УДК 004.7.056.53

ОПЕРАТОРНОЕ ПРЕДСТАВЛЕНИЕ ПРОЦЕДУРЫ ИЗВЛЕЧЕНИЯ ДАННЫХ О СЕССИЯХ ТРАФИКА ПО ДАННЫМ СЕРВЕРА УДАЛЕННОГО ВЗАИМОДЕЙСТВИЯ (RSH) И СИМУЛЯТОРА СЧЕТЧИКА ПАКЕТОВ CISCO IP ACCOUNTING (IPCAD) НА ТРАНЗИТНОМ LINUX МАРШРУТИЗАТОРЕ

А. Ю. Телков, Е. А. Крохин

Воронежский государственный университет

Поступила в редакцию 25.04.2017 г.

Аннотация. В статье рассмотрено операторное представление процедуры извлечения данных о сессиях одиночных хостов с привлечением элементов линейной алгебры и элементов языка SQL из структуры, соответствующей базе данных, формирующейся и обновляющейся в результате работы счетчика пакетов Cisco IP Accounting (IPCAD) на транзитном маршрутизаторе с операционной системой LINUX.
Ключевые слова: SQL, сервер удаленного взаимодействия (RSH), Cisco IP Accounting.
Annotation. The article deals with the operator representation of the procedure for extracting data about sessions of single hosts with the use of elements of linear algebra and elements of the SQL language from the structure corresponding to the database that is generated and updated as a result of the work of the Cisco IP Accounting packet counter (IPCAD) on a transit router with The LINUX operating system.
Keywords: SQL, Remote Interaction Server (RSH), Cisco IP Accounting.

ВВЕДЕНИЕ

При возможности детального анализа трафика информационных систем на аномальность, положительные результаты дает анализ трафика на уровне сессий сетевого взаимодействия [1–4]. Поскольку одиночные сетевые узлы (хосты) могут одновременно участвовать в нескольких сеансах сетевого обмена, в некоторый фиксированный момент времени на каждом из хостов может быть активно множество сессий обмена данными с разными узлами назначения по одному или нескольким протоколам сетевого взаимодействия. Поэтому, чтобы исследовать аномальность сетевого поведения узла при его обмене данными по некоторому прикладному протоколу на интервале наблюдения, требуется:

во-первых, выяснить номер порта на транспортном уровне, который закреплен за этим протоколом;

во-вторых, определить сеансы обмена узла с другими узлами, в том числе идентифицировать эти узлы;

в-третьих, выделить сессии, соответствующие сеансам обмена исследуемого узла с этими узлами;

в-четвертых, для каждой сессии исследовать поведение узла на аномальность с помощью некоторого алгоритма.

Например, чтобы исследовать взаимодействие узла с участниками сетевого обмена по протоколу https, нужно принять во внимание, что сервис https работает на 443 порту TCP, найти все сессии, которые начинаются на тестируемом узле и заканчиваются на других узлах на порту 443 TCP, найти все эти другие узлы и исследовать каждый сеанс обмена тестируемого узла с удаленными узлами для всех сессий.

Целью данной работы является представление в операторном [5] виде процедуры извлечения данных о сессиях некоторого хоста по информации, получаемой с транзитного

маршрутизатора, через который проходит сетевой трафик этого узла, с помощью счетчика пакетов Cisco IP Accounting (IPCAD) [6, 7] и сервера удаленного взаимодействия (RSH).

ОПЕРАТОРНОЕ ПРЕДСТАВЛЕНИЕ ИЗВЛЕЧЕНИЯ ДАННЫХ О СЕССИЯХ

Сессию принято [4] идентифицировать набором пары сокетов (сокет – это IP адрес + порт) источника и назначения, т. е. IP_s, $Port_s$, IP_d, $Port_d$ и указывать при этом протокол взаимодействия (TCP, UDP и т. п.). Рассматривая трафик в рамках сессии как функцию количества передаваемых данных S^D от времени t, т. е. $S^D(t)$, сессию на интервале времени $[T_b;T_e]$, $T_e > T_b$ можно описать вектором

$$S^D = [S_1^D, S_2^D, S_3^D, \ldots, S_M^D]^T, \qquad (1)$$

разбив интервал наблюдения $t \in [T_b;T_e]$ на M отрезков с шагом $\Delta t = \frac{T_e - T_b}{M}$.

