ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ И СИСТЕМЫ, ВЫЧИСЛИТЕЛЬНАЯ ТЕХНИКА

УДК 004.056.55
DOI: 10.17586/0021-3454-2020-63-9-777-785

ИССЛЕДОВАНИЕ ВЕРОЯТНОСТИ ВОЗНИКНОВЕНИЯ ОДНОГО ТИПА ОШИБОК В СИСТЕМЕ ОПРЕДЕЛЕНИЯ ИСТОЧНИКА ИНФОРМАЦИОННЫХ ПАКЕТОВ

М. О. Таныгин

Юго-Западный государственный университет, 305040, Курск, Россия,
E-mail: tanygin@yandex.ru

Представлен метод определения источника информационных пакетов в системах, использующих протоколы связи, не позволяющие формировать сообщения длиной более 10—15 байтов. Метод основан на определении принадлежности конкретного полученного приемником пакета структурированному множеству пакетов фиксированного размера, ассоциированному с источником. Для определения принадлежности каждый информационный пакет содержит поле хеш-последовательности, сформированной на основе данных других пакетов источника и уникального числового идентификатора источника. Приемник на основе анализа этой последовательности и последовательностей других информационных пакетов формирует множество пакетов, образующих единое сообщение источника и определяет место каждого пакета в сообщении. На основе теории вероятностей и теории случайных процессов получены численные характеристики вероятности возникновения ошибки определения источника, сформировавшего пакет. Получены зависимости вероятности такой ошибки от параметров алгоритма формирования информационного пакета: длины поля хеша и мощности множества информационных пакетов, для которого определяется принадлежность источнику. Показана применимость метода для разделения приемником сообщений от множества источников, сформулированы зависимости между длиной поля хеша и количеством источников сообщений.

Ключевые слова: *определение источника сообщений, теория вероятностей, хеш, ошибка передачи, параметры алгоритма формирования сообщений*

В настоящее время существует множество подходов к решению задачи определения аутентичности источника сообщений, передаваемых по открытым каналам связи. Обычно в их основе лежит решение различных криптографических задач источником и приемником сообщений [1—3]. Для целого класса программно-аппаратных систем, в которых взаимодействие программных и аппаратных компонентов происходит через подсистему ввода-вывода операционной системы, характерной особенностью канала связи является небольшой размер пакета данных, передаваемых между источником и приемником [4—6]. Такой размер не позволяет реализовывать криптографические преобразования над ним, обеспечивающие требуемый уровень достоверности аутентификации и целостности [7—10]. В то же время нарушение корректности потока данных между компонентами таких систем может оказать существенное влияние на их работоспособность в целом [7, 8].

Для изоляции подобных каналов связи и повышения вероятности правильного опознавания источника сообщений целесообразно использовать способы и алгоритмы, основанные на формировании из отдельных информационных пакетов структур пакетов, которые ассоциированы с конкретным отправителем [11, 12]. Целостность и аутентичность отдельного пакета определяется его принадлежностью к такой структуре, этот подход использован в известной технологии блокчейн [13, 14]. Наиболее вероятной ошибкой определения принадлежности информационного пакета такой структуре является ошибочное включение в ее состав пакета, сформированного и переданного в приемник не целевым отправителем, а посторонним источником. Природа этого отправителя в данном случае неважна: он может быть как злоумышленником, осуществляющим деструктивные воздействия на систему, так и просто абонентом, передающим собственные пакеты в приемник. В результате цепочка информационных пакетов от легального источника не будет обработана [15].

В настоящей работе рассмотрим использование методов определения источника информационных пакетов на основе связанности пакетов в единое структурированное множество, которое будем называть цепочкой. Определены следующие принципы работы системы передачи информационных пакетов (ИП):

— на источнике генерируются информационные пакеты и передаются по каналу связи в приемник;

— в приемник поступают не только ИП от легального источника, но и посторонние ИП,

— посторонние ИП по формату неотличимы от легальных,

— целостность пакетов сообщений в процессе передачи не нарушается,

— все передаваемые ИП закодированы с использованием уникальной (известной только отправителю и приемнику) числовой последовательности, в приемнике происходит их декодирование, что позволяет считать посторонние ИП сформированными случайным образом,

— последовательность ИП восстанавливается в приемнике с известной вероятностью на основе алгоритмов обработки входящих сообщений.

