

Сети телекоммуникаций и информатика

УДК 621.39

Модель установления сессии по протоколу SIP с учётом повторных передач

М. М. Нсангу, Н. В. Серебренникова

*Кафедра систем телекоммуникаций
Российский университет дружбы народов
ул. Миклуто–Маклая, д.6, Москва, 117198, Россия*

В статье построена модель установления сессии по протоколу SIP с учётом повторной передачи сообщений, когда на транспортном уровне для передачи сообщений применяются протоколы TCP и RTP/RTCP. Математическая модель разработана в виде цепи Маркова, предложен метод её анализа и получены формулы для расчёта её вероятностно-временных характеристик. Теоретические результаты проиллюстрированы численно.

Ключевые слова: SIP, TCP, цепь Маркова, вероятностно-временные характеристики, среднее время установления сессии.

1. Введение

Сеть связи следующего поколения (Next Generation Network, NGN) — это концепция построения сетей связи, обеспечивающих предоставление практически неограниченного набора услуг с гибкими возможностями по их управлению, персонализации и созданию новых услуг. Основная технологическая идея сетей NGN состоит в разделении транспортных процессов и процессов управления вызовами. Протокол SIP (Session Initiation Protocol) — один из основных протоколов NGN — позволяет в полной мере осуществлять необходимые функции управления вызовами. Являясь протоколом прикладного уровня, протокол SIP поддерживает управление сеансами между абонентами NGN и обеспечивает базу для предоставления услуг связи [1, 2]. Для организации услуг VoIP (Voice over IP) протокол SIP может использоваться с протоколами транспортного уровня TCP (Transmission Control Protocol) и RTP/RTCP (Real-Time Protocol / RTP Control Protocol) [3]. Протокол TCP предназначен для надёжной передачи данных, а протокол RTP рекомендован IETF для использования с SIP для передачи мультимедийной информации. Протокол RTCP работает только совместно с протоколом RTP и позволяет контролировать его работу с целью обеспечения качества передачи сообщений. Несмотря на широкое распространение, для протокола SIP в настоящее время не стандартизованы методики, позволяющие проводить полный анализ параметров его производительности [4]. Поэтому проблема построения и анализа математических моделей, описывающих процессы работы протокола SIP, является весьма актуальной.

В данной статье построена математическая модель процесса установления сессии по протоколу SIP поверх транспортных протоколов TCP и RTP/RTCP и предложены формулы для расчёта среднего времени установления сессии с учётом ретрансляции сообщений из-за различного рода сбоев и отказов при их передаче по IP-сети. В первом разделе статьи проведён анализ процедур установления сессии по протоколу SIP с учётом повторных передач. Во втором разделе в виде однородной цепи Маркова (ЦМ) построена математическая модель процесса установления сессии SIP и найдены формулы для расчёта её вероятностно-временных характеристик (ВВХ). На основе полученных результатов выведена формула для расчёта среднего времени установления соединения на уровнях протоколов SIP, TCP и RTP/RTCP. В третьем разделе представлены результаты численного анализа.

2. Постановка задачи исследования

Схема сигнального обмена, необходимого для установления сессии по протоколу SIP, показана на рис. 1. Цифры на рисунке соответствуют порядку отправки сигнальных сообщений между пользователями и прокси-серверами, с одной стороны, и между прокси-серверами и шлюзами (сервером DNS и сервером локализации), с другой стороны.