При этом элементы вектора S^D такие что,

$$S_i^D = \begin{matrix} t = T_b + (i+1)\Delta t \\ t = T_b + i\Delta t \end{matrix} \Bigg| \sum_{k=1}^{K_i} S_{i,k}^D, \quad i = 0,1,\ldots,(M-1),$$

то есть каждый элемент S_i^D представляют собой сумму размеров $S_{i,k}^D$ (либо количества) пакетов, переданных в рамках данной сессии в i-м интервале времени Δt. Сам вектор (1) представляет собой гистограмму трафика на временной оси, и чем меньше интервал Δt, тем ближе эта гистограмма к реальной функции $S^D(t)$.

В ряде случаев, например, при сетевом взаимодействии приложений наблюдаемого хоста с удаленным узлом по протоколу TCP на транспортном уровне, помимо сессии в прямом направлении, существует сессия в обратном направлении, характеризующаяся IP_d, $Port_d$, IP_s, $Port_s$. Указывая на первом месте адрес и порт назначения, мы подразумеваем обратное направление передачи данных. По аналогии с (1), сессию в обратном направлении можно описать вектором

$$S^I = [S_1^I, S_2^I, S_3^I, \ldots, S_M^I]^T, \qquad (2)$$

разбивая тот же интервал наблюдения $t \in [T_b;T_e]$ на M отрезков с шагом $\Delta t = \frac{T_e - T_b}{M}$.

При этом элементы вектора S^I такие что,

$$S_i^I = \begin{matrix} t = T_b + (i+1)\Delta t \\ t = T_b + i\Delta t \end{matrix} \Bigg| \sum_{k=1}^{K_i} S_{i,k}^I, \quad i = 0,1,\ldots,(M-1),$$

то есть каждый элемент S_i^I представляют собой сумму размеров $S_{i,k}^I$ (либо количества) пакетов, переданных в рамках обратной сессии в i-м интервале времени Δt. Сам вектор (2) представляет собой гистограмму трафика обратной сессии на временной оси, и чем меньше интервал Δt, тем ближе эта гистограмма к реальной функции $S^I(t)$.

При работе счетчика пакетов Cisco IP Accounting (IPCAD) запуская механизм подсчета на транзитном маршрутизаторе на i-м интервале времени, через время Δt будем иметь данные, которые можно представить в виде матрицы $\mathbf{B}_i$. Размерность матрицы $Q_i \times 7$, где количество строк $Q_i \sum_{z=1}^{Z} S_z$ определяется количеством хостов Z, на которых в i-й интервал времени Δt существует трафик через рассматриваемый транзитный маршрутизатор, причем на каждом хосте S_z сессий, $z = 1 \ldots Z$.

$$\mathbf{B}_i = [\mathbf{A}_i^s, \mathbf{P}_i^s, \mathbf{A}_i^d, \mathbf{P}_i^d, \mathbf{N}_i, \mathbf{V}_i, \mathbf{R}_i], \qquad (3)$$

где $\mathbf{A}_i^s = [A_{i,1}^s, A_{i,2}^s, \ldots, A_{Q_i}^s]^T$ – вектор-столбец IP адресов источников,

$\mathbf{P}_i^s = [P_{i,1}^s, P_{i,2}^s, \ldots, P_{Q_i}^s]^T$ – вектор-столбец номеров портов источников,

$\mathbf{A}_i^d = [A_{i,1}^d, A_{i,2}^d, \ldots, A_{Q_i}^d]^T$ – вектор-столбец IP адресов назначения,

$\mathbf{P}_i^d = [P_{i,1}^d, P_{i,2}^d, \ldots, P_{Q_i}^d]^T$ – вектор-столбец номеров портов назначения,

$\mathbf{N}_i = [N_{i,1}, N_{i,2}, \ldots, N_{Q_i}]^T$ – вектор-столбец количества пакетов,

$\mathbf{V}_i = [V_{i,1}, V_{i,2}, \ldots, V_{Q_i}]^T$ – вектор-столбец сумм размеров пакетов,

$\mathbf{R}_i = [R_{i,1}, R_{i,2}, \ldots, R_{Q_i}]^T$ – вектор-столбец кодов протокола взаимодействия.

