.

Тензор, записанный через инварианты - главные значения, можно представить в виде:

$$T = \begin{pmatrix} S1 & 0 & 0 \\ 0 & S2 & 0 \\ 0 & 0 & S3 \end{pmatrix}.$$

В свою очередь, девиатор может быть соответственно (учетом выражения для *T)* определен из соотношения для его квадрата следа:

$$Tr[Dv \cdot Dv] = \left(\frac{1}{3}(S1-S2)^2 + (S1-S3)^2 + (S3-S1)^2\right).$$

Из последнего выражения девиатор можно записать следующим образом:

$$Dv = T - \frac{1}{3}Tr[T] \cdot \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Воспользуемся еще одним положением, нашедшем широкое применение в механике, утверждающем, что тензор удобно изображать геометрически в трехмерном декартовом пространстве главных напряжений. В этом случае, каждый тензор может быть представлен как трехмерный вектор. Исследования показали, что скалярное произведение для ранее полученных в виде векторов значений для девиаторной и шаровой компонент имеет нулевой результат. Это является свидетельством ортогональ-

ности шаровой и девиаторной частей тензора, представленных как вектора в пространстве главных напряжений. Т. к. главные значения, или собственные числа матрицы, являются инвариантами, то и любые комбинации из их будут инвариантами. В качестве инвариантов предложено выбрать функции:

$$\sigma = \frac{1}{3}\mathrm{Tr}[\mathrm{T}] = \frac{1}{3}(S1 + S2 + S3),$$

$$\tau = \left(\frac{1}{2}\left(\mathrm{Tr}[\mathrm{T}\cdot\mathrm{T}] - \frac{1}{3}Tr[T]^2\right)\right)^{1/2} = \left(\frac{1}{2}(\mathrm{Tr}[\mathrm{Dv}\cdot Dv])\right)^{1/2} = \left(\frac{1}{6}\left((\mathrm{S1}-\mathrm{S2})^2 + (\mathrm{S1}-\mathrm{S3})^2 + (\mathrm{S2}-\mathrm{S3})^2\right)\right)^{1/2}.$$

Смысл в рассматриваемом случае естественно должен меняться в зависимости вот выбора системы отсчета параметров трафика.

Рассмотрим возможности, предоставляемые «плоской» моделью трафика. В этом случае возможность формулирования системы правил «*если A, то B, иначе C*» с целью идентификации и анализа трафика. Обращаясь к изложенному выше, отмечаем, что каждому напряженному состоянию в двухмерном пространстве инвариантов (σ, τ) (в нашем случае трафику) будет соответствовать точка на плоскости. Если в трехмерном пространстве трафика известна некоторая точка, соответствующая тензору, то всегда можно вычислить ее проекцию на гидростатическую вот через скалярное произведение на вектор *n*. С другой стороны через это же скалярное произведение можем определить угол между вектором *n* и вектором трафика

$$\frac{T\cdot n}{\sqrt{T\cdot T}} = \cos\alpha\,.$$

Следовательно, можно нанести точку на плоскость (σ, τ), или указать направление, на котором будут находиться все точки для трафика в определенном состоянии – атака или ее отсутствие, определенного с точностью к меченный параметров трафика.

Приведем также другие формы представления инвариантов тензора, которые целесообразно использовать в последующем анализе и формулировании правил:
длина обобщенного вектора, характеризующего трафик:

$$l = (pg^2 + pd^2)^{1/2},$$

где $pg = \left(\frac{1}{\sqrt{3}}\ \frac{1}{\sqrt{3}}\ \frac{1}{\sqrt{3}}\right)\cdot T\cdot\begin{pmatrix}\frac{1}{\sqrt{3}}\\ \frac{1}{\sqrt{3}}\\ \frac{1}{\sqrt{3}}\end{pmatrix}\cdot[[1,\ 1]]$, $\mathrm{pd} = \left(\frac{1}{2}(Tr[T\cdot T] - \frac{1}{3}Tr[T]^2)\right)^{1/2}$.

Проведенные исследования показали высокую эффективность применения указанных параметров в качестве обобщенной модели трафика. Рассмотрим конкретные представления ТТ для ситуаций, когда трафик имеет вид тензора второго ранга [T_{ij}]:

$$\mathrm{T} = \{T_{ij}\} = \begin{pmatrix} T_{11} & T_{12} & T_{13} \\ T_{21} & T_{22} & T_{23} \\ T_{31} & T_{32} & T_{33} \end{pmatrix},$$

где T_{ij} - отдельные компоненты тензора - параметры трафика, т.е. $T_{11} = \boldsymbol{x_1}$, $T_{12}=\boldsymbol{x_2}$, … и т.д.,

Для проверки эффективности нейросетевых технологий сигнатуры атак определялись согласно [2] и включали такие параметры: $\mathbf{x} = \{x_i\}$, i =1,9**:**

x_1 – *Protocol ID* протокол, связанный с событием *(TCP=0, UDP=1, ICMP=2, unknown=3);*

x_2 – номер порта источника;

x_3 – номер порта хоста назначения;

x_4 – *IP*-адреса источника;

x_5 – *IP*-адреса приемника;

x_6 – *ICMP Type* тип *ICMP*-пакет *(Echo Request or Null);*

x_7 – *ICMP Code* кодовое поле из *ICMP*-пакета *(None or Null);*

x_8 – *Raw Data Length* длина данных в пакете;

x_9 – *Raw Data* порция данных в пакете.

