УДК 004.832.3

Ю.П. НИКОЛАЕВА

Севастопольский национальный технический университет, Украина

ИССЛЕДОВАНИЕ СТАТИСТИЧЕСКИХ ХАРАКТЕРИСТИК ПЕРЕДАЧИ ШИФРОВАННЫХ ИНФОРМАЦИОННЫХ ПОТОКОВ ЧЕРЕЗ ТРАНЗИТНЫЙ УЗЕЛ СЕТИ

Рассматривается задача исследования взаимосвязи между статистическими характеристиками исходного IPv6 потока и потока, зашифрованного при помощи технологии IPSec. По условиям эксперимента зашифрованный трафик собирается на транзитном узле информационной сети. Генерируется трафик, параметры которого имеют законы распределения специального вида и исследуются эмпирические характеристики трансформированного потока. Появляется дополнительная возможность накопления информации об исходном потоке на основании статистического исследования значений поля «Payload Length» заголовка зашифрованного IPv6 пакета для дальнейшего использования в систему поддержки принятия решений (СППР) для классификации и фильтрации потоков данных.

Ключевые слова: *IPSec, IPv6, ESP, ESP Payload, закон распределения.*

Введение

В настоящее время усиливается тенденция к максимизации объемов трафика, передаваемого по информационным сетям в зашифрованном виде. Помимо положительного эффекта – повышения безопасности коммуникаций – наблюдается и отрицательный, заключающийся в потере контроля администратора сети над информационными потоками. Это приводит, с одной стороны, к усложнению контроля над сетевыми вторжениями и вредоносным программным обеспечением. С другой стороны, усложняется управление нежелательными действиями легитимных пользователей: распространением файлов, передачи голосовых и мгновенных текстовых сообщений.

Для идентификации и последующей классификации информационных потоков широко используются средства машинного обучения и искусственного интеллекта [1 – 3], эвристические подходы [4, 5], а также средства статистического анализа параметров информационных потоков [6, 7].

В работах [8] и [9] используется визуальный анализ шаблонов трафика. Ни один из данных подходов не обеспечивает полного решения проблемы классификации информационных потоков. Наиболее перспективным направлением является построение информационной технологии, включающей в себя систему поддержки принятия решений (СППР), основанную на результатах многоверсионного анализа информационных потоков.

В данной работе ставится цель исследовать статистические параметры IPv6 потока, зашифрованного при помощи технологии IPSec. Предлагается рассмотреть законы распределения параметров заголовков зашифрованного IPv6 пакета и сравнить с законами распределения длины поля данных исходного, нетрансформированного, потока.

Результаты анализа трансформированного информационного потока являются входными данными проектируемой СППР по классификации трафика.

1. Постановка задачи

Объектом исследования является информационная система специального вида, осуществляющая передачу и обработку информационных потоков. Информационная система включает в себя передающие и принимающие узлы, а так же узел наблюдения, для которого принципиально ограничены возможности сбора информации о некоторых параметрах информационных потоков.

Цель исследования – при помощи методов статистического анализа выявить зависимость между параметрами передаваемых информационных потоков и параметрами потоков, доступными для измерения в узле наблюдения.

Формальная постановка задачи статистического анализа следующая:

Пусть информационная сеть, состоящая из n+1 узла, задается графовой структурой специального вида со следующей матрицей смежности:

$$G_{ij} = \begin{cases} 1, & i = 0, j = 1, n; \\ 1, & j = 0, i = 1, n; \\ 0, & \text{иначе.} \end{cases}$$

Соответствующий граф представлен на рис. 1.

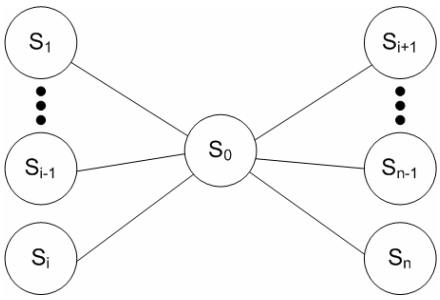

Рис. 1. Граф-схема информационной сети с узлом наблюдения S_0

Узел S0 является узлом наблюдения, узлы S1 – Sn являются логическими источниками и приемниками информационного потока.