Получив матрицы $\mathbf{B}_i$ для M интервалов наблюдения Δt запишем

$$\boldsymbol{\psi} = \begin{bmatrix} \begin{bmatrix} \mathbf{W}_0 \\ \mathbf{W}_1 \\ \ldots \\ \mathbf{W}_{M-1} \end{bmatrix} & \begin{matrix} \mathbf{B}_0 \\ \mathbf{B}_1 \\ \ldots \\ \mathbf{B}_{M-1} \end{matrix} \end{bmatrix}, \qquad (4)$$

где составной вектор $\mathbf{W} = [\mathbf{W}_0, \mathbf{W}_1, \ldots, \mathbf{W}_{M-1}]^T$ образован векторами $\mathbf{W}_i = [\mathbf{W}_{i,1}, \mathbf{W}_{i,2}, \ldots, \mathbf{W}_{i,Q_i}]^T$

$$\begin{bmatrix} \begin{bmatrix} W_{0,1} \\ W_{0,2} \\ \cdots \\ W_{0,Q_0} \end{bmatrix} & \begin{bmatrix} A^s_{0,1} \\ A^s_{0,2} \\ \cdots \\ A^s_{0,Q_0} \end{bmatrix} & \begin{bmatrix} P^s_{0,1} \\ P^s_{0,2} \\ \cdots \\ P^s_{0,Q_0} \end{bmatrix} & \begin{bmatrix} A^d_{0,1} \\ A^d_{0,2} \\ \cdots \\ A^d_{0,Q_0} \end{bmatrix} & \begin{bmatrix} P^d_{0,1} \\ P^d_{0,2} \\ \cdots \\ P^d_{0,Q_0} \end{bmatrix} & \begin{bmatrix} N_{0,1} \\ N_{0,2} \\ \cdots \\ N_{0,Q_0} \end{bmatrix} & \begin{bmatrix} V_{0,1} \\ V_{0,2} \\ \cdots \\ V_{0,Q_0} \end{bmatrix} & \begin{bmatrix} R_{0,1} \\ R_{0,2} \\ \cdots \\ R_{0,Q_0} \end{bmatrix} \\ \begin{bmatrix} W_{1,1} \\ W_{1,2} \\ \cdots \\ W_{1,Q_1} \end{bmatrix} & \begin{bmatrix} A^s_{1,1} \\ A^s_{1,2} \\ \cdots \\ A^s_{1,Q_0} \end{bmatrix} & \begin{bmatrix} P^s_{1,1} \\ P^s_{1,2} \\ \cdots \\ P^s_{1,Q_0} \end{bmatrix} & \begin{bmatrix} A^d_{1,1} \\ A^d_{1,2} \\ \cdots \\ A^d_{1,Q_0} \end{bmatrix} & \begin{bmatrix} P^d_{1,1} \\ P^d_{1,2} \\ \cdots \\ P^d_{1,Q_0} \end{bmatrix} & \begin{bmatrix} N_{1,1} \\ N_{1,2} \\ \cdots \\ N_{1,Q_1} \end{bmatrix} & \begin{bmatrix} V_{1,1} \\ V_{1,2} \\ \cdots \\ V_{1,Q_1} \end{bmatrix} & \begin{bmatrix} R_{1,1} \\ R_{1,2} \\ \cdots \\ R_{1,Q_1} \end{bmatrix} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \begin{bmatrix} W_{M-1,1} \\ W_{M-1,2} \\ \cdots \\ W_{M-1,Q_{M-1}} \end{bmatrix} & \begin{bmatrix} A^s_{M-1,1} \\ A^s_{M-1,2} \\ \cdots \\ A^s_{M-1,Q_{M-1}} \end{bmatrix} & \begin{bmatrix} P^s_{M-1,1} \\ P^s_{M-1,2} \\ \cdots \\ P^s_{M-1,Q_{M-1}} \end{bmatrix} & \begin{bmatrix} A^d_{M-1,1} \\ A^d_{M-1,2} \\ \cdots \\ A^d_{M-1,Q_{M-1}} \end{bmatrix} & \begin{bmatrix} P^d_{M-1,1} \\ P^d_{M-1,2} \\ \cdots \\ P^d_{M-1,Q_{M-1}} \end{bmatrix} & \begin{bmatrix} N_{M-1,1} \\ N_{M-1,2} \\ \cdots \\ N_{M-1,Q_{M-1}} \end{bmatrix} & \begin{bmatrix} V_{M-1,1} \\ V_{M-1,2} \\ \cdots \\ V_{M-1,Q_{M-1}} \end{bmatrix} & \begin{bmatrix} R_{M-1,1} \\ R_{M-1,2} \\ \cdots \\ R_{M-1,Q_{M-1}} \end{bmatrix} \end{bmatrix} \quad (5)$$

