УДК: 81.93.29

РАСПОЗНАВАНИЕ ТИПА ПАКЕТОВ, ПЕРЕДАВАЕМЫХ В ТРАНСПОРТНОМ ПОТОКЕ ПИРИНГОВОЙ МУЛЬТИСЕРВИСНОЙ СЕТИ «SKYPE»

RECOGNITION OF THE TYPE OF PACKETS TRANSMITTED IN «SKYPE» PEER-TO-PEER MULTISERVICE NETWORK TRAFFIC

Канд. техн. наук Д.А. Клецков, В.В. Кузьмин, А.С. Егоров

PhD D.A. Kletskov, V.V. Kuzmin, A.S. Egorov

Военный университет радиоэлектроники

В сложной геополитической обстановке, в том числе из-за активизации различных террористических организаций, деятельности иностранных спецслужб, а также вмешательства государств во внутренние дела других стран по средствам использования сети общего пользования Интернет актуальной задачей является мониторинг сетевого трафика, передаваемого в данных сетях. Широкое распространение получили сервисы, предоставляющие возможность абонентам передавать в сети мультимедийную информацию, такие как пиринговые сети «Skype», «Telegram», «Viber», «WhatsApp». Статья посвящена вопросам применения статистических методов обработки результатов наблюдения для распознавания типа пакетов, передаваемых в транспортном потоке пиринговой мультисервисной сети «Skype». Приводится обоснование порога принятия решения на основе критерия минимума вероятности суммы ошибок первого и второго рода.
Ключевые слова: пиринговая сеть «Skype», коэффициенты асимметрии и эксцесса, распознавание, трафик, критерий.

In a complex geopolitical situation, including due to the activation of various terrorist organizations, the activities of foreign intelligence services, as well as state interference in the internal Affairs of other countries through the use of the public Internet, monitoring of network traffic transmitted in these networks is an urgent task. Services that allow subscribers to transmit multimedia information to the network, such as the peer-to-peer networks Skype, Telegram, Viber, and WhatsApp, have become widespread. The article is devoted to the application of statistical methods for processing observation results for recognizing the type of packets transmitted in the transport stream of the Skype peer-to-peer multiservice network. The decision threshold is justified based on the minimum probability criterion for the sum of errors of the first and second types.
Keywords: peer-to-peer network «Skype», skewness and kurtosis coefficients, recognizing, traffic, criterion.

В статье рассмотрен вопрос распознавания речевых пакетов и пакетов без информационной нагрузки в транспортном потоке пиринговой сети «Skype» с целью повышения оперативности обработки речевых сообщений. Особенностью пиринговой сети является то, что каждый абонент (узел) сети может выполнять функции выделенного сервера. Данная организационная структура сети работоспособна при любом количестве и сочетании доступных узлов и эффективнее классических архитектур клиент/сервер. Следует отметить, что стремительное развитие

IP-телефонии началось с момента появления данной сети. На алгоритмах и интерфейсах, схожих с «Skype», функционирует множество других сетей подобного типа. В настоящее время сеть «Skype» остается одной из самых популярных систем IP-телефонии. Данный сервис был представлен в 2003 году и к сентябрю 2011 года насчитывал около 663 миллионов пользователей. К 2014 году доля трафика пиринговой сети «Skype» составляла 40 % от международного телефонного рынка. Динамика роста числа пользователей данной сети представлена на рис. 1 [3, 8].

Функционирование сети «Skype» основано на совокупности методов аутентификации пользователей, специально созданных протоколах криптографического закрытия информации с использованием технологии динамической смены ключей. Вариант структуры пиринговой сети «Skype» представлен на рис. 2.

Пиринговая сеть «Skype» представляет собой иерархическую систему. В состав сети входят следующие элементы: сервер аутентификации (СА); суперузлы (СУ); абоненты сети (узлы) [6, 10]. Центральным элементом сети является

