

## ОБ ОДНОМ МЕТОДЕ УПРАВЛЕНИЯ БУФЕРОМ ДЛЯ ПРЕДОТВРАЩЕНИЯ ПЕРЕГРУЗКИ В SIP-СЕРВЕРАХ В ПОДСИСТЕМЕ IMS

Гамиль Абдуллах

*СПбГУТ «Санкт-Петербургский государственный университет телекоммуникаций им. проф. М.А. Бонч-Бруевича», Санкт-Петербург, Россия (193232, Санкт-Петербург, пр. Большевиков д.22, корпус 1), e-mail: obad-85a@mail.ru*

---

Важность исследования функционирования подсистемы IMS в настоящий момент достаточно очевидна. Она включает задачу анализа производительности серверов управления сеансами (P-CSCF, I-CSCF, S-CSCF) и показателей эффективности их функционирования. Для установления, модификации и разъединения сессий пользователей используется протокол установления сессий SIP. Задача оценки объема сигнальной нагрузки и времени установления сеанса в подсистеме остается малоисследованной, так как для протокола SIP до настоящего времени не стандартизованы методики, позволяющие проводить анализ параметров его производительности. В статье описано влияние функций управления сеансами на время установления сеанса, показана необходимость дополнительного исследования обработки сообщений функцией S-CSCF ввиду ее высокой загруженности и возможности возникновения перегрузки в час наибольшей нагрузки. Разработан метод управления буфером с целью предотвращения перегрузок функции S-CSCF посредством введения классификации и последующей приоритизации сигнальных сообщений, а также введения пороговых значений в очереди поступающих на узел S-CSCF сигнальных сообщений.

Ключевые слова: IMS, качество обслуживания, установление сеанса связи, время пребывания, SIP, перегрузка

## ON THE BUFFER MANAGEMENT METHOD TO CONGESTION AVOIDANCE IN SIP SERVERS IN IP MULTIMEDIA SUBSYSTEM

Gamil Abdullah

*The Bonch-Bruevich Saint-Petersburg State University of Telecommunications (SPbSUT), Saint-Petersburg, Russia (193232, Saint-Petersburg, Prospekt Bolshevikov, 22, korpus 1), e-mail: obad-85a@mail.ru*

---

Currently the importance of investigation for functioning IMS subsystem is fairly obvious. It includes the task of the performance analyzing management server by sessions (P-CSCF, I-CSCF, S-CSCF) and performance metrics of their functioning. To establish, modify, and disconnect user sessions used session Initiation Protocol SIP. The assessing task of the volume of the load signaling and session setup time in the subsystem is less studied, since the SIP protocol has not yet been standardized techniques to analyze the parameters of its performance. This paper considers the functions impact for the management session at the time of the session setup show the need for further study of the function messages processing S-CSCF due to its high workload and the possibility of congestion in the busy hour. Also has been developed a method buffer management in order to prevent overload function S-CSCF by the introduction classification and prioritization followed signaling messages, as well as the introduction of threshold values in the queue incoming signaling messages to node S-CSCF.

Keywords: IMS, quality of service, session establishment, sojourn time, SIP, overload

В настоящее время подсистема IMS (*IP Multimedia Subsystem*) рассматривается как возможное решение для построения сетей следующего поколения (СПП) и как основа конвергенции мобильных и стационарных сетей на платформе IP. При этом при инициации и модификации сеансов связи в сетях СПП международными организациями по стандартизации был принят протокол SIP (*Session Initiation Protocol*). Данный протокол характеризуется относительно небольшой чувствительностью к параметрам качества обслуживания (Quality of service, QoS). Основные показатели QoS в подсистеме IMS, применяемые как в мобильных, так и в фиксированных сетях, были приведены в [2].

Основными функциональными элементами IMS являются функция управления сеансами связи и маршрутизацией (Call Session Control Function, CSCF) и сервер домашних абонентов (Home Subscriber Server, HSS). Функциональность CSCF включает три составные части – функцию обслуживания (Serving CSCF, S-CSCF), функцию запроса (Interrogating CSCF, I-CSCF) и функцию прокси-сервера (Proxy CSCF, P-CSCF) [4].