Базовые векторы (“атака присутствует” - “атака отсутствует”) приведены в табл.1

Таблица 1

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	y
0	2314	80	1573638018	-1580478590	1	1	401	3758	0, 1
0	1611	6101	8801886082	-926176166	1	1	0	2633	1, 0

Собственные значения ТТ «атака - отсутствие атаки» составляют:

$\lambda 1$ =1.0e+009 * { -1.5805, 0.0000 0.0000}, $\lambda 2$= 1.0e+008 * {-9.2619, 0.0002 0.0000}. Координаты соответствующих векторов: *sigma*1=-5.2682e+008, *tau*1= =4.1632e+017; *sigma*2=3.0872e+008, *tau*2 = 1.4297e+017. На рис. 3. приведены графические отображения ТТ «атака - отсутствие атаки» и их векторные аналоги, построенные на базе систем ортогональных инвариантов (*sigma*, *tau*). Из представленных графических изображений видно, насколько эффективно использование векторных аналогов ТТ, построенные на базе систем ортогональных инвариантов (*sigma*, *tau*) для идентификации аномальных режимов КС.

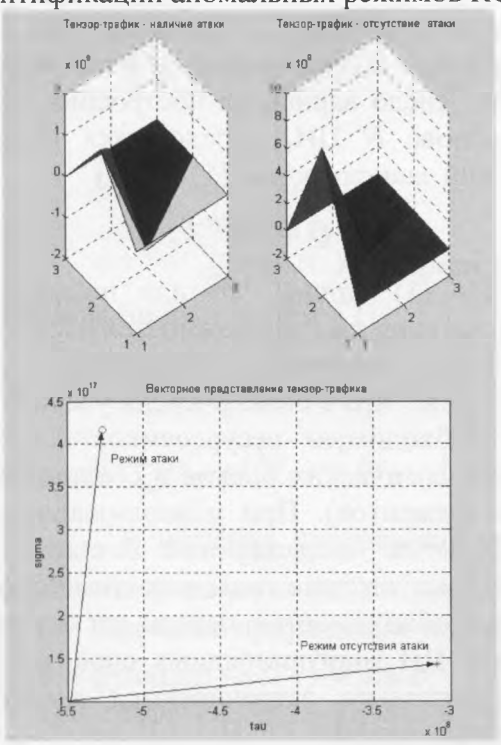


Рис.3. Векторные аналоги ТТ, построенные на базе систем ортогональных инвариантов (*sigma*, *tau*)

Выводы.

1. Трафик компьютерной сеты, представляемый в виде совокупности параметров (x_1 – *Protocol ID*; x_2 – номер порта источника; x_3 -порта хоста назначения; x_4 – *IP*-адрес источника; x_5 – *IP*-адрес приемника; x_6 – *ICMP Type* тип *ICMP*-пакет; x_7 – *ICMP Code* кодовое поле из *ICMP*-пакета; x_8 – *Raw Data Length* длина данных в пакете; x_9 – *Raw Data* порция данных в пакете) может быть представлен обобщенным параметром, в качестве которого рационально использовать ортогональные инварианты тензор-трафика, вычисленные на основе собственных значений.

2. Представление трафика КС в виде обобщенного параметра позволяет существенно повысить эффективность идентификации аномальных режимов КС.

Литература

1. *А.Беляев, С.Петренко* Системы обнаружения аномалий: новые идеи в защите информации. - Экспресс-Электроника, №2, 2004.

2. *Sushmita Mitra, Sankar K. Pal, Pabitra Mitra* Data Mining in Soft Computing Framework: A Survey. - IEEE TRANSACTIONS ON NEURAL NETWORKS, VOL. 13, NO. 1, JANUARY 2002 p.

3. *Ю.Н.Минаев, О.Ю.Филимонова, Н.Н.Гузий.* Интеллектуальные технологии в системах идентификации и прогнозирования атак на компьютерные сети. Электронное моделирование, № 6 , 2005. - 12 с.

4. *S. Hartmann* Computational Aspects of the Symmetric Eigenvalue Problem of Second Order Tensors. - TECHNISCHE MECHANIK, Band 23, Heft 2-4, (2003), 283 – 294. Manus-kripteingang: 15. August 2003.

5. *Ф.Пинежанинов* Математика в механической прочности. Интернет-ресурс: *http://pinega.da.ru/.*