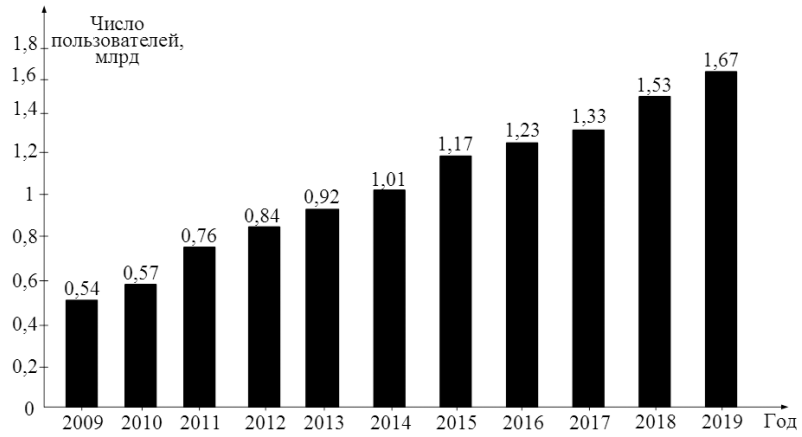


Рис. 1. Динамика роста числа пользователей пиринговой сети «Skype»

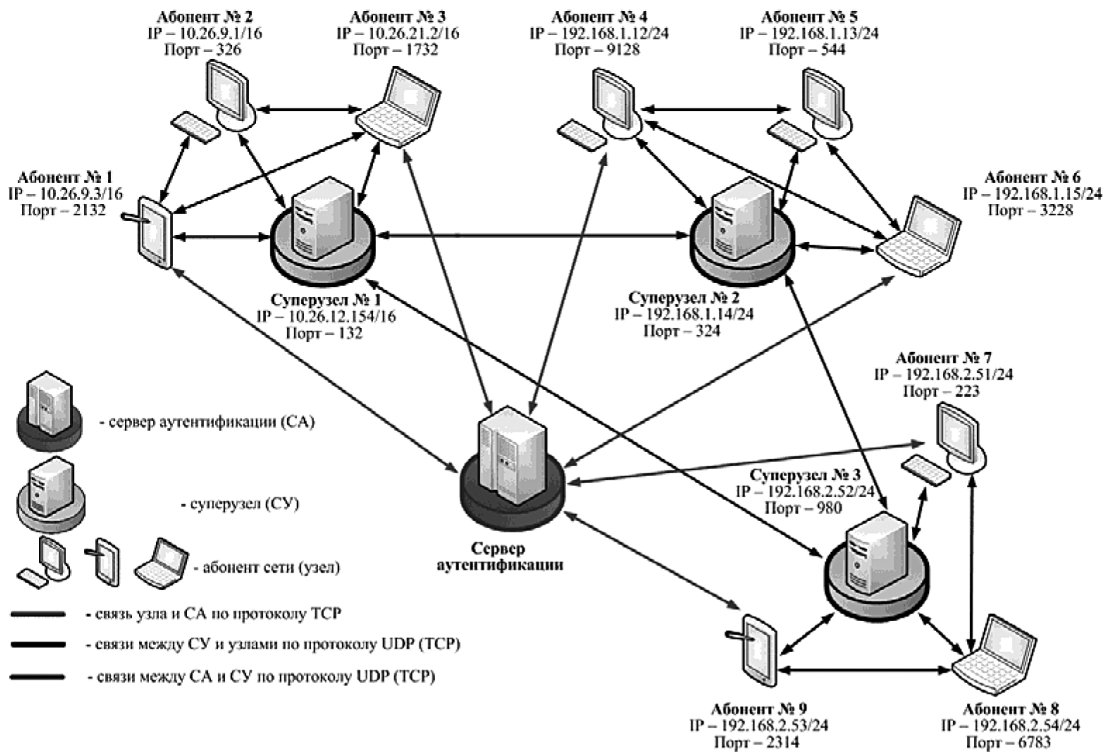


Рис. 2. Вариант взаимодействия структурных элементов сети «Skype»