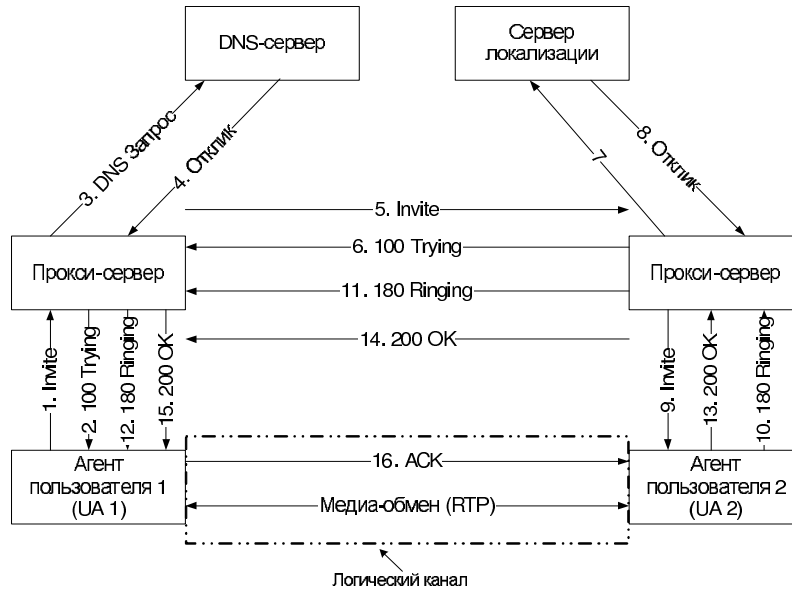


Рис. 1. Схема обмена сигнальными сообщениями

Механизм SIP, использующийся для обмена сообщениями между шлюзами и прокси-серверами, обеспечивает высокую надёжность передачи [1]. Поэтому, с точки зрения анализа возможных потерь и ошибок, наибольший интерес представляет обмен сообщениями между агентами пользователя (User Agent, UA). На рис. 2 построена упрощённая диаграмма установления соединения в условиях нормального функционирования сети. В случае возникновения различных сбоев запускается механизм повторной передачи сообщения, показанный на рис. 3. Отметим, что ретрансляция предусмотрена для сообщений Invite, 200-ОК и ACK. Механизм передачи сообщений 100-Trying и 180-Ringing является надёжным [5], поэтому будем предполагать, что эти сообщения передаются без ошибок.