Информационным потоком между i-м и j-м узлами графа G за интервал времени ΔT будем называть конечное множество специальных элементов ν - «информационных сообщений»:

$$I_{\Delta T}^{i,j} = \left\{ \nu \middle| V_{\Delta T}(\nu) \right\}.$$

Характеристическое свойство множества может быть сформулировано следующим образом:

$$V_{\Delta T}(\nu) = V_S^i(\nu) \vee V_D^i(\nu) \vee V_T^{i,j}(\nu, \Delta T),$$

где $V_S^i(\nu)$ и $V_D^i(\nu)$ - функции, характеризующие источник и приемник информационного сообщения ν, $V_T^{i,j}(\nu, \Delta T)$ - характеристика временного интервала.

$V_S^i(\nu)$ = {1, если i является узлом источником для ν; 0 иначе}. $V_D^i(\nu)$ = {1, если j является узлом приемником для ν, 0 иначе}. $V_T^{i,j}(\nu, \Delta T)$ = {1, если ν зафиксирован в узле S0 за интервал ΔT; 0 иначе}.

Шифрование T является функцией на множестве I информационных сообщений, трансформирующей поток $I_{\Delta T}^{i,j}$ в $I_{\Delta T}^{*i,j}$:

$$T : I_{\Delta T}^{i,j} \to I_{\Delta T}^{*i,j}. \qquad (1)$$

В узле S_0 доступны для измерения только параметры трансформированного, зашифрованного потока $I_{\Delta T}^{*i,j}$.

Требуется определить условия, при которых по статистическим характеристикам потока $I_{\Delta T}^{*i,j}$ возможны статистически значимые выводы об исходном потоке $I_{\Delta T}^{i,j}$.

2. Методика исследования

Для исследования зависимости между исходным и зашифрованным информационными потоками исследовалась локальная сеть, использующая в качестве протокола сетевого уровня IPv6. Между отправителем и получателям трафика настроен виртуальный канал по технологии IPSec в режиме «tunnel» [9]. Протокол аутентификации не использовался.

Схема трансформации IP-пакетов, использующаяся в режиме «tunnel», приведена на рис. 2.

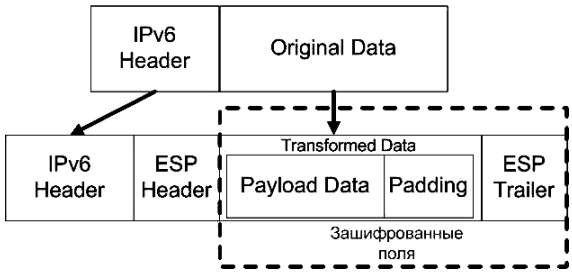

Рис. 2. Трансформация исходного IPv6 пакета при помощи IPSec в режиме «tunnel»

Использованы следующие протоколы шифрования, объединенные в три группы по результатам трансформации размера исходного пакета (табл. 1).

Таблица 1

Используемые протоколы

Группа	Протоколы
I	1) 3DES-CBC (192 bit) [RFC2451] 2) DES-CBC (64 bit) [RFC2405] 3) BLOWFISH-CBC (128 bit) [RFC2451]
II	1) AES-CTR (224 bit)[RFC3686]
III	1) TWOFISH-CBC (256 bit) 2) CAMELLIA-CBC (256 bit) [RFC3713] 3) AES-CBC(256 bit) [RFC3602]

В квадратных скобках для каждого протокола шифрования приведен соответствующий стандарт IETF [10].

Для генерации потоков данных использовались протоколы ICMP (скрипты на основе утилиты «ping6») и UDP (набор инструментов D-ITG [11]).

Формат поля «Original data» (рис. 2) при использовании этих инструментов принимает следующий вид (рис. 3):

Для исследования статистических параметров трансформированного потока использовалось поле «Payload Length» заголовка IPv6 пакета. Поле содержит длину содержимого пакета, следующего за заголовком IPv6. В условиях проводимого эксперимента значение поля «Payload Length» эквивалентно длине поля «Encapsulation Security Payload» («ESP payload»).