СА, в котором содержится база данных с учетными записями пользователей сети. При прохождении процедуры аутентификации, пользователь отправляет СА запрос в виде служебного сообщения, в котором указываются уникальные логин и пароль, подтверждающие его право на подключение к сети. Алгоритм установления соединения между СА и абонентами при прохождении процедуры аутентификации основан на использовании протокола транспортного уровня Transmission Control Protocol (TCP) и включает в себя три этапа: запрос абонента на аутентификацию с СА, ответ СА на запрос абонента, подтверждение успешного прохождения процедуры аутентификации и подключение абонента к сети «Skype». Необходимо отметить, что среднее время установления соединения составляет 0,05 секунды. Последовательность этапов при прохождении процедуры аутентификации проходит в несколько этапов. На первом этапе абонент, аутентифицирующийся в сети «Skype», отправляет СА пакет с запросом на установление соединения, обозначая начало установления соединения. При этом происходит назначение параметров сеанса связи, в виде формирования счетчика полезной нагрузки абонента. На втором этапе СА формирует ответ на запрос абонента с счетчиком полезной нагрузки сервера, увеличивая значение счетчика полезной нагрузки абонента на единицу. На третьем этапе абонент отправляет СА служебный пакет с подтверждением заявки на установление соединения, инкрементируя значение счетчика полезной нагрузки СА на единицу. После выполнения этапов аутентификации абонента в сети «Skype» соединение считается установленным. После успешного прохождения процедуры аутентификации СА по запросу абонентов сети устанавливает соединение между ними. Если установление соединения не представляется возможным, используется СУ, являющийся посредником между пользователями сети «Skype». Ранее роль СУ в сети мог выполнять любой пользователь с внешним IP-адресом и открытым TCP-портом. С 2012 года, после приобретения сети «Skype» компанией Microsoft, роль СУ выполняют только серверы этой компании. Это объясняется тем, что при предоставлении пользователю прав СУ требуется значительное количество трафика. В этом режиме пользователь сети пропускал через себя «чужой» трафик по принципу работы пиринговой сети [1, 5, 9].

В «Skype» не используется какого-либо метода подавления молчания, то есть, если речь во время разговора отсутствует, информационные пакеты продолжают формироваться и передаваться. Отсутствие поддержки подавления молчания введено специально для достижения лучшего качества голоса и поддержания связи с помощью протоколов пользовательских датаграмм (UDP) и NAT (Network address translation — преобразование сетевых адресов).

При выявлении информативного признака с целью распознавания типа информационных пакетов (речь или речевая пауза), передаваемых в сети «Skype», проведён статистический анализ сетевого трафика при осуществлении аудиозвонка «Skype». Для перехвата сетевого трафика сети «Skype» использовалось специальное программное обеспечение «WireShark 3.2.3». По результатам анализа сетевого трафика сети «Skype» выдвинута гипотеза о том, что существует взаимосвязь между длинной передаваемого пакета и типом нагрузки в данном пакете (речь или речевая пауза) сети «Skype». Длина речевого пакета, передаваемого по протоколу UDP, в зависимости от того, говорит ли пользователь в данный момент или нет, имеет различное значение. Обязательным условием для определения типа передаваемого пакета является использование информативного признака, основанного на нахождении порогового значения длины передаваемого пакета, при которой будет установлена его принадлежность к тому или иному типу передаваемых данных — пакетам с речевыми сообщениями (ПРС) или пакетам с речевыми паузами (ПРП). С целью определения порогового значения длины пакета, обеспечивающей переход в разряд пакетов с речевыми сообщениями, проведем анализ законов распределения длин рассматриваемых ПРС и ПРП, а также найдем основные параметры распределений. Для оценивания закона распределения случайной величины $\hat{L}$ (безразмерная величина) и основных числовых характеристик распределения использованы методы проверки гипотез о законе распределения [2, 6], в основе которых лежит исследование функции распределения. Качество оценивания функции распределения требует задания соответствующей точности (доверитель-

ный интервал ε) и надежности (доверительная вероятность β). Пусть доверительный интервал равен 0,01, а доверительная вероятность составляет 0,95. Требуемое число испытаний $N_{\text{тр}}$, обеспечивающее заданное качество оценивания функции распределения предполагаемого закона распределения случайной величины $\hat{L}$, определяется неравенством вида

$$N_{\text{тр}} \geq \frac{t_\beta^2}{4\varepsilon^2},$$

где t_β — табличная функция.

Тогда требуемое число испытаний составляет $N_{\text{тр}} = 9600$ для ПРС и ПРП. Для проведения анализа осуществлена выборка $N = 10625$ ПРС и $N = 10125$ ПРП, где N — реальное число испытаний. Гистограммы распределения случайной величины $\hat{L}$ для случая передачи пакетов обоих типов представлены на рис. 3.