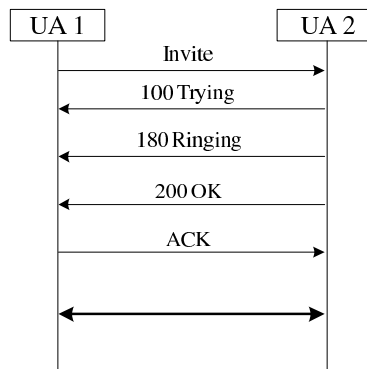


Рис. 2. Упрощённая диаграмма установления сессии

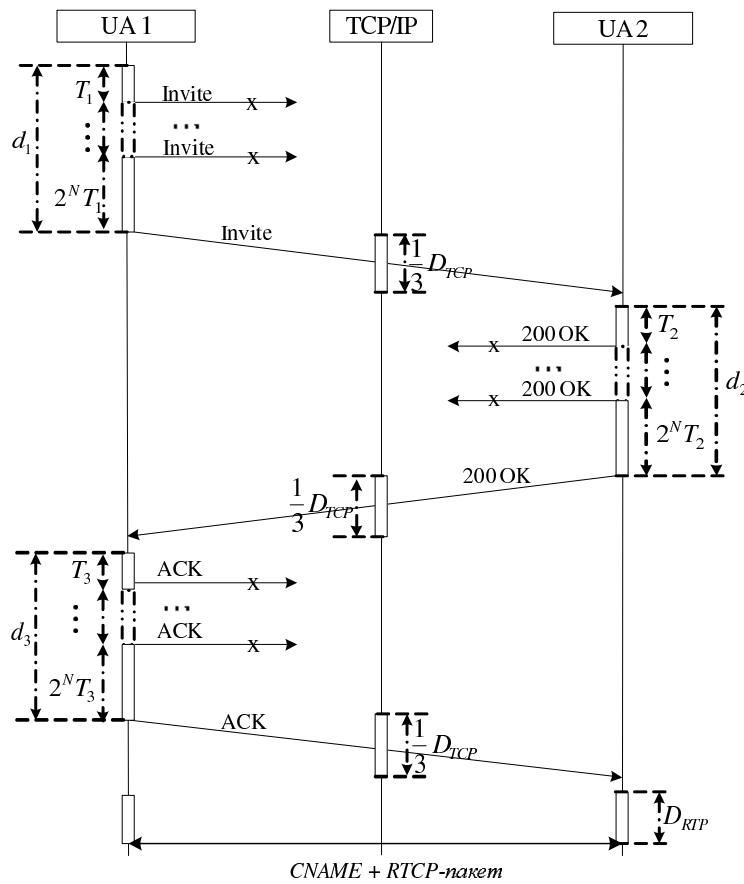


Рис. 3. Диаграмма установления сессии с учётом повторных передач

Повторная передача сообщений на уровне протокола SIP происходит с использованием таймеров [6], контролирующих процесс ретрансляции. Так, в момент передачи сообщения Invite запускаются таймер T_1 и таймер T_{INV} , ограничивающий время ожидания ответа (сообщения типа 100-Trying и 180-Ringing). При срабатывании таймера T_1 происходит повторная передача запроса Invite протоколом SIP на транспортный уровень, и таймер запускается заново, но уже с удвоенным значением. Этот процесс продолжается после каждой повторной передачи до истечения таймера T_{INV} . Аналогично процесс повторной передачи организован и для сообщений 200-OK и ACK.

На практике сообщения 100-Trying и 180-Ringing приходят к агенту пользователя UA1 почти одновременно. Их источником является прокси-сервер (физически он может быть реализован, например, как программный коммутатор), который определяет состояние вызываемого SIP-телефона на основе имеющейся у него информации. Задержку при обработке данных сообщений учитывать в расчёте не будем, поскольку она определяется действиями вызываемого абонента (его статус в момент поступления сообщения Invite, время реакции абонента и др.), а не характеристиками сети связи. Таким образом, среднее время D_{SIP} передачи всех сообщений SIP может быть вычислено по формуле:

$$D_{SIP} = d_1 + d_2 + d_3, \tag{1}$$

где, как показано на рис. 3, величины d_1, d_2, d_3 являются средними значениями временных задержек, возникающих с учётом ретрансляций при передаче сообщений Invite, 200-OK, ACK соответственно. Тогда среднее время установления

сессии D рассчитывается как сумма

$$D = D_{\text{SIP}} + D_{\text{ТСР}} + D_{\text{RTP}}, \quad (2)$$

где $D_{\text{ТСР}}$ — средняя задержка, возникающая в ожидании подтверждения успешной передачи сообщений SIP, D_{RTP} — средняя задержка, возникающая при установлении сессии по протоколу RTP с момента передачи SIP-терминалом сообщения АСК до момента приёма первого сообщения CNAME и приклеенного к нему RTP-пакета [7]. Заметим, что согласно [5] можно принять, что $D_{\text{ТСР}} = \frac{3}{2}\text{RTT}$, где RTT (Round Trip Time) — средняя величина сетевой задержки в петле связи.

3. Математическая модель расчёта ВВХ

Построим в виде ЦМ модель установления сессии по протоколу SIP. Ограничим величиной N число повторных передач сообщений Invite, 200-ОК, АСК и заметим, что по стандарту $N = 7$ [6]. Обозначим g вероятность повторной передачи SIP-сообщения в прямом направлении (от UA1 к UA2), и f — вероятность повторной передачи сообщения в обратном направлении.

Будем считать, что успешная передача SIP-терминалом сообщений Invite, 200-ОК или АСК происходит по истечении таймеров $T_1^{n_1}$, $T_2^{n_2}$, $T_3^{n_3}$ соответственно, где n_1 , n_2 и n_3 , $1 \leq n_1, n_2, n_3 \leq N$ — номера успешных попыток повторной передачи. Обозначим состояния протокола SIP в соответствии с шагами процедуры установления сессии и построим пространство \aleph состояний ЦМ (табл. 1).