В табл. 2 показано соответствие между размером поля данных «Data» исходного пакета и длиной поля «ESP payload» зашифрованного пакета для каждой из групп протоколов шифрования.

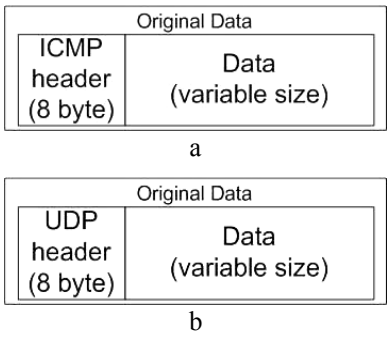

Рис. 3. Содержимое поля «Original data» исходного пакета при использовании протоколов ICMP(a) и UDP (b)

Таблица 2

Соответствие значений полей данных при трансформации информационных потоков

Размер поля «Data» исходного пакета, байт	Размер поля «ESP payload» трансформированного пакета, байт		
	Группа протоколов шифрования		
	I	II	III
1	32	28	40
…	…	…	…
998	1024	1024	1032
999	1032	1028	1048
1000	1032	1028	1048
1001	1032	1028	1048
1002	1032	1028	1048
1003	1032	1032	1048
1004	1032	1032	1048
1005	1032	1032	1048
1006	1032	1032	1048
1007	1040	1036	1048
1008	1040	1036	1048

Анализ табл. 1 показывает, что преобразование T (1) отображает исходное множество размеров поля данных на множество с меньшей мощностью, т.е. преобразование происходит с потерей информации об исходных параметрах потока.

Потеря информации вызвана тем, что алгоритмы шифрования оперируют блоками, кратными некоторому предопределенному числу, зависящему от алгоритма. Поле данных дополняется полем Padding, содержащим незначащие символы, и затем шифруется [12].

Размер поля данных «Data» варьировалось в диапазоне [1 … 1414], что позволяет исследовать статистические параметры трансформированного потока без учета особенностей, накладываемых фрагментацией IPv6 пакета [13].

Для контроля и наблюдения за параметрами зашифрованного фрагмента сообщения применялся Wireshark с установленным модулем GCript [14].

3. Анализ характеристик трансформированных информационных потоков

Рассматривались случаи, когда распределение размера поля данных («Data») исходного, нешифрованного потока, подчиняется нормальному, гамма и равномерному законам. Выбор количества интервалов осуществлялся в соответствии с рекомендациями [15]. Объем выборки во всех экспериментах принимался равным 10000 пакетов.

Для проверки гипотезы о законе распределения во всех экспериментах использовался критерий согласия Пирсона [16].

Гипотеза о виде распределения переменной не противоречит статистическими данным, если выполняется соотношение

$$P(\chi^2_{\alpha,r} > \chi^2_0) = p > \alpha = 0,05,$$

где $\chi^2_{\alpha,r}$ - табличное значение критерия для уровня значимости α и r степеней свободы, χ^2_0 - расчетное значение критерия, p - расчетный уровень значимости. Уровень значимости α во всех экспериментах принимается равным 0.05.

3.1. **Нормальный закон распределения** длины поля данных «Data» исходного потока.

Параметры исходного генерируемого потока приняты следующими: $\sigma = 200$, $\mu = 800$. Теоретическая функция плотности вероятности:

$$f(x,\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\left(\frac{(x-\mu)^2}{2\sigma^2}\right)}.$$

На рис. 4 представлены графики эмпирической плотности распределения вероятности для параметров исходного пакета («Payload») и пакетов, трансформированных при помощи I-й, II-й и III-й групп протоколов шифрования, соответственно. Визуальный анализ позволяет сделать предположение о нормальном распределении параметров трансформированного потока.