С целью обоснования гипотезы о законе распределения, которому подчинены результаты экспериментов, в соответствии с методикой [2, 3, 7] по полученным экспериментальным данным необходимо рассчитать основные числовые характеристики распределения, такие как оценка математического ожидания — МО ($\tilde{m}_{\hat{L}}$), оценка дисперсии ($\tilde{D}_{\hat{L}}$) и оценка среднего квадратического отклонения — СКО ($\tilde{\sigma}_{\hat{L}}$), а также оценки коэффициента асимметрии ($\tilde{e}_{\hat{L}}$) и коэффициента эксцесса ($\tilde{e}_{\hat{L}}$) по следующим выражениям:

$$\tilde{m}_{\hat{L}} = \sum_{i=1}^{\phi} \overline{L}_i \cdot p_i^*, \; p_i^* = \frac{z_i}{N_{\text{тр}}};$$

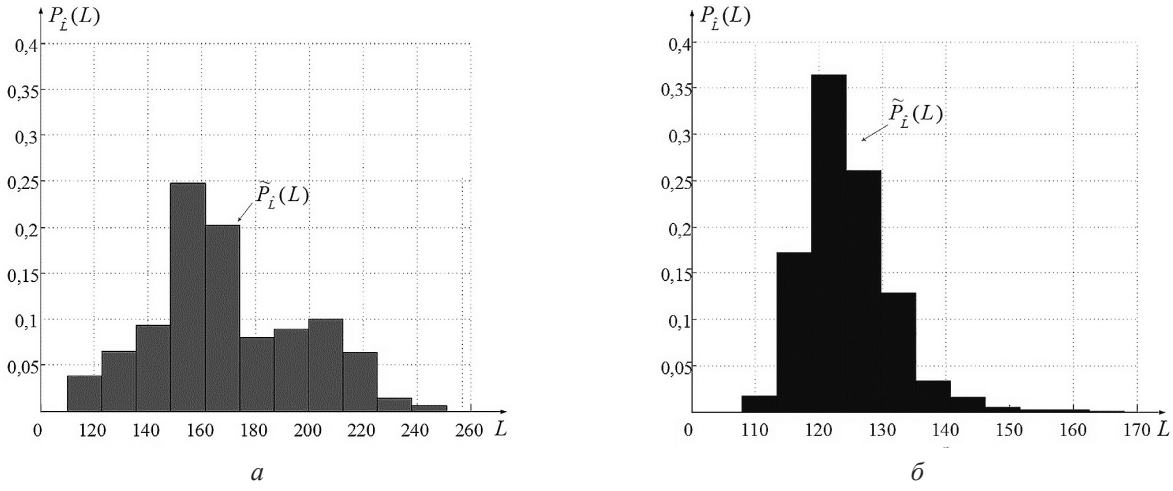
$$\tilde{D}_{\hat{L}} = \sum_{i=1}^{\phi} \left(\overline{L}_i - \tilde{m}_{\hat{L}}\right) \cdot p_i^*, \; \tilde{\sigma}_{\hat{L}} = \sqrt{\tilde{D}_{\hat{L}}};$$

$$\tilde{a}_{\hat{L}} = \frac{\tilde{\mu}_3\left[\hat{L}\right]}{\left(\tilde{\sigma}_{\hat{L}}\right)^3}, \; \tilde{e}_{\hat{L}} = \frac{\tilde{\mu}_4\left[\hat{L}\right]}{\left(\tilde{\sigma}_{\hat{L}}\right)^4} - 3,$$

где $\tilde{\mu}_3\left[\hat{L}\right]$ — оценка центрального момента третьего порядка; $\tilde{\mu}_4\left[\hat{L}\right]$ — оценка центрального момента четвертого порядка; p_i^* — частота попадания вариантов случайной величины $\hat{L}$ в i-й разряд гистограммы; z_i — число значений (вариантов) случайной величины $\hat{L}$, попавших в i-й разряд гистограммы; ϕ — число разрядов, на которые делится интервал наблюдения данной случайной величины; $\overline{L}_i$ — середина i-го разряда гистограммы. Наличие оценок коэффициентов асимметрии и эксцесса позволяет выдвинуть гипотезу о гипотетическом законе распределения. Основные числовые характеристики распределения и результаты расчета оценок коэффициентов асимметрии и эксцесса по статистическим данным представлены в табл. 1.

В работе [2] показано, что каждому закону свойственно определенное соотношение между коэффициентами асимметрии и эксцесса (рис. 4).