Таблица 1

Соответствие пространства состояний \aleph состояниям протокола SIP

Номер состояния ЦМ	Состояние протокола SIP	Описание
0	«старт»/ «финиш»	ожидание запроса из внешней среды / транзакция заблокирована
1	«I»	Передача сообщения Invite
$n_1 + 1$, $1 \leq n_1 \leq N$	« $T_1^{n_1}$ »	n_1 -ая повторная передача сообщения Invite, $1 \leq n_1 \leq N$
$N + 2$	«200»	Передача сообщения 200-ОК
$n_2 + N + 2$, $1 \leq n_2 \leq N$	« $T_2^{n_2}$ »	n_2 -ая повторная передача сообщения 200-ОК, $1 \leq n_2 \leq N$
$2N + 3$	«АСК»	Передача сообщения АСК
$n_3 + 2N + 3$, $1 \leq n_3 \leq N$	« $T_3^{n_3}$ »	n_3 -ая повторная передача сообщения АСК, $1 \leq n_3 \leq N$
$3N + 4$	«setup»	Сессия установлена

Отметим, что при передаче сообщений возможны два варианта:

1. Успешная передача с вероятностью $(1 - g)$ для Invite и АСК и с вероятностью $(1 - f)$ для 200-ОК.
2. Сбой передачи с вероятностью g для Invite и АСК и с вероятностью f для 200-ОК. В этом случае происходит ретрансляция (повторная передача) сообщения.

Тогда процесс установления сеанса описывается однородной ЦМ, диаграмма переходов которой на множестве состояний \aleph показана на рис. 4.

Обозначим n число шагов ЦМ до первого попадания в состояние $3N + 4$ («Setup») из состояния 1 («Invite»), $n = 3, \dots, 3N + 3$, и введём множество

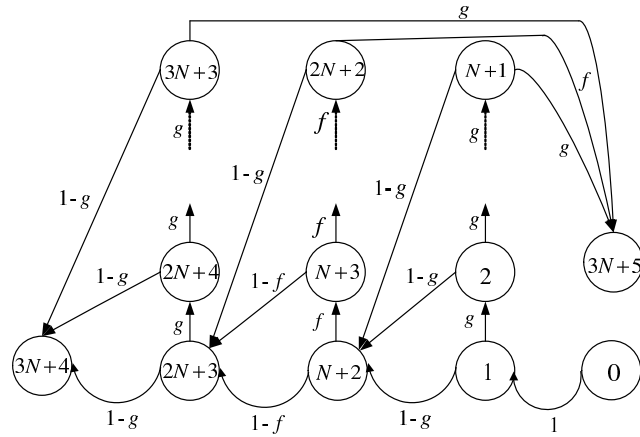


Рис. 4. Диаграмма переходов состояний ЦМ

$\mathbf{N}(n) = \{(n_1, n_2, n_3) : n_1 + n_2 + n_3 = n - 3; 0 \leq n_1, n_2, n_3 \leq N\}$. Пусть $f_{1,3N+4}^{(n_1, n_2, n_3)}$ — вероятность перехода ЦМ из состояния 1 в состояние $3N+4$ за $n = n_1 + n_2 + n_3 + 3$ шагов (соответствует вероятности успешного установления сессии за n шагов) такого, чтобы число повторных передач сообщения Invite было равно n_1 , сообщения 200-ОК — n_2 и сообщения АСК — n_3 . Тогда вероятность $P(f, g)$ успешного установления сессии по протоколу SIP с учётом повторных передач может быть вычислена по формуле:

$$P(f, g) = \sum_{n=3}^{3N+3} \sum_{(n_1, n_2, n_3) \in \mathbf{N}(n)} f_{1,3N+4}^{(n_1, n_2, n_3)}. \quad (3)$$