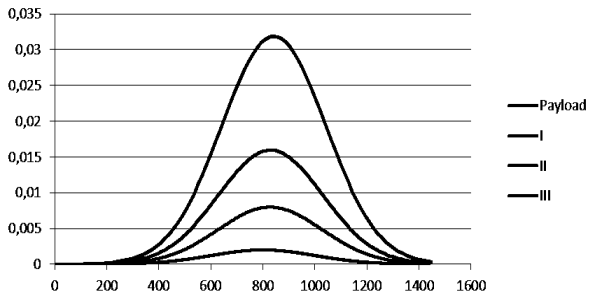

Рис. 4. Эмпирическая плотность распределения параметров исходного и трансформированного потоков. Исходный закон распределения – нормальный

В табл. 3 представлены уровни значимости p для критерия согласия Пирсона. Полученные значения для всех групп протоколов шифрования превышают уровень значимости $\alpha = 0,05$, что позволяет принять гипотезу о нормальном распределении длин поля «ESP Payload» зашифрованного IPv6 потока.

Таблица 3

Уровень значимости, математическое ожидание и дисперсия для трансформированных потоков. Исходный закон распределения – нормальный

	I	II	III
p	0,91	0,77	0,97
M[X]	829	827	840
$\sqrt{D[X]}$	198	198	198

3.1. **Гамма-распределение** длины поля данных «Data» исходного потока.

Параметры потока $\alpha = 2$, $\beta = 150$. (Вообще говоря, в данном случае гамма-распределение является распределением Эрланга 2 порядка).

Теоретическая функция плотности вероятности:

$$f(x,\alpha,\beta) = \frac{1}{\beta^{\alpha}\Gamma(\alpha)} x^{\alpha-1} e^{-\frac{x}{\beta}}.$$

На рис. 5 приведены графики плотности распределения для исходного потока и для значений длины поля «ESP Payload» трансформированных потоков. Визуальный анализ показывает, что графики для зашифрованных потоков имеют особенность в начале оси абсцисс. Это связано с тем, что ни для каких параметров исходного потока результирующие параметры не принимает значения, меньшие 32, 28 или 40 для I, II и III групп протоколов шифрования соответственно (табл. 1). Поэтому далее рассматриваются смещенные эмпирические функции распределения, полученные следующим образом:

$$F^{*}[X] = F[X - \min\{x_i\} + 1]. \qquad (2)$$

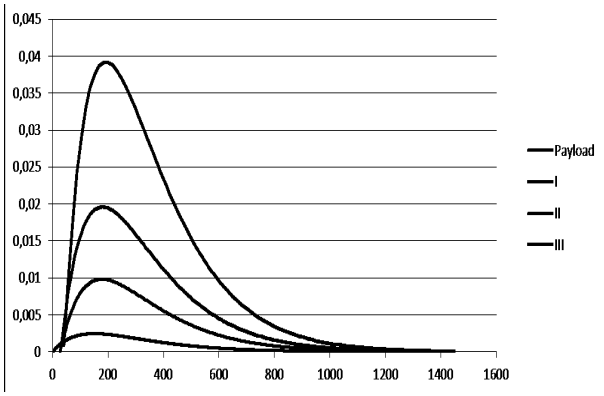

Рис. 5. Эмпирическая плотность распределения параметров потоков. Исходный закон распределения – гамма

Для трансформированного потока гипотеза о распределении по гамма- закону не подтверждается. Однако для смещенного потока, преобразованного по правилу (2), статистические данные не противоречат гипотезе о гамма-распределении генеральной совокупности.

В табл. 4 приведены значения уровней значимости p, превышающие заданный уровень значимости α для всех групп протоколов шифрования.

Таблица 4

Уровень значимости, математическое ожидание и дисперсия для трансформированных потоков. Исходный закон распределения – гамма

	I	II	III
p	0,97	0,32	0,44
M[X]	298	300	302
$\sqrt{D[X]}$	209	207	208

3.1. **Равномерный закон распределения** длины поля данных «Data» исходного потока.

Параметры исходного потока

$$a = 1,\ b = 1414.$$

Теоретическое распределение:

$$f(x,a,b) = \begin{cases} \frac{1}{b-a}, x \in [a,b]; \\ 0, x \notin [a,b]. \end{cases}$$

Анализ эмпирических функций плотности распределения (рис. 6) для параметров исходного и трансформированных потоков позволяет сделать предположение о равномерном законе распределения длин поля «ESP payload» зашифрованного IPv6 потока.