Полученные значения оценок коэффициентов асимметрии и эксцесса (табл. 1) находятся вблизи от прямой 3 и точки 4, что позволяет выдвинуть начальные гипотезы о нормальном и логарифмически нормальном законах распределения случайной величины $\hat{L}$ при передаче информационных пакетов обоих типов. Для



а *б*

Рис. 3. Гистограммы распределения случайной величины $\hat{L}$: а — ПРС; б — ПРП

Таблица 1

Сводная таблица числовых характеристик распределения

Тип пакетов	Объем выборки ($N_{\text{тр}}$)	МО ($\tilde{m}_{\hat{L}}$)	СКО ($\tilde{\sigma}_{\hat{L}}$)	Коэффициент асимметрии ($\tilde{a}_{\hat{L}}$)	Коэффициент эксцесса ($\tilde{e}_{\hat{L}}$)
Пакеты с речевыми сообщениями (ПРС)	10625	169,58	27,76	0,34	– 0,46
Пакеты с речевыми паузами (ПРП)	10125	124,45	6,85	1,22	3,33

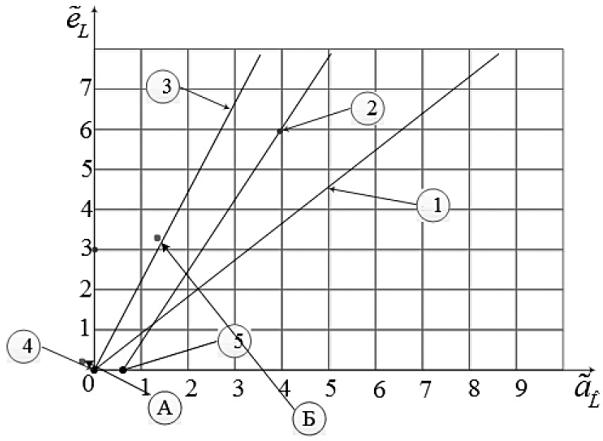


Рис. 4. Взаимосвязь коэффициентов асимметрии и эксцесса для разных законов распределения:
1 — для закона Пуассона;
2 — для показательного закона;
3 — для логарифмически нормального распределения;
4 — для нормального закона;
5 — для равномерного закона;
А — зависимость оценок коэффициентов асимметрии и эксцесса полученные для ПРП;
Б — зависимость оценок коэффициентов асимметрии и эксцесса полученные для ПРС

проверки гипотез о законах распределения воспользуемся методом А. Колмогорова. В этом случае, показатели согласованности распределения опытных данных по нормальному ($\hat{u}_{\text{норм}}$) и логарифмически-нормальному ($\hat{u}_{\text{логнорм}}$) законам будут описываться следующими выражениями:

$$\hat{u}^{(\text{Колм})}_{\text{норм}} = \sqrt{N_{\text{тр}}}\max_{L}\left|\hat{F}^{*(\text{норм})}_{\hat{L}}(L) - F^{(\text{норм})}_{\hat{L}^{\Gamma}}(L)\right|,$$

$$\hat{u}^{(\text{Колм})}_{\text{логнорм}} = \sqrt{N_{\text{тр}}}\max_{L}\left|\hat{F}^{*(\text{логнорм})}_{\hat{L}}(L) - F^{(\text{логнорм})}_{\hat{L}^{\Gamma}}(L)\right|,$$

где $\hat{F}^{*(\text{норм})}_{\hat{L}}(L)$, $F^{(\text{норм})}_{\hat{L}^{\Gamma}}(L)$, $\hat{F}^{*(\text{логнорм})}_{\hat{L}}(L)$, $F^{(\text{логнорм})}_{\hat{L}^{\Gamma}}(L)$ — статистическая и гипотетическая (теоретическая) функции распределения наблюдаемой случайной величины $\hat{L}$, вычисленные по логарифмически нормальному и нормальному законам распределения. Значение уровня значимости ζ для метода Колмогорова примем равным 0,01, а соответствующую ему критическую границу $u^{(\text{Колм})}_{\zeta} = 1{,}627$. Полученные показатели согласованности для опытных данных сведены в табл. 2.

Полученные значения показателей согласованности при предположении о нормальном и логарифмически нормальном законах распределения длин пакетов для пакетов с речью и паузами не превышают критическую границу при проверке соответствующим методом, т.е. $\tilde{u}^{(\text{Колм})}_{\text{норм}} < u^{(\text{Колм})}_{\zeta}$ и $\tilde{u}^{(\text{Колм})}_{\text{логнорм}} < u^{(\text{Колм})}_{\zeta}$. Тем не менее, исследуемые функции имеют значительные коэффициенты асимметрии, что не позволяет отнести их к нормальному закону распределения. Таким образом, принимается гипотеза о логарифмически нормальном законе распределения. Статистические и теоретические функции распределения представлены на рис. 5.