Нетрудно показать, что

$$f_{1,3N+4}^{(n_1, n_2, n_3)} = \alpha g^{n_1+n_3} f^{n_2}, \quad (4)$$

где множитель $\alpha = (1-g)^2(1-f)$ равен вероятности успешной передачи сообщений Invite, 200-ОК и АСК. Подставляя (4) в (3), получаем:

$$P(f, g) = \alpha \left(\sum_{n_1=0}^N g^{n_1} \right) \left(\sum_{n_2=0}^N f^{n_2} \right) \left(\sum_{n_3=0}^N g^{n_3} \right) = \alpha \left(\frac{1-g^{N+1}}{1-g} \right)^2 \left(\frac{1-f^{N+1}}{1-f} \right), \quad (5)$$

$$\bar{P}(f, g) = 1 - P(f, g) = 1 - (1-f^{N+1})(1-g^{N+1})^2. \quad (6)$$

Заметим, если принять равными вероятности сбоя при передаче сообщений в прямом и обратном направлениях, т.е. $f = g$, то $P(g) = (1-g^{N+1})^3$.

Перейдём к анализу временных задержек установления сессии. Введём $\gamma(n_1, n_2, n_3)$ — доля сессий, установленных за $n = n_1 + n_2 + n_3 + 3$ шагов, т.е.

$$\gamma(n_1, n_2, n_3) = \frac{f_{1,3N+4}^{(n_1, n_2, n_3)}}{P(f, g)}. \quad (7)$$

Задержка $d_1(n_1)$ передачи сообщения Invite с учётом числа n_1 ретрансляций, $0 \leq n_1 \leq N$, может быть найдена по формуле:

$$d_1(n_1) = \sum_{l=0}^{n_1} 2^l T_1 = (2^{n_1} - 1) T_1. \quad (8)$$

Аналогичные формулы могут быть получены и для задержек $d_2(n_2)$ и $d_3(n_3)$ передачи сообщений 200-ОК и АСК. Тогда общая задержка $D_{\text{SIP}}(n_1, n_2, n_3)$ с учётом повторных передач рассчитывается как

$$D_{\text{SIP}}(n_1, n_2, n_3) = d_1(n_1) + d_2(n_2) + d_3(n_3) = (2^{n_1} - 1)T_1 + (2^{n_2} - 1)T_2 + (2^{n_3} - 1)T_3, \quad (9)$$

а среднее значение D_{SIP} задержки установления сессии с учётом повторных передач сообщений Invite, 200-ОК и АСК по протоколу SIP может быть найдено по формуле

$$D_{\text{SIP}} = \sum_{n=3}^{3N+3} \sum_{(n_1, n_2, n_3) \in \mathbf{N}(n)} \gamma(n_1, n_2, n_3) D_{\text{SIP}}(n_1, n_2, n_3). \quad (10)$$

Далее будем считать, что начальные значения таймеров $T_1 = T_2 = T_3 = T$, и, тогда (9) принимает вид:

$$D_{\text{SIP}}(n_1, n_2, n_3) = T(2^{n_1} + 2^{n_2} + 2^{n_3} - 3). \quad (11)$$

Подставляя (4), (7), (11) в (10), получаем, что

$$D_{\text{SIP}} = T \sum_{n=3}^{3N+3} \sum_{(n_1, n_2, n_3) \in \mathbf{N}(n)} \frac{\alpha g^{n_1+n_3} f^{n_2}}{P(f, g)} (2^{n_1} + 2^{n_2} + 2^{n_3} - 3) = T \left(\frac{\alpha}{P(f, g)} \chi - 3 \right), \quad (12)$$

$$\text{где } \chi = 2 \left(\frac{1-(2g)^{N+1}}{1-2g} \right) \left(\frac{1-f^{N+1}}{1-f} \right) \left(\frac{1-g^{N+1}}{1-g} \right) + \left(\frac{1-g^{N+1}}{1-g} \right) \left(\frac{1-(2f)^{N+1}}{1-2f} \right) \left(\frac{1-g^{N+1}}{1-g} \right).$$

При одинаковых вероятностях сбоев в сети в прямом и обратном направлениях, т.е. при $f = g$, получаем

$$D_{\text{SIP}} = 3T \left(\frac{(1-g)}{(1-g^{N+1})} \frac{(1-(2g)^{N+1})}{(1-2g)} - 1 \right). \quad (13)$$