Действительно, анализ значений расчетных уровней значимости p (табл. 5) позволяет сделать вывод о том, что статистические данные не противоречат гипотезе о равномерном распределении.

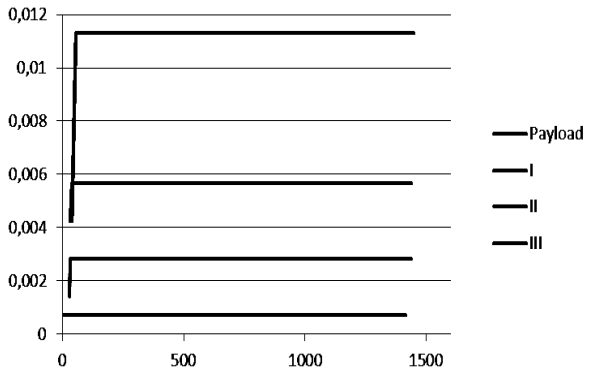

Рис. 6. Эмпирическая плотность распределения исходного и модифицированного потоков

Таблица 5

Уровень значимости, математическое ожидание и дисперсия для трансформированных потоков. Исходный закон распределения – равномерный

	I	II	III
p	0,47	0,68	0,34
M[X]	737	735	749
$\sqrt{D[X]}$	408	408	408

Выводы

В ходе проведенных экспериментальных исследований было показано, что для нормального и равномерного распределений длины поля данных исходного IPv6 пакета, распределение значения длины поля «ESP Payload» при заданных условиях подчиняется тому же закону, что и для исходного потока. Для исходного потока, параметры которого распределены согласно гамма-закону, статистические данные по смещенному при помощи преобразования (2) результирующему потоку не противоречат гипотезе о гамма-распределении.

Таким образом, появляется дополнительная возможность накопления информации об исходном потоке на основании статистического исследования значений поля «Payload Length» заголовка зашифрованного IPv6 пакета для дальнейшего использования в СППР для классификации и фильтрации потоков данных.

Планируется продолжение исследования зависимости между параметрами заголовка исходного IPv6 пакета и параметрами зашифрованного потока для потоков, не подчиняющихся стандартным законам распределения (медиаданные, передача голосовых сообщений, P2P-транзакции).

Литература

1. Bar-Yanai, Roni. Realtime Classification for Encrypted Traffic [Text] / Roni Bar-Yanai, Michael Langberg, David Peleg, Liam Roditty // Experimental Algorithms: 9th International Symposium, SEA. – 2010. – P. 373 –385.

2. Alshammari, R. Machine Learning Based Encrypted Traffic Classification: Identifying SSH and Skype [Text] / R. Alshammari, A.N. Zincir-Heywood // Computational Intelligence for Security and Defense Applications, 2009. CISDA 2009. – 8-10 July 2009. – P. 1 – 9.

3. Identifying the traffic of SSH-encrypted applications [Electronic resource] / M. Dusi, A. Este, F. Gringoli, L. Salgarelli. – Url: http://www.gtti.it/GTTI09/files/papers/Reti/Reti_12.10_ Dusi.pdf. – 23.02.2012.

4. Dewes, C. Ananalysis of Internet chat systems [Текст] / C. Dewes, A. Wichmann, and A. Feldmann // Proceedings of the 3rd ACM SIGCOMM conference on Internet measurement. Miami Beach, FL, USA, October 2003. – P. 51 – 64.

5. Gu, Ch., Encrypted Internet Traffic Classification Method based on Host Behavior [Text] / Chengjie Gu, Shunyi Zhang, Xiaozhen Xue // JDCTA: International Journal of Digital Content Technology and its Applications, Vol. 5, No. 3, 2011. – C. 167– 174.

6. Sun, G. Novel Hybrid Method for Effectively Classifying Encrypted Traffic [Text] / Guang-Lu Sun, Yibo Xue, Yingfei Dong, Dongsheng Wang and Chenglong LiA // Global Telecommunications Conference (GLOBECOM 2010), 2010. – 6–10 Dec. 2010. – P. 1–5.