$i = 0,1,\ldots,(M-1)$, где каждый элемент играет вспомогательную роль для идентификации любой из строк матрицы $\boldsymbol{\psi}$.

В развернутом виде матрица $\boldsymbol{\psi}$ представлена формулой (5) и имеет размерность $X \times 8$, где $X = \sum_{i=0}^{M-1} Q_i$.

В матрице (5) содержится избыточная по отношению к исходно поставленной задаче информация. Чтобы для каждого исследуемого узла из J возможных с IP адресами A^P_j, $j = 1,2,\ldots,J$ решить поставленную задачу нужно:

1. Получить список сессий исследуемого j-го узла, $j = 1,2,\ldots,J$ для заданного протокола, определенного портом назначения P^p и кодом протокола, в виде

$$\mathbf{C} = F\{\boldsymbol{\psi}\} \quad (6)$$

$\mathbf{C} = [C_1, C_2, \ldots, C_U]^T$ – вектор столбец с отобранными (с исключением повторяющихся) элементами векторов $\mathbf{P}^s_i = [P^s_{i,1}, P^s_{i,2}, \ldots, P^s_{Q_i}]$, $i = 0,1,\ldots,N-1$, для которых оператор $\mathbf{F}$ определяет вхождение элемента в результирующий вектор-столбец $\mathbf{C}$ при выполнении двух условий: $P^d_{i,q_i} = P^p$, $i = 0,1,\ldots,N-1$; $q_i = 1,2,\ldots,Q$ и $A^s_{i,q_i} = A^p_j$, $i = 0,1,\ldots,N-1$; $q_i = 1,2,\ldots,Q$.

2. Для каждой сессии, то есть для каждого элемента C_u, $u = 1,2,\ldots,U$ получить вектор ненулевых значений счетчиков в виде:

$$S^{D^*} = H^D_{C_u}\{\boldsymbol{\psi}\},\ S^{I^*} = H^I_{C_u}\{\boldsymbol{\psi}\} \quad (7)$$

$S^{D^*} = [S^{D^*}_1, S^{D^*}_2, \ldots, S^{D^*}_{m^D}]^T$, где $m^D < M$ так сессия может существовать только на части отрезков Δt общего интервала наблюдения $t \in [T_b; T_e]$ и $S^{I^*} = [S^{I^*}_1, S^{I^*}_2, \ldots, S^{I^*}_{m^I}]^T$, где $m^I < M$ по той же причине что и выше (но для сессии в обратном направлении).

3. Отобразить полученные элементы вектора S^{D^*} в элементы вектора S^D при совпадающих временных интервалах, в отсутствии соответствия элементы вектора S^D заполнить нулями.

$$S^D = G^D\{S^{D^*}\},\ S^I = G^I\{S^{I^*}\}. \quad (8)$$

4. Окончательно, можно написать в виде:

$$S^D = G^D\left\{H^D_{C_u}\{\boldsymbol{\psi}\}\right\},\ S^I = G^I\left\{H^I_{C_u}\{\boldsymbol{\psi}\}\right\},\quad \mathbf{C} = F\{\boldsymbol{\psi}\},\ u = 1,2,\cdots,U, \quad (9)$$

где S^D и S^I – искомые векторы-столбцы с количеством элементов M, определенные нами ранее формулами (1) и (2) для найденной u-й сессии, $u = 1,2,\ldots,U$; G^D, G^I – некоторые операторы отображения, $H^D_{C_u}$, $H^I_{C_u}$ – операторы, извлекающие строки с данными из матрицы $\boldsymbol{\psi}$, определенной нами ранее формулами (4) и (5) для u-й сессии в прямом и обратном направлении, соответственно. Конкретный вид операторов, о которых здесь идет речь в случае, когда представление матрицы $\boldsymbol{\psi}$ реализуется в виде таблицы базы данных, будет приведен ниже в конкретных примерах.