Однако с увеличением числа услуг, характеризующихся высокими требованиями к производительности серверов управления сеансами связи, предоставляемых с использованием протокола SIP на базе сети IMS, значительно увеличивается и нагрузка на этих серверах, порождающая, соответственно, возможность (с некоторой вероятностью) перегрузки сервера. За последние годы опубликован ряд работ, посвященных смежным исследованиям, а именно методам контроля перегрузок в серверах протокола установления сеансов в SIP-сети. Так, комитет IETF в зависимости от типа перегрузок выделяет ряд решений проблемы (например, в RFC 7339 (2014-09-23) и RFC 7415 (2015-02-05) и др.), среди которых: увеличение числа SIP-серверов, механизм 503, метод просеивания потока, метод снижения скорости. Однако оптимального решения для управления перегрузок на SIP-сервере не найдено.

Анализ сигнальных сообщений протокола SIP в подсистеме IMS, представленных в [1], показал, что функция S-CSCF представляет собой критичный к перегрузке элемент, «узкое место» сети по сравнению с другими серверами IMS. В этой статье приведены результаты анализа возможной перегрузки функцией S-CSCF, описана исследуемая модель СМО функционирования сервера и предложен метод управления буфером этой функции с целью предотвращения этой ситуации.

### **Метод управления очередью в S-CSCF**

Предлагается упрощенный механизм контроля перегрузок, который позволяет осуществить управление нагрузкой на SIP-сервер путем применения метода классификации и приоритетов очереди SIP-сообщений и введения порогов для сообщений, поступающих на S-CSCF. Для упрощения моделей, поступающих на сервер, сигнальные сообщения были разделены на три класса. Первый класс включает в себя типы SIP-сообщений, которые завершают сеанс связи, такие как BYE или CANCEL, и те сообщения, которые возникают только после успешной передачи сообщения: ACK или 2xx. Этот класс сообщений имеет наивысший приоритет в условиях перегрузки. Сообщения REGISTER, MESSAGE, PUBLISH, NOTIFY и SUBSCRIBE отнесены ко второму классу, и им назначается средний приоритет. Третий класс содержит типы SIP-сообщений, которые инициируют сеанс связи (INVITE), и те, которые обеспечивают промежуточные ответы, такие как 1xx. Третьему классу предоставляется наименьший приоритет.

## Построение модели СеМО предлагаемого метода

Рассмотрим упрощенную модель IMS, включающую в себя только единичные случаи различных CSCF. Построение модели основано на исследовании четырех процедур установления сеансов: регистрация, установление/прекращение сеанса, обмен мгновенными сообщениями, а также представление информации о присутствии пользователей. Поток сигналов, передаваемых в сети маршрутизации, попадает в функциональные элементы сети (один из CSCF) несколько раз, т. е. каждый функциональный объект в сети маршрутизации обрабатывает поток сигналов трафика сигнализации несколько раз [6]. Функциональные объекты обслуживают сигнальные потоки в узле, а затем направляют трафик дальше.

Функционирование системы будем моделировать в виде открытой сети массового обслуживания (СеМО), состоящей из узлов  $M = \{1, 2, 3\}$ , в которой циркулируют заявки нескольких классов из множества  $N = \{1, 2, 3\}$ , где каждый исследуемый узел реализован в виде системы массового обслуживания (рис. 1). Функции P-CSCF и I-CSCF представляют собой СМО типа M/M/1 с дисциплиной обслуживания «FIFO – первый пришел, первый ушел», а функция S-CSCF как СМО типа M/G/1, где выбор дисциплины обслуживания зависит от того, обнаружена или нет перегрузка. Поступающие в систему заявки трех классов образуют простейшие потоки с интенсивностями  $\lambda_1$ ,  $\lambda_2$  и  $\lambda_3$  соответственно.