Таблица 2

Значения показателей согласованности для опытных данных

Тип пакетов	Показатель согласованности $\hat{u}^{(\text{Колм})}_{\text{норм}}$	Показатель согласованности $\hat{u}^{(\text{Колм})}_{\text{логнорм}}$
ПРС	1,19	1,243
ПРП	1,496	1,598

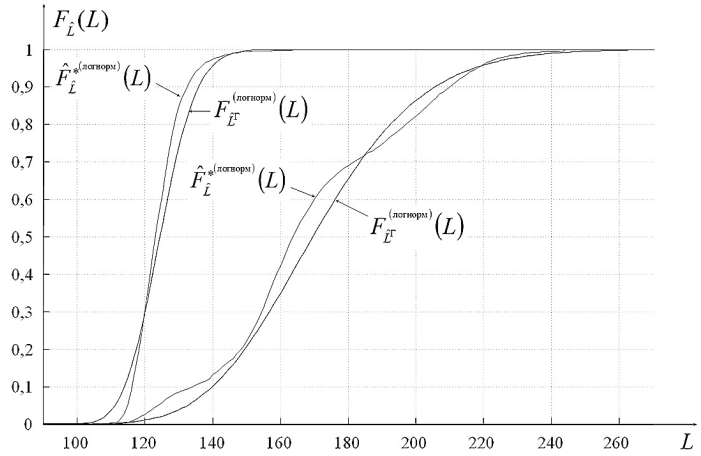
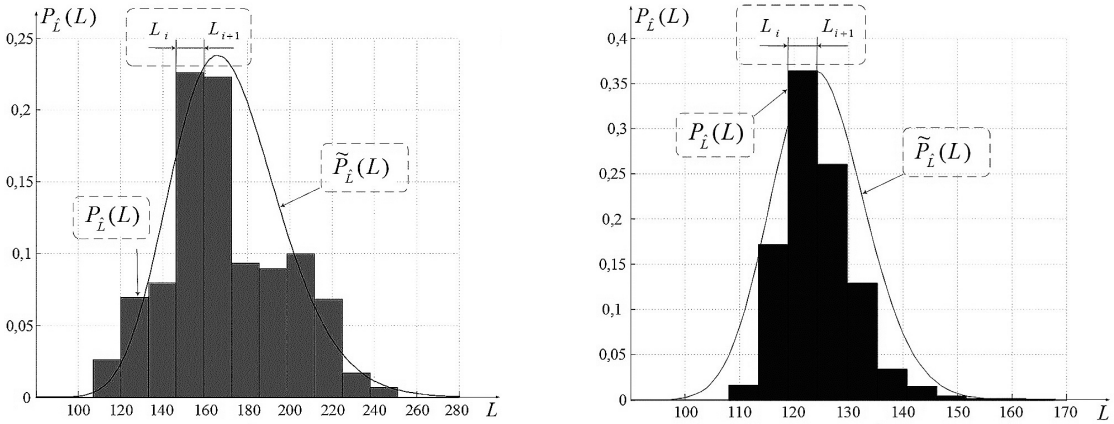


Рис. 5. Статистическая и теоретическая функции распределения длин пакетов опытных данных для логарифмически нормального закона распределения

В силу этого, аппроксимацию полученных статистических распределений допустимо осуществлять с помощью логарифмически нормального закона распределения. Гистограммы и аппроксимирующие кривые распределения для случая передачи ПРС и ПРП представлены на рис. 6).

В силу того, что поведение информативного признака $\hat{L}$ носит случайный характер, задача определения типа передаваемых данных сводится к задаче статистической проверки альтернативных гипотез: Γ_0 — временной ряд соответствует передаче ПРП; Γ_1 — временной ряд соответствует передаче ПРС, рис. 7. Рассмотренные события являются случайными, им могут быть поставлены в соответствие вероятности наступления описанных событий. Их вероятности обозначим следующим образом: p_{11} — вероятность правильного решения о факте передачи ПРС; p_{12} — вероятность ложного решения о факте передачи ПРС; p_{21} — вероятность ложного решения о факте передачи ПРП; p_{22} — вероятность правильного решения о факте передачи ПРП. Величины вероятностей $p_{11}, p_{12}, p_{21}, p_{22}$ зависят от размеров и расположения области D_0 допустимых значений и критической области D_1, рис. 7. Поэтому, предъявляя соответствующие требования к вероятностям $p_{11}, p_{12}, p_{21}, p_{22}$, необходимо определить расположение и размеры данных областей, т.е. критическую границу. В рассматриваемом случае, когда из априорной информации известны только условные законы распределения результатов наблюдений случайной величины $\hat{L}$,