Согласно сказанному, на приемнике получается последовательность, длина которой равна сумме отправленных и посторонних информационных блоков. Источник формирует множество ИП в виде $S^{\text{source}} = \left\{S_1^s, S_2^s, \ldots, S_{N^p}^s\right\}$, где N^p — число пакетов в передаваемой цепочке. На вход приемника поступает множество ИП $S^{\text{rec}} = \left\{S_1^r, S_2^r, \ldots, S_{N^p+N^f}^r\right\}$, где N^f — число посторонних пакетов, выданных злоумышленником (рис. 1; пакет, сформированный посторонним источником, выделен штриховкой).

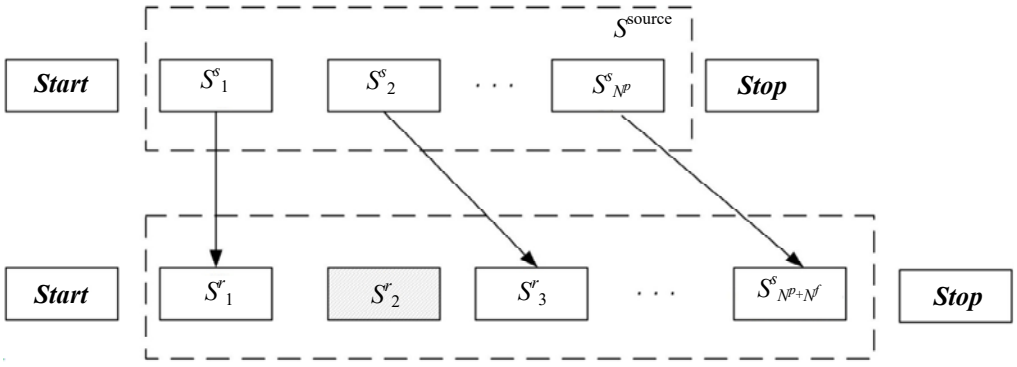

Рис. 1

По условию, заданному моделью, все пакеты формируются источником и поступают на вход приемника между нулевым пакетом, или старт-словом, и (N^f+1)-м пакетом, или стоп-словом. Данные пакеты не входят в состав цепочки, а служат лишь граничными пакетами,

семантически связанными с содержимым ИП цепочки. Из-за возможности задания специального формата для них считаем, что формирование случайным образом старт-слова или стоп-слова — событие с пренебрежимо малой вероятностью наступления [15].

Задача приемника заключается в установлении взаимно-однозначного соответствия между элементами множества S^{source} и множества S^{rec} (рис. 1) за счет анализа хеш-последовательностей, содержащихся в каждом ИП. Цепочкой S^{chain} длиной l будем называть подмножество множества S^{rec} мощностью l, элементы которого удовлетворяют условию:

$$S_i^{\text{hash}} = F_{\text{hash}}\left(S_{i-1}^{\text{inf}}\right), \quad i = \overline{1, \ldots, l}, \tag{1}$$

где $S_i = \left\{S_i^{\text{inf}} \mid S_i^{\text{hash}}\right\}$ — пакет из множества S^{rec}, записываемый в цепочку на позицию i, состоящий из информационной части S_i^{inf} и хеша S_i^{hash}, F_{hash} — хеш-функция, S_{i-1}^{inf} — информационная часть пакета, находящегося на позиции i–1 цепочки, S_0 — старт-слово.

Критерием правильности построения цепочки и корректности определения источника является возможность построения цепочки длиной N^p из пакетов множества S^{rec} и ее единственность. Две цепочки считаем различными, если они различаются хотя бы одним элементом.