7. Wright, Ch. Using Visual Motifs to Classify Encrypted Traffic [Text] / Charles V. Wright, Fabian Monrose, Gerald M. Masson. // VizSEC '06 Proceedings of the 3rd international workshop on Visualization for computer security. 2006. – P. 34 – 37.

8. Lian, W. Traffic Classification Using Visual Motifs: An Empirical Evaluation [Text] / Wilson Lian Fabian Monrose, John McHugh // VizSec'10. Proceedings of the Seventh International Symposium on Visualization for Cyber Security. 2010. – P. 45 – 52.

9. Bollapragada, V. IPSec VPN Design [Text] / Vijay Bollapragada, Mohamed Khalid, Scott Wainner. – Cisco Press. 2005. – 384 p.

10. Request for Comments (RFC) [Electronic resource]. – Url: http://www.ietf.org/rfc.html. – 23.02.2012.

11. D-ITG, Distributed Internet Traffic Generator [Electronic resource]. – Url: http://www.grid.unina.it /software/ITG. – 23.02.2012.

12. Rhee, M. Internet Security. Cryptographic Principles, Algorithms and Protocols [Text] / Man Young Rhee. John Wiley & Sons Ltd, 2003. – 405 p.

13. Lewis, M. Comparing, Designing, and Deploying VPNs / M. Lewis. – Cisco Press. 2006. – 1080 c.

14. ESP Payload Decryption / ESP Authentication Checking [Electronic resource]. – Url: http://wiki. wireshark.org/ESP_Preferences. – 23.02.2012.

15. Прикладная статистика. Правила проверки согласия опытного распределения с теоретическим [Электронный ресурс]. – Режим доступа: http://www.complexdoc.ru/ntdtext/541018/1. – 23.02.2012.

16. Куприенко, Н.В. Статистика. Методы анализа распределений. Выборочное наблюдение. [Текст] / Н.В. Куприенко, О.А. Пономарева, Д.В. Тихонов. – СПб.: Изд-во Политехн. ун-та, 2009. – 138 с.

Поступила в редакцию 5.03.2012

Рецензент: д-р техн. наук, проф. И.А. Жуков, Национальный авиационный университет, Киев, Украина.

ДОСЛІДЖЕННЯ СТАТИСТИЧНИХ ХАРАКТЕРИСТИК ПЕРЕДАЧІ ШИФРОВАНИХ ІНФОРМАЦІЙНИХ ПОТОКІВ ЧЕРЕЗ ТРАНЗИТНИЙ ВУЗОЛ МЕРЕЖІ

Ю.П. Ніколаєва

Розглядається завдання дослідження взаємозв'язку між статистичними характеристиками вихідного потоку і потоку, зашифрованого за допомогою технології IPSec, IPv6. За умовами експерименту зашифрований трафік збирається на транзитному вузлі інформаційної мережі. Генерується трафік, параметри якого мають закони розподілу спеціального вигляду і досліджуються емпіричні характеристики трансформованого потоку. З'являється додаткова можливість накопичення інформації про вихідний потоці на підставі статистичного дослідження значень поля «Payload Length» заголовка зашифрованого IPv6 пакета для подальшого використання в СППР для класифікації та фільтрації потоків даних.

Ключові слова: IPSec, IPv6, ESP, ESP Payload, закон розподілу

STATISTICAL ANALYSIS OF ENCRYPTED TRAFFIC TRANSMISSION THROUGH TRANSIT NETWORK NODE

Y.P. Nikolaieva

Connection between IPv6 payload and ESP payload of encrypted traffic is investigated. Statistic is collected on a transit network node. Standard-based probability distributions are used to generate experimental traffic. An additional opportunity to generate information about the original thread on the basis of a statistical study of field values «Payload Length» IPv6 packet header encrypted for use in decision support system for the classification and filtering of data streams.

Key words: IPSec, IPv6, ESP, ESP Payload, probability distribution

Николаева Юлия Петровна – ассистент кафедры кибернетики и вычислительной техники Севастопольского национального технического университета, Севастополь, Украина.