а *б*

Рис. 6. Гистограммы и аппроксимирующие кривые распределения случайной величины L: а — ПРС; б — ПРП

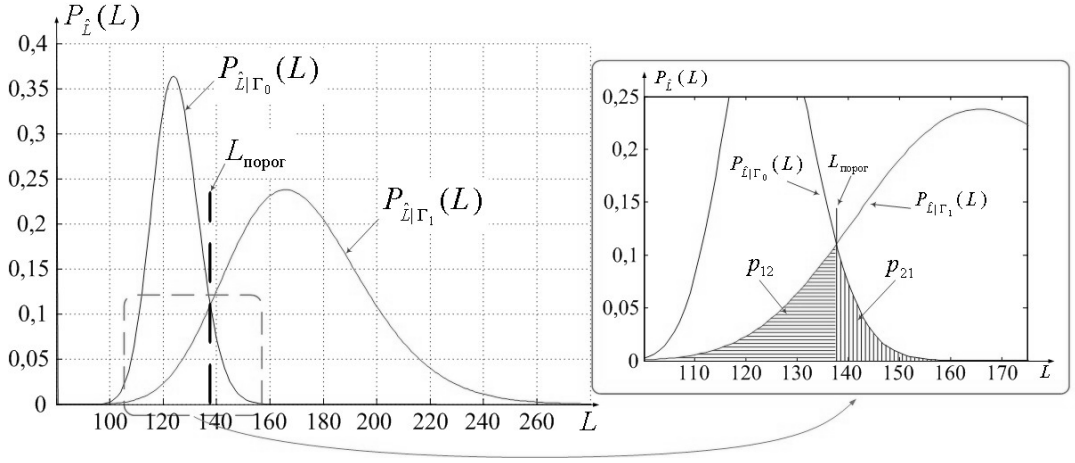


Рис. 7. Пороговое значение длин пакетов

удобно использовать критерий минимальной вероятности ошибки. Его сущность состоит в том, что минимизируется вероятность принятия неправильного решения, т.е. сумма вероятностей ошибок первого и второго рода должна быть минимальной [2, 4]. Согласно критерию минимума вероятности ошибки критическая точка $L_{\text{порог}}$, рис. 7, выбирается таким образом, чтобы сумма вероятностей ошибок первого и второго рода была минимальной, т.е. выполнялось условие

$$L_{\text{порог}} = \arg\min_{\{L_{\text{порог}}\}} \left(p_{12} + p_{21}\right).$$

В результате подстановки числовых значений, проведенных преобразований и решения квадратного уравнения относительно искомого параметра значение порога составляет $L_{\text{порог}} = 138,27$. Значения вероятности ложного решения о факте передачи пакетов с речью $p_{12} = 8,68 \cdot 10^{-2}$, а вероятности ложного решения о факте передачи пакетов с паузами $p_{21} = 6,26 \cdot 10^{-2}$. Соответственно вероятность правильного решения о факте передачи речевых пакетов $p_{11} = 9,132 \cdot 10^{-1}$, а вероятность правильного решения о факте передачи пакетов с паузами $p_{22} = 9,374 \cdot 10^{-1}$. Полученные результаты пороговых значений длин пакетов представлены на рис. 7. Учитывая малую величину суммы вероятностей ошибки, можно сделать вывод о надежности выявленного, основанного на нахождении порогового значения длины передаваемого пакета информативного признака, который предназначен для определения типа передаваемых данных.

Таким образом, можно сделать вывод о том, что выдвинутая гипотеза о различных значениях длины пакетов с речевой нагрузкой и паузами в трафике пиринговой сети «Skype» подтверждена практически. На основе этого возможно выполнить распознавание информационных пакетов в сетевом трафике пиринговой сети «Skype», целью которого будет разделение пакетов на два типа и выделение пакетов с информационной нагрузкой. Преимущество данного информативного признака будет заключаться в более оперативной обработке трафика, т.к. анализу будет подвергаться не совокупность информационных пакетов с речевой нагрузкой и паузами, а только информационные речевые пакеты.