Пусть последний элемент цепочки длиной l есть элемент j множества S^{rec}:

$$S_l \in S^{\text{chain}} = S_j^r \in S^r. \tag{2}$$

Процесс добавления ИП в цепочку заключается в выделении в множестве S^{rec} подмножества s_j подряд идущих после S_j^r пакета ИП:

$$s_j \in S^{\text{chain}} = \left\{S_{j+1}^r, S_{j+2}^r, \ldots, S_{j+m}^r\right\}. \tag{3}$$

Мощность этого подмножества определяется из выражения:

$$m = \min\left(N^p + N^f - j, N^f\right). \tag{4}$$

Элемент S подмножества s_j добавляется в цепочку S^{chain}, если для него выполняется равенство (1). Если среди элементов подмножества s_j имеется несколько ИП S, S', S'', S''', для которых выполняется данное равенство, то вместо одной цепочки формируется несколько, каждая из которых состоит из исходной цепочки S^{chain} с добавленными к ней соответственно ИП S, S', S'', S'''... В дальнейшем добавление ИП к этим вновь образовавшимся цепочкам происходит независимо от других.

Если отсутствует хотя бы один элемент подмножества s_j, удовлетворяющий условию (1), и выполняется неравенство $l < N^p$, цепочка считается сформированной частично или полностью из посторонних ИП и ликвидируется. Цепочка длиной N^p входит в множество цепочек как легальная, т.е. составленная исключительно из слов множества S^{source}. В случае правильного завершения алгоритма должна остаться одна такая цепочка. Входящие в нее информационные пакеты считаются выданными легальным источником.

Ошибочное определение источника сообщения, при котором анализ всех ИП покажет, что сформированных цепочек длиной N^p более одной, назовем, по аналогии с [15], коллизией. Наиболее простой случай такой ошибки, когда две цепочки различаются только одним i-м ИП, представлен на рис. 2. Условие возникновения подобной коллизии запишется в виде:

$$S1_i^{\text{hash}} = S1_i^{\text{hash}} = F_{\text{hash}}\left(S_{i-1}^{\text{inf}}\right),$$
$$F_{\text{hash}}\left(S1_i^{\text{inf}}\right) = F_{\text{hash}}\left(S2_i^{\text{inf}}\right) = S_{i+1}^{\text{hash}}. \quad (5)$$

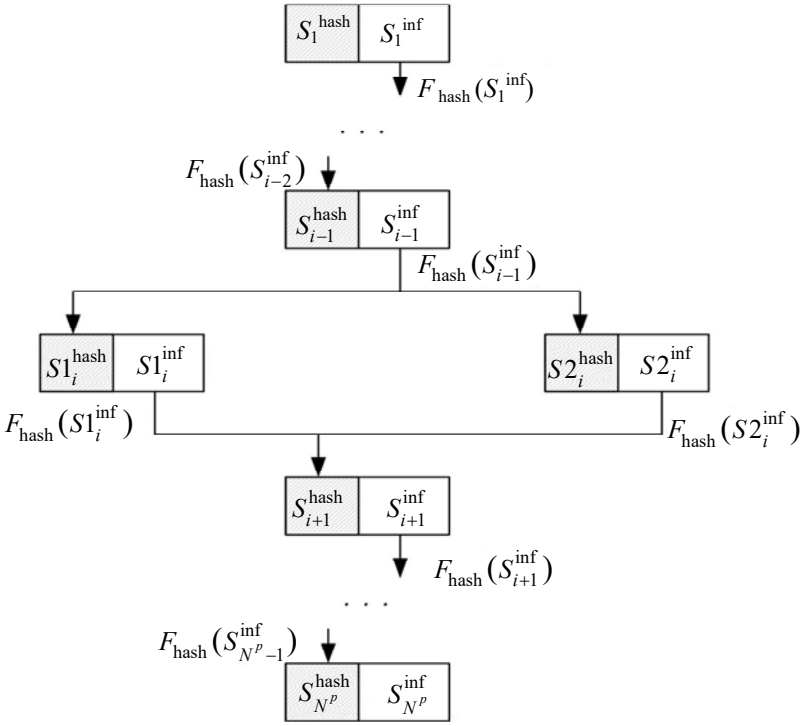

Рис. 2

В случае возникновения коллизии задача выделения из множества цепочек той, которая соответствует множеству S^{source}, не может быть решена исходя из оговоренных условий передачи и используемых алгоритмов. В результате потребуется повторная передача всех ИП цепочки.