Найдём теперь среднее значение D_{RTP} задержки установления сессии по протоколу RTP. Сессия RTP начинается после успешной передачи подтверждения АСК по протоколу SIP. Задержка D_{RTP} состоит из времени успешной передачи сигнального сообщения CNAME и прикреплённого к нему RTCP-пакета. Пусть τ — интервал времени между передачей фреймов RTCP-пакета, g' — вероятность потери фрейма, k — число фреймов в RTCP-пакете, $q = 1 - (1 - g')^k$ — вероятность потери RTCP-пакета. Согласно [6], время передачи сообщения CNAME и RTCP-пакета по протоколу RTP может быть найдено по формуле:

$$D_{\text{RTP}} = \left[\frac{2 - (1 - g')^k}{(1 - g')^k} \right] \frac{T_{\text{RTCP}}}{2} + \frac{\text{RTT}}{2} + (k - 1) \tau. \quad (14)$$

Здесь $\frac{T_{\text{RTCP}}}{2}$ — среднее время передачи RTCP-пакета; $\frac{\text{RTT}}{2}$ — время успешной передачи сообщения CNAME по IP-сети.

Будем предполагать, что $g' = f = g$ и $k = 1$. Тогда с учётом (13) и (14), формула (2) принимает вид:

$$D = 2RTT + \left[\frac{1+g}{1-g} \right] \frac{T_{\text{RTCP}}}{2} + 3T \left(\left(\left(\frac{1-g}{1-g^{N+1}} \right) \left(\frac{1-(2g)^{N+1}}{1-2g} \right) \right) - 1 \right). \quad (15)$$

Таким образом, формула (15) определяет среднее время установления соединения по протоколу SIP поверх протоколов TCP и RTP/RTCP с учётом повторных передач.

4. Пример численного анализа

Исходные данные для приведённого численного анализа разработаны с использованием [6, 7]. Будем считать, что $g' = f = g \leq 10^{-1}$, $T = 0,5$ с, $T_{RTSP} = 5$ с, $RTT = 0,425$ с, число фреймов в RTSP-пакете $k = 1$.

На рис. 5 построен график зависимости вероятности потери вызова от вероятности g повторной передачи сигнала сообщения. Можно убедиться в эффективности установления сессии по протоколу SIP, поскольку даже при критической ситуации, когда из десяти пакетов теряется один ($f = g = 10^{-1}$), вероятность потери вызова не превышает 10^{-7} , что не противоречит требованию международных стандартов — вероятность потери IP-пакетов в сетях следующего поколения не должна превышать 10^{-3} [8].

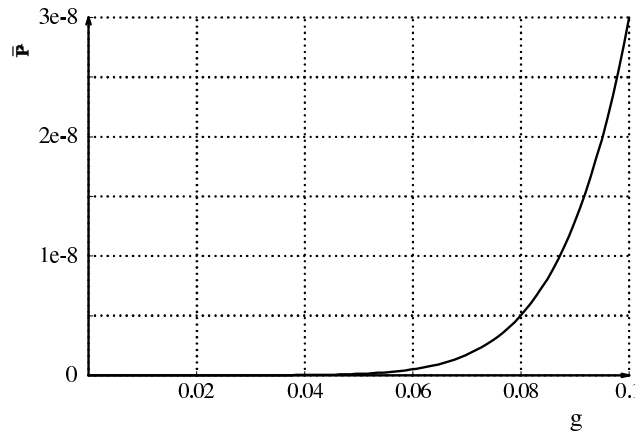


Рис. 5. Зависимость вероятности потери вызова от вероятности повторной передачи g при установлении соединения по протоколу SIP

На рис. 6 показана зависимость времени ретрансляции SIP-сообщений от максимального числа повторных передач. Из графика видно, что при увеличении вероятности g потери сигнала сообщения и числа возможных ретрансляций наблюдается рост задержки D_{SIP} . При $g = 10^{-1}$ и $N = 7$ задержка ретрансляции сообщений SIP не превышает 0,19 с, что не противоречит нормам международного стандарта [9], ограничивающего время установления сессии без учёта протоколов TCP и RTP/RTSP значением 3 с. Общая задержка установления соединения, рассчитанная по формуле (15), при $g = 10^{-1}$ и $N = 7$ достигает 4,093 с.