Литература

1. Mazurchik V., Homansard A. Information hiding in communication networks // The Institute of Electrical and Elecronics Engineers, Inc. 2016. Vol. 172. P. 172–175.

2. Юсупов Р.М. Статистические методы обработки результатов наблюдений / Р.М. Юсупов. — Л.: МО СССР. 1984. 563 с.

3. Number of estimated «Skype» users worldwide from 2009 to 2024 [Электронный ресурс] // Statistics.com :URL:https://www.statista.com/statistics/820384/estimated-number-»Skype»-users-worldwide (дата обращения: 22.11.2020).

4. Кремер Н.Ш. Теория вероятностей и математическая статистика: Учебник для вузов. — М.: ЮНИТИ-ДАНА. 2004. 573 с.

5. Кузьмин В.В. Структурно-функциональная модель объекта радиомониторинга на основе комплексирования информации с применением математического аппарата теории информационного поиска // Вопросы оборонной техники. Серия 16. Технические средства противодействия терроризму. 2018. № 9–10 (123–124). С. 138–144.

6. Олифер В.Г., Олифер Н.А. Компьютерные сети. Принципы, технологии, протоколы: Учебник для вузов. 4-е изд. — СПб.: Питер. 2010. 944 с.

7. Чабан Л.Н. Теория и алгоритмы распознавания образов. Учебное пособие. — М.: МИИГАиК. 2004. 70 с.

8. Егоров А.С., Клецков Д.А. Применение методов технического маскирования в пиринговых файлообменных сетях // Инновационная деятельность в Вооруженных Силах Российской Федерации. Труды всеармейской научно-практической конференции. Министерство обороны Российской Федерации, Международная академия авторов научных открытий и изобретений (Санкт-Петербургское отделение). Военная академия связи. 2019. С. 116–119.

9. Еремеев И.Ю., Клецков Д.А., Шишкалов А.В. Информативный признак для распознавания служебных пакетов канала управления спутниковых систем связи с многостанционным доступом и временным разделением абонентов // Информация и космос. 2015. № 3. С. 50–61.

10. Кузьмин В.В., Козлов С.Ю. Методика построения маршрутов движения объектов в районе сбора информации // Известия Тульского государственного университета. Технические науки. 2020. № 9. С. 225–231.

References

1. Mazurchik V., Homansard A. Information hiding in communication networks // The Institute of Electrical and Elecronics Engineers, Inc. 2016. Vol. 172. P. 172–175.

2. Usupov R.M. Statistical methods for processing the results of observations // R.M. Usupov. — L.: MO SSSR. 1984. 563 p.

3. Number of estimated «Skype» users worldwide from 2009 to 2024 [Electronic resource] // Statistics.com URL:https://www.statista.com/statistics/820384/estimated-number-»Skype»-users-worldwide (accessed: 22.11.2020).

4. Kremer N.SCH. Theory and mathematical statistics]:Uchebnik dlja vuzov. — M.: UNITI-DANA. 2004. 573 p.

5. Kuzmin V.V. Structural and functional model of the monitoring object based on information aggeration using the mathematical apparatus of information search theory // Military Enginery. Issue 16. Counter-terrorism technical devices. 2018. № 9–10 (123–124). P. 138–144.

6. Olifer V.G., Olifer N.A. Computer networks. Principles, technologies, protocols: Textbook for universities. 4th ed. — SPb.: Piter. 2010. 944 p.

7. Chaban L.N. Theory and algorithms of pattern recognition. Training manual. — M.: MIIGAiK. 2004. 70 p.

8. Egorov A.S., Kletskov D.A. Application of methods of technical masking in peer-to-peer file-sharing networks // Innovative activity in the Armed forces of the Russian Federation. Proceedings of the All-Army Scientific and Practical conference. Ministry of Defense of the Russian Federation, International Academy of Authors of Scientific Discoveries and Inventions (St. Petersburg Branch). Military Academy of Communications. 2019. P. 116–119.

9. Eremeev I.U., Kletskov D.A., Shishkalov A.V. Informative sign for the recognition of service packets of the control channel of satellite communication systems with multi-station access and time division of subscribers // Information and Space. 2015. № 3. P. 50–61.

10. Kuzmin V.V., Kozlov S.U. Methodology for building routes of objects in the area of information collection // Proceedings of the Tula State University. Technical sciences. 2020. № 9. P. 225–231.