Для определения вероятности возникновения коллизий рассмотрим момент, когда в результате анализа поступивших в приемник ИП сформирована цепочка ИП из множества S^{source} длиной l. Приемник анализирует сформированное по описанным правилам подмножество s_j путем проверки хешей его ИП и определения того, который будет добавлен в цепочку (l+1)-м элементом (рис. 3; $S0$ — последний добавленный пакет в цепочку, выданный источником ИП, $S1$ и $S2$ — следующие за ним легальные ИП, Sf_1 – Sf_r — ИП посторонних источников, полученные в приемнике в промежутке между пакетами $S0$ и $S2$. Первые m из них принадлежат множеству s_j, анализируются приемником, могут быть добавлены в цепочку после пакета $S0$ и могут вызвать простейшую коллизию).

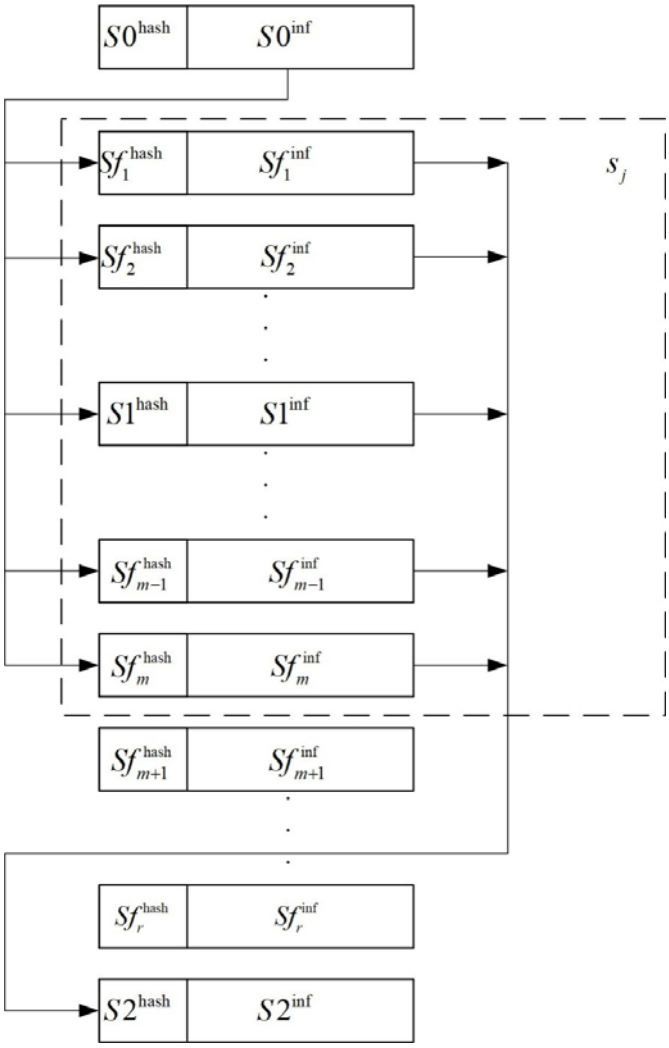

Рис. 3

Представим процесс формирования простой коллизии в виде цепи Маркова с дискретным временем (рис. 4; S_0 — начальное состояние; $S_{1,i}, i=\overline{1,\ldots,m}$ — состояния, соответствующие получению приемником между пакетами $S0$ и $S2$ ровно i ИП посторонних источников; $S_{2,i}, i=\overline{1,\ldots,m}$ — состояния, соответствующие тому, что хеш ровно из i посторонних ИП подмножества s_j удовлетворяет первой части условия (5) и может быть включен в цепочку после пакета $S0$; S_f —состояние, соответствующее тому, что хотя бы один из посторонних ИП подмножества s_j, хеш которого совпал с $F_{\text{hash}}\left(S0^{\text{inf}}\right)$, удовлетворил второй части условия (5) и может быть включен в цепочку между $S0$ и $S2$, формируя цепочку, отличающуюся от легальной на один ИП; S_t — состояние, соответствующее безошибочному добавлению к легальной цепочке пакета $S1$ и отсутствию коллизии на данном этапе). Данное представление адекватно, так как совпадение хешей и информационных частей каждого из посторонних ИП $Sf_1 - Sf_r$ с требуемым для включения его в цепочку между пакетами $S0$ и $S2$ есть случайное событие, не зависящее от того, совпали или нет хеши и информационные части остальных посторонних ИП подмножества s_j [16].