НЕКОТОРЫЕ ЧАСТНЫЕ РЕЗУЛЬТАТЫ И КОНКРЕТНЫЙ ВИД ОПЕРАТОРОВ

На практике для хранения элементов матрицы ψ на Linux маршрутизаторе с операционной системой CentOS использована СУБД PostgreSQL. Структура БД показана в табл. 1, пример данных в табл. 2.

В табл. 2. показана структура, которая соответствует матрице (5), отличие заключается лишь в том, что строки БД формируются с момента старта механизма записи информации в БД до момента останова. Столбец с названием «code» в БД соответствует вспомогательному вектору из формулы (4), элементы которого с целью обеспечения уникальных значений сформированы так: первые четыре цифры – год, далее 2 цифры – месяц, далее 2 цифры – день, далее 2 цифры – час, далее 2 цифры – минуты, затем 2 цифры секунды. Последние 5 цифр необходимы для хранения номера сессии (может быть от 0 до 65 536), присвоенной в рамках интервала записи. Столбец БД с названием «sourceip» соответствует вектору-столбцу $\mathbf{A}_i^s$, столбец БД с названием «sourceport» соответствует вектору-столбцу $\mathbf{P}_i^s$, столбец БД с названием «destip» соответствует вектору-столбцу $\mathbf{A}_i^d$, столбец БД с названием «destport» соответствует вектору-столбцу $\mathbf{P}_i^d$, столбец БД с названием «packets» соответствует вектору-столбцу $\mathbf{N}_i$, столбец с названием «bytes» соответствует вектору-столбцу $\mathbf{V}_i$, код протокола, который описывается с помощью вектора-столбца $\mathbf{R}_i$ в (3), в БД представлен столбцом с названием «protocol».

Таблица 1

Имя поля	Тип
code	bpchar(19)
sourceip	bpchar(15)
sourceport	bpchar(15)
destip	bpchar(15)
destport	bpchar(15)
packets	int4
bytes	int4
protocol	bpchar(5)
interface	bpchar(5)
info	bpchar(16)
ipcdate	date
ipctime	interval(2147418114)

Теперь разберем конкретный вид операторов. Чтобы получить список клиентских портов, соответствующих https сессиям узла с IP адресом 10.227.11.78 за 5 декабря в интервале с 10:00 до 10:02 нужно выполнить следующий вопрос к БД:

```
SELECT DISTINCT sourceport
FROM ipcaddump
WHERE((ipcdate='20161205') AND
      (ipctime>='10:00:00') AND
      (ipctime<='10:02:00') AND
      (sourceip='10.227.11.78') AND
      (destport='443'))
```
(10)

Запрос (10) как раз показывает, что должен сделать оператор **F** из (6). В нашем примере он вернул следующий список:

```
55333 55356 55189 49624 49261 55331
49303 54251 55359 55350 55353
```
(11)

Таблица 2

code	sourceip	sourceport	destip	destport	packets	bytes	protocol
2016120510000000000	10.227.11.75	137	10.227.11.255	137	1	78	17
2016120510000000001	10.227.11.75	138	10.227.11.255	138	1	202	17
2016120510000000003	10.227.11.246	50588	217.69.141.152	443	1	41	6
2016120510000000005	10.227.11.86	45491	93.158.134.119	443	1	41	6
2016120510000000007	10.227.11.86	45517	217.69.136.176	443	2	1905	6
2016120510000000008	10.227.11.57	49034	192.168.67.11	10050	1	60	6
2016120510000000009	10.227.11.2	53831	10.8.2.233	445	1	48	6