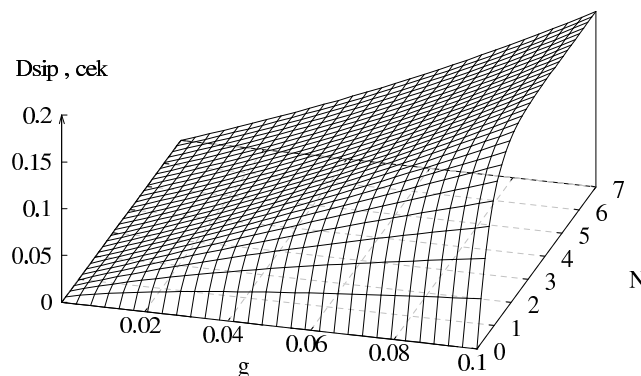


Рис. 6. Зависимость времени ретрансляции SIP-сообщений от числа повторных передач и вероятности повторной передачи

5. Заключение

В статье представлена математическая модель процесса установления сессии SIP с учётом повторных передач поверх протоколов TCP и RTP/RTCP. На основе результатов анализа модели получена формула для расчёта общей задержки, возникающей на уровнях протоколов SIP, TCP и RTP/RTCP. Для проверки адекватности предложенных формул проведён численный эксперимент, результаты которого не противоречат требованиям международных стандартов к нормативным значениям задержек установления сессии по протоколу SIP.

Литература

1. Гольдштейн Б. С., Зарубин А. А., Саморезов В. В. Протокол SIP: справочник. — СПб.: БХВ-Санкт-Петербург, 2005.
2. Системы сигнализации в сетях с коммутацией каналов и пакетов / А. И. Летников, А. П. Пшеничников, Ю. В. Гайдамака, А. В. Чукарин. — М.: МТУСИ, 2008.
3. Schulzrinne H., Casner S., Frederick R., Jacobson V. A Transport Protocol for Real-Time Applications. RFC 1889. — 1996.
4. Серебренникова Н. В., Хатуниев А. Б. Эволюция нормирования параметров качества протокола SIP // Вестник связи. — 2009. — № 5. — С. 10–14.
5. Самуйлов К. Е., Лузгачев М. В., Плаксина О. Н. Разработка вероятностной модели для анализа показателей качества протокола инициирования сеансов связи // Вестник РУДН, серия «Математика. Информатика. Физика». — 2007. — № 3.
6. De Marco G., Iacovoni G., Barolli L. A Technique to Analyze Session Initiation Protocol Traffic // Proceedings of the 11th International Conference on Parallel and Distributed Systems 11th. — 2005. — Vol. 2. — Pp. 595–599.
7. Performance Optimization of VoIP Calls over Wireless Links using H.323 Protocol / S. K. Das, E. Lee, K. Basul et al // IEEE Transactions on Computers. — 2003. — Vol. 52, No 6. — Pp. 742–752.
8. ITU-T Recommendation Y.1541, Network Performance Objectives for IP-Based Services. — 2002.
9. ITU-T Recommendation E.721, Network Grade of Service Parameters and Target Values for Circuit-Switched Services in the Evolving ISDN. — 1991.

UDC 621.39

SIP-Session Model with the Message Retransmission Mechanism

M. M. Nsangou, N. V. Serebrennikova

*Telecommunication Systems Department
Peoples' Friendship University of Russia
6, Miklukho-Maklaya str., 117198, Moscow, Russia*

In this paper we propose a SIP-session model with the message retransmission mechanism and TCP and RTP/RTCP as the underlying protocols. The mathematical model is developed using the Markov chain theory, the analysis method is described. The theoretical results are numerically illustrated.

Key words and phrases: SIP, TCP, Markov chain, performance parameters, mean session setup delay.