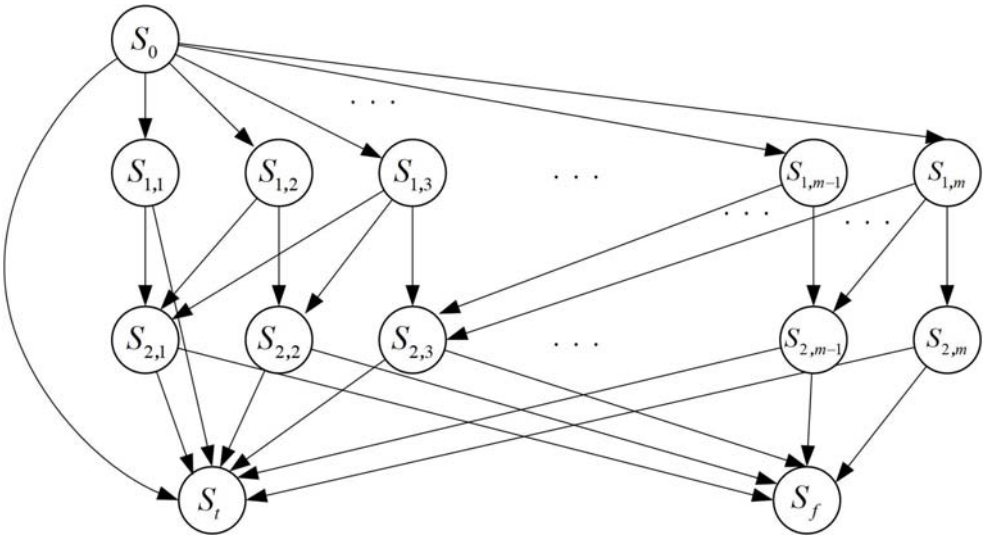

Рис. 4

Так как число посторонних ИП, полученных между несколькими легальными, распределено по закону Пуассона, то вероятность перехода между состояниями S_0 и $S_{1,i}, i=\overline{1,\ldots,m}$, запишется в виде:

$$p\left(S_0, S_{1,i}\right)=\begin{cases}\dfrac{\left(2\times N^f / N^p\right)^i}{i!}e^{-\frac{2N^f}{N^p}}, & i=\overline{1,\ldots,m-1},\\ \displaystyle\sum_{j=i}^{\infty}\frac{\left(2\times N^f / N^p\right)^i}{i!}e^{-\frac{2N^f}{N^p}}, & i=m,\end{cases} \tag{6}$$

где $2\times N^f / N^p$ — среднее число посторонних ИП, полученных в промежутке между пакетами *S*0 и *S*2. Соответственно вероятность перехода между состояниями S_0 и S_t равна $e^{-2\times N^f/N^p}$. Вероятность того, что из i посторонних ИП ровно у j хеш совпадет с хешем, сформированным из информационной части ИП *S*0, имеет биномиальное распределение. Тогда вероятность перехода между состояниями $S_{1,i}, i=\overline{1,\ldots,m}$, и $S_{2,j}, j=\overline{1,\ldots,i}$:

$$p\left(S_{1,i}, S_{2,j}\right)=C_i^j\left(2^{-H}\right)^j\left(1-2^{-H}\right)^{i-j}, \tag{7}$$

где H — длина поля хеша информационного пакета.

Вероятность перехода $p\left(S_{2,i}, S_t\right)$ между состояниями $S_{2,j}, j=\overline{1,\ldots,m}$, и S_t равна вероятности того, что из i посторонних ИП ни у одного хеш информационной части не совпадет с хешем *S*2: $\left(1-2^{-H}\right)^i$, а вероятность $p\left(S_{2,i}, S_f\right)$ между $S_{2,j}, j=\overline{1,\ldots,m}$, и S_f — $1-\left(1-2^{-H}\right)^i$.