Теперь информацию о каждой сессии можно извлечь запросами такого вида, это пример действия оператора $\mathbf{H}^{D}_{C_u}$ из (7):

```
SELECT*
FROM ipcaddump
WHERE ((ipcdate='20161205') AND
       (ipctime>='10:00:00') AND
       (ipctime<='10:02:00') AND
       (sourceip='10.227.11.78') AND
       (destport='443') AND
       (sourceport='55189 '))
```

(12)

А информацию об инверсной сессии (трафик в обратном направлении – от узла назначения к узлу-источнику) извлечь так, это пример действия оператора $\mathbf{H}^{I}_{C_u}$ из (7):

```
SELECT*
FROM ipcaddump
WHERE ((ipcdate='20161205') AND
       (ipctime>='10:00:00') AND
       (ipctime<='10:02:00') AND
       (destip='10.227.11.78') AND
       (destport='55189 ') AND
       (sourceport='443'))
```

(13)

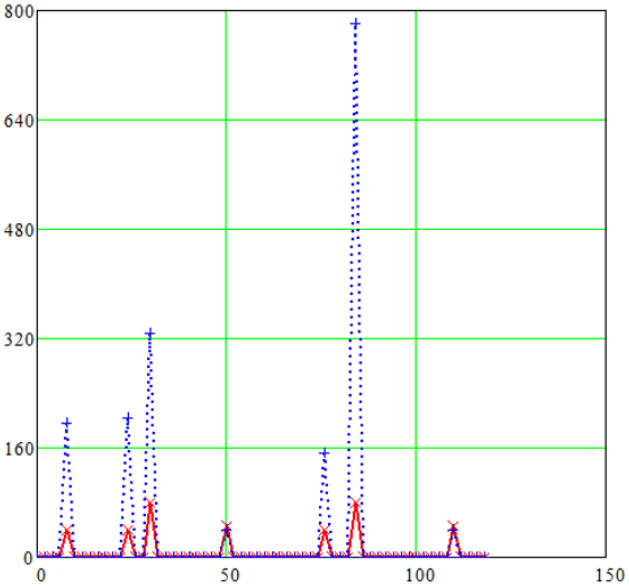

Рис. 1. HTTPS трафик (количество байт от времени)

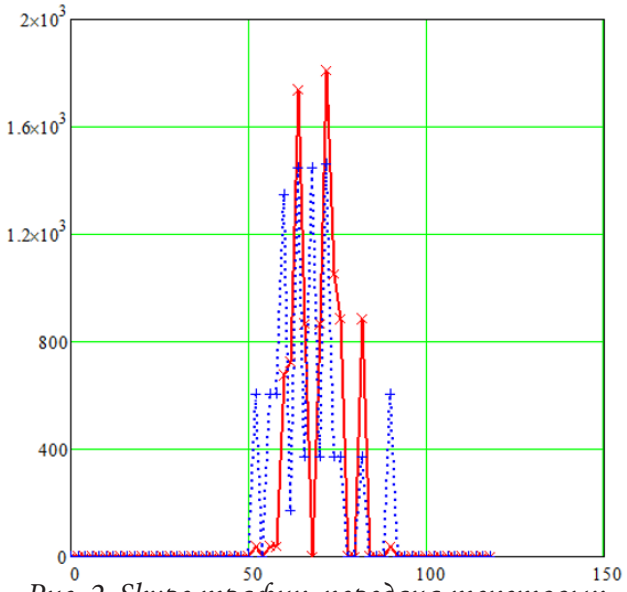

Рис. 2. Skype трафик, передача текстовых сообщений (кол-во байт от времени)

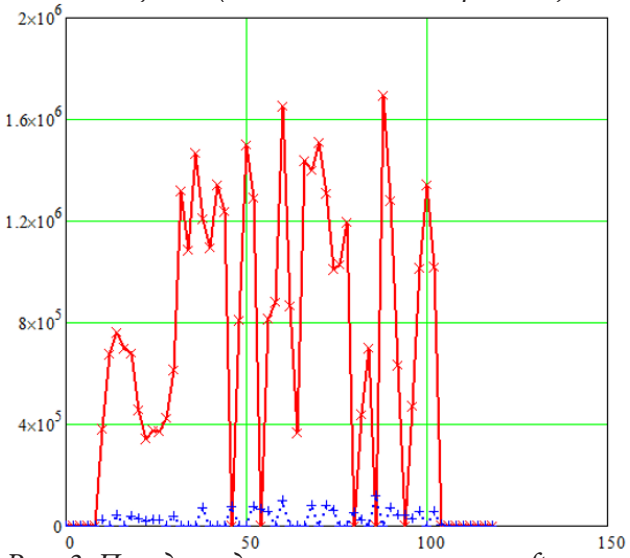

Рис. 3. Передача данных по протоколу ftp внутри OpenVPN тоннеля (кол-во байт от времени)