Вероятность попадания системы в состояние S_f по истечении достаточного времени P_f равна вероятности простой коллизии для одного из N^p ИП множества. Общая вероятность P_{sc} возникновения подобной ошибки при передаче всей цепочки определится выражением

$$P_{sc}=1-\left(1-P_f\right)^{N^p}. \tag{8}$$

На рис. 5 приведены зависимости вероятности возникновения ошибки при передаче полной цепочки из-за простой коллизии от длины поля хеша и отношения числа посторонних ИП к числу легальных (*1* — *H*=6, *2* — 7). Видно, что с ростом как длины цепочки N^p, так и активности источников посторонних ИП, выражаемой соотношением N^f / N^p, вероятность ошибки асимптотически приближается к единице. Но в определенных диапазонах она может быть аппроксимирована прямой, т.е. для практического использования в реальных системах можно предположить, что вероятность ошибки опознавания источника ИП прямо пропорциональна длине цепочки и активности источников посторонних ИП или числу посторонних ИП, получаемых за время сеанса передачи цепочки ИП. Если построить данный график, зафиксировав параметры N^p и N^f / N^p, изменяя лишь длину поля хеша *H* (рис. 6; *1* — $N^f / N^p = 3$; *2* — 6; *3* — 9, *4* — 16), взяв ось ординат в логарифмическом масштабе, видно, что вероятность возникновения ошибки пропорциональна значению показательной функции от длины поля хеша. В результате имеем следующую пропорцию для вероятности возникновения ошибки:

$$P_{sc} \sim N^p \times e^{-H} \times N^f / N^p = e^{-H} \times N^f . \qquad (9)$$

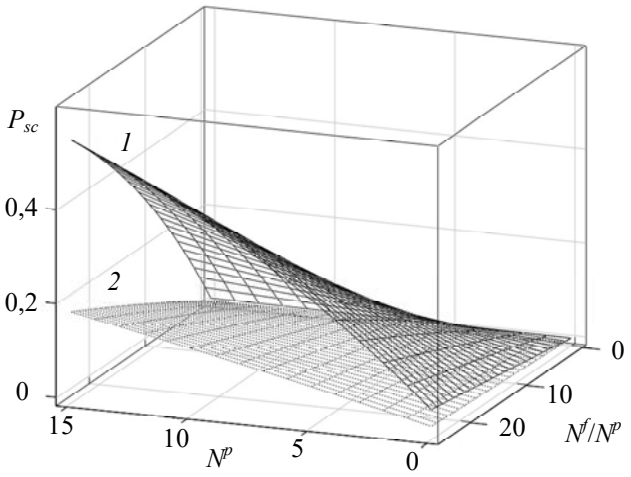

Рис. 5

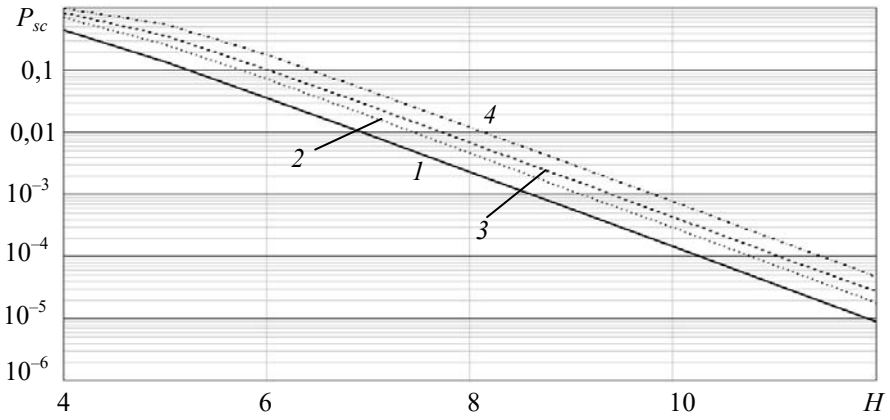

Рис. 6

Иными словами, вероятность возникновения ошибки типа „простая коллизия“ в системе можно изменять, варьируя только длину поля хеша, а не длину цепочки. Полученное соотношение позволяет подбирать значение поля хеша исходя из условий передачи. Если перед системой стоит задача разделения ИП, поступающих от нескольких источников, то по формуле $N^f / N^p + 1$ можно определить их число. Тогда для сохранения требуемого уровня достоверности определения источника каждого ИП и отнесения его к соответствующему

источнику длину поля хеша сообщения необходимо увеличивать пропорционально натуральному логарифму от числа источников информации.