В указанном интервале времени существовала следующая сетевая активность на узле: обычный http трафик, сессия skype, в рамках которой шел обмен текстовыми сообщениями, и, кроме этого, была организована передача данных в исходящем направлении по протоколу ftp внутри OpenVPN тоннеля, установленного на 443 порт VPN сервера (маскировка под https). Интервал записи данных о сессиях до очередного обнуления счетчика пакетов составлял 2 секунды. Графики для указанных сессий показаны на рис. 1–3, на всех этих рисунках сплошная линия – исходящий трафик.

ЗАКЛЮЧЕНИЕ

В работе рассмотрен подход, позволяющий представить извлечение данных о сессиях одиночных хостов в операторном виде. Исходная информация для последующего представления может быть получена с помощью сервера удаленного взаимодействия (rsh) и эмулятора счетчика пакетов Cisco IP Accounting Daemon (ipcad) с транзитного маршрутизатора. Приведены общие операторные отношения, конкретный вид операторов для частных случаев, развернутый вид матрицы для представления данных, а также примеры операторов получения данных прямых и обратных сессий одиночных хостов. Применение предложенного подхода, позволяет, кроме этого, получать зависимости для трафика сессий в виде гистограмм (зависимость объема переданных данных от времени).

Рассмотренный подход может найти свое применение в ходе создания новых алгоритмов мониторинга сетевого трафика и обнаружения сетевых аномалий.

СПИСОК ЛИТЕРАТУРЫ

1. *Деарт В. Ю.* Статистические характеристики трафика современного провайдера доступа в сеть Интернет / В. Ю Деарт, В. А. Маньков, А. В. Пилюгин // T-Comm. – 2008. – №4. – С. 54–57.

2. *Joachim Charzinski*. HTTP/TCP connection and flow characteristics. / Charzinski Joachim // Performance evaluation. – 2000. – № 42. – С. 149–162.

3. *Choi H.* A Behavioral Model of Web Traffic / H. Choi, J. Limb // International Conference of Networking Protocol'99. – 1999.

4. *Xie Y.* Modelling Web Session for Detecting Pseudo HTTP Traffic / Y. Xie, X. Huang, C. Tang // Journal of Computers. – 2013. – Vol. 8, № 2. – С. 341–348.

5. *Прэтт У.* Цифровая обработка изображений: в 2-х томах / У. Прэтт // Пер. с англ.– Москва : Мир, 1982. – Т. 1.

6. A Technical Overview [Электронный ресурс] // Introduction to Cisco IOS NetFlow. – 2012. – Режим доступа: http://www.cisco.com/c/en/us/products/collateral/ios-nx-os-software/ios-netflow/prod_white_paper0900aecd80406232.html.

7. Cisco Systems NetFlow Services Export Version 9 [Электронный ресурс] // IETF. – 2004. – Режим доступа: https://www.ietf.org/rfc/rfc3954.txt.

Телков Александр Юрьевич – канд. физ.-мат. наук, доцент кафедры электроники, физический факультет, Воронежский государственный университет.
E-mail: telkov@dpo-it.ru

Крохин Евгений Александрович – студент второго курса магистратуры кафедры электроники, физический факультет, Воронежский государственный университет.
E-mail: e.krohin@gmail.com

Telkov Alexandr Yurievich – Master of Science, the faculty of physics, the Department of electronics, Voronezh state University.
E-mail: telkov@dpo-it.ru

Krokhin Evgeny Alexandrovich – a second year student of the magistracy, the faculty of physics, the Department of electronics, Voronezh state University.
E-mail: e.krohin@gmail.com