СПИСОК ЛИТЕРАТУРЫ

1. *Bellare M., Canetti R., Krawczyk H.* Keying hash functions for message authentication // Advances in Cryptology. Lecture Notes in Computer Science. 1996. Vol. 1109. P. 1—15.
2. *Black J.*, *Rogaway P*. CBC MACs for arbitrary-length messages: The three-key constructions // J. Cryptol. 2015. Vol. 18, N 2. P. 111—131.
3. *Stallings W.* NIST Block Cipher Modes of Operation for Confidentiality // Cryptologia. 2010. N 34(2). P. 163—175.
4. PCI Special Interest Group. PCI Express® Base Specification Revision 3.0 [Электронный ресурс]: <http://citeseerx.ist.psu.edu/viewdoc/download? doi=10.1.1.694.7081&rep=rep1&type=pdf>. (дата обращения 15.10.2019)
5. *Слободин Р. С., Добрица В. П., Таныгин М. О., Талдыкин Е. В., Непочатых Е. В.* Способ обмена данными между контроллерами защиты информации по протоколу PCI–Express // Телекоммуникации. 2019. № 8. С. 21—26.
6. *Tanygin M. O., Tipikin A. P.* Methods of authentication of information protection systems and controlling software // Telecommunications and Radio Engineering. 2007.Vol. 66, N 5. P. 453—463.
7. *Типикин А. П., Глазков А. С.* Метод и функциональная организация контроля обращений и закрытия доступа к секторам файлов при хищении накопителя информации // Информационные технологии. 2010. № 5. С. 25—30.
8. *Типикин А. П., Глазков А. С., Муратов С. А.* Организация пользовательской системы защиты информации, хранящейся на жестком магнитном диске // Телекоммуникации. 2009. №10. С.33—37.
9. *Лоднева О. Н., Ромасевич Е. П.* Анализ трафика устройств интернета вещей // Современные информационные технологии и ИТ-образование. 2018. Т. 14, № 1. С. 149—169.
10. *Муравьева-Витковская Л. А.* Оценка структурных параметров маршрутизатора при приоритетном управлении неоднородным трафиком с произвольным распределением длин пакетов // Изв. вузов. Приборостроение. 2017. Т. 60, № 10. С. 951—956.
11. *Karri R.*, *Rajendran J., Rosenfeld K., Tehranipoor M.* Trustworthy hardware: Identifying and classifying hardware Trojans // Computer. 2010. Vol. 43, is. 10. P. 39—46. DOI: 10.1109/MC.2010.299.
12. *Ganesh Kumar, Sriman B., Murugan A., Muruganantham B.* IoT – smart contracts in data trusted exchange supplied chain based on block chain // Intern. J. of Electrical and Computer Engineering. 2020. Vol. 10, is.1. P. 438—446. DOI: 10.11591/ijece.v10i1.pp438-446.
13. *Shangping W., Dongyi L., Yaling Z., Juanjuan C.* Smart Contract-Based Product Traceability System in the Supply Chain Scenario // IEEE Access. 2019. Vol. 7. DOI:10.1109/ACCESS.2019.2935873.
14. *Таныгин М. О., Алшаиа Х. Я., Алтухова В. А.* Об одном методе контроля целостности передаваемой поблоково информации // Телекоммуникации. 2019. № 3. С. 12—21.
15. *Вентцель Е. С., Овчаров Л. А.* Теория случайных процессов и ее инженерные приложения. М.: Наука, 1991. 384 с.

Сведения об авторе

Максим Олегович Таныгин — канд. техн. наук, доцент; Юго-Западный государственный университет, кафедра информационной безопасности; заведующий кафедрой; E-mail: tanygin@yandex.ru

Поступила в редакцию
28.02.2020 г.

Ссылка для цитирования: *Таныгин М. О.* Исследование вероятности возникновения одного типа ошибок в системе определения источника информационных пакетов // Изв. вузов. Приборостроение. 2020. Т. 63, № 9. С. 777—785.

INVESTIGATION OF ONE TYPE OF ERRORS PROBABILITY IN SYSTEM DETERMINING INFORMATION PACKETS SOURCE

M. O. Tanygin

South-West State University, 305040, Kursk, Russia,
E-mail: tanygin@yandex.ru

A method for determining information packets source in systems using communication protocols that do not allow generating messages longer than 10—15 bytes, is described. The method is based on determining whether a particular received packet belongs to a structured set of fixed size packets associated with the source. To identify the ownership, each information packet contains a hash sequence field formed on data from other packets of the source and the source unique numeric identifier. As a result of analysis of this sequence and sequences of other information packets, the receiver generates a set of packets that form a single source message and determines location of each packet in the message. Using the probability theory and the theory of random processes, numerical characteristics of the probability of occurrence of error in determining the packet generated source are derived. The error probability dependences on the information packet generating algorithm parameters — the hash field length and the power of the analyzed set of information packets — are obtained. Applicability of the method for separating messages from multiple sources is demonstrated, and the dependencies between the hash field length and the number of message sources are formulated.

Keywords: messages source determining, probability theory, hash, transmission error, the message generation algorithm parameters

REFERENCES

1. Bellare M., Canetti R., Krawczyk H. *Advances in Cryptology. Lecture Notes in Computer Science*, 1996, vol. 1109, pp. 1–15.
2. Black J., Rogaway P. *J. Cryptol*, 2015, no. 2(18), pp. 111–131.
3. Stallings W. *Cryptologia*, 2010, no. 34(2), pp. 163–175.
4. *PCI Special Interest Group. PCI Express® Base Specification Revision 3.0*,http://citeseerx.ist.psu.edu/viewdoc/download? doi=10.1.1.694.7081&rep=rep1&type=pdf.
5. Slobodin R.S., Dobritsa V.P., Tanygin M.O., Taldykin E.V., Nepochatykh E.V. *Telekommunikatsii* (Telecommunications), 2019, no. 8, pp. 21–26. (in Russ.)
6. Tanygin M.O., Tipikin A.P. *Telecommunications and Radio Engineering*, 2007, no. 5(66), pp. 453–463.
7. Tipikin A.P., Glazkov A.S. *Information Technologies* (Informacionnye Tehnologii), 2010, no. 5, pp. 25–30. (in Russ.)
8. Tipikin A.P., Glazkov A.S., Muratov S.A. *Telekommunikatsii* (Telecommunications), 2009, no. 10, pp. 33–37.
9. Lodneva O.N., Romasevich E.P. *Modern Information Technologies and IT-education*, 2018, no. 1(14), pp. 149–169. (in Russ.)
10. Murav'yeva-Vitkovskaya L.A. *Journal of Instrument Engineering*, 2017, no. 10(60), pp. 951–956. (in Russ.)
11. Karri R., Rajendran J., Rosenfeld K., Tehranipoor M. *Computer,* 2010, no. 10(43), pp. 39–46, DOI: 10.1109/MC.2010.299.
12. Ganesh Kumar, Sriman B., Murugan A., Muruganantham B. *Journal of Electrical and Computer Engineering*, 2020, no. 1(10), pp. 438–446,DOI: 10.11591/ijece.v10i1.pp438-446.
13. Shangping W., Dongyi L., Yaling Z., Juanjuan C. *IEEE Access*, 2019, vol. 7, DOI:10.1109/ACCESS.2019.2935873.
14. Tanygin M.O., Alshaia Kh.Ya., Altukhova V.A. *Telekommunikatsii* (Telecommunications), 2019, no. 3, pp. 12–21.
15. Venttsel' E.S., Ovcharov L.A. *Teoriya sluchaynykh protsessov i yeye inzhenernyye prilozheniya* (Theory of Stochastic Processes and Its Engineering Applications), Moscow, 1991, 384 p. (in Russ.)

Data on author

Maxim O. Tanygin — PhD, Associate Professor; South-West State University, Department of Information Security; Head of the Department; E-mail: tanygin@yandex.ru

For citation: Tanygin M. O. Investigation of one type of errors probability in system determining information packets source. *Journal of Instrument Engineering*. 2020. Vol. 63, N 9. P. 777—785 (in Russian).

DOI: 10.17586/0021-3454-2020-63-9-777-785