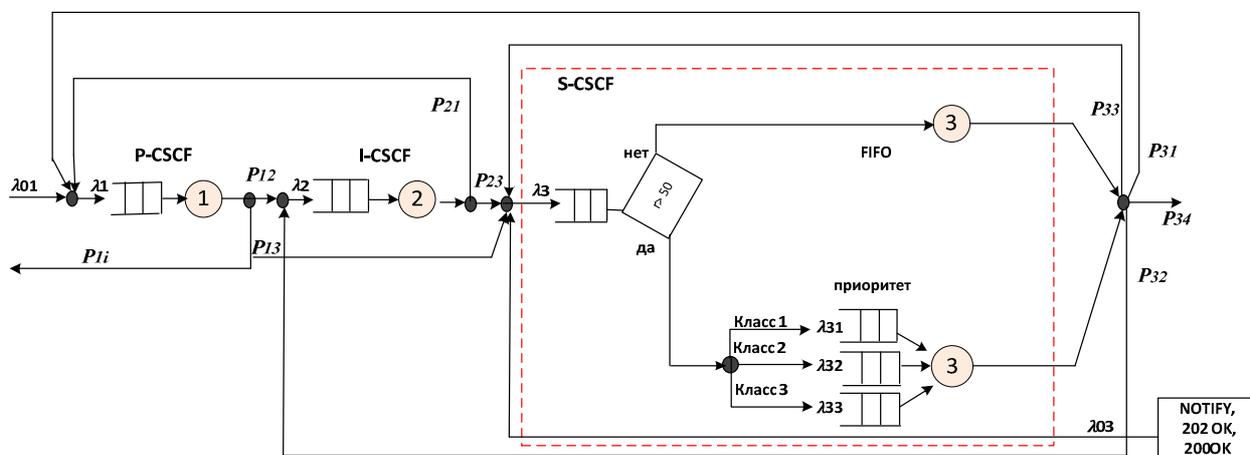


Рис. 1. Модель СеМО управления очередями в S-CSCF

Согласно рис.1, можно составить систему уравнений (1) – (3):

$$\lambda_1 = \lambda_{01} + P_{21}\lambda_2 + P_{31}\lambda_3 \quad (1)$$

$$\lambda_2 = P_{12}\lambda_1 + P_{32}\lambda_3 \quad (2)$$

$$\lambda_3 = \lambda_{03} + P_{13}\lambda_1 + P_{23}\lambda_2 + P_{33}\lambda_3 \quad (3)$$

где,  $\lambda_1$  – интенсивность прибытия заявок в P-CSCF,  $\lambda_2$  – интенсивность прибытия заявок в I-CSCF,  $\lambda_3$  – интенсивность прибытия заявок в S-CSCF,  $P_{ij}$  – характеризует долю заявок, переходящих из узла  $i$  в узел  $j$ ,  $ij \in M$ . Интенсивность прибытия заявок  $\lambda_{01}$  и  $\lambda_{03}$  определяется следующим:

$$\lambda_{01} = \lambda_{INVITE} + \lambda_{RINGING} + \lambda_{100TRYING} + \lambda_{MESSAGE} + \lambda_{SUBSCRIBE} + \lambda_{REGISTER} + \lambda_{PUBLISH} + \lambda_{ACK} + \lambda_{BYE} + \lambda_{200OK} + \lambda_{200OK} \quad (4)$$

$$\lambda_{03} = \lambda_{NOTIFY} + \lambda_{200OK} + \lambda_{202OK} \quad (5)$$

В соответствии с построенной моделью можно найти интенсивность прибытия заявок разных классов в S-CSCF таким образом:

$$\lambda_{3n} = P_n \lambda_3, \quad n=1,2,3$$

и  $\sum_{n=1}^3 P_n = 1$ ,  $P_n$  – вероятность прибытия заявок класса,  $n$ .

По сути, для успешного управления перегрузками требуется введение порогового управления очередью на обработку сигнальных сообщений сервером S-CSCF, которое позволит определить начало перегрузки – порог обнаружения перегрузки  $r$ ,  $r = 50$ . Пока общее число сообщений в очереди не превышает порога  $r$ , сервер функционирует в режиме нормальной нагрузки с дисциплиной обслуживания FIFO. При  $r > 50$  S-CSCF работает в режиме перегрузки с дисциплиной обслуживания по приоритету. Ожидающие обслуживания заявки разнесены по разным накопителям ограниченной емкости. Между заявками разных классов установлены относительные приоритеты, означающие, что всякий раз из накопителей на обслуживание выбирается заявка с самым высоким приоритетом. Математическое ожидание времени пребывания заявок первого и второго приоритета –  $\bar{T}_1$  и  $\bar{T}_2$  можно представить в следующем виде [3]:

$$\bar{T}_1 = \frac{2 - \rho_1 + \rho_2}{2(1 - \rho_1)} \cdot \bar{X}$$

$$\bar{T}_2 = \frac{(\rho_1 + \rho_2) + 2(1 - \rho_1)(1 - \rho_1 - \rho_2)}{2(1 - \rho_1)(1 - \rho_1 - \rho_2)} \cdot \bar{X}$$

где  $\rho_1 = \lambda_{31} \bar{X}$  и  $\rho_2 = \lambda_{32} \bar{X}$  – нагрузка первого и второго класса заявок, и  $\bar{X}$  – время обслуживания заявок.

### Результаты

На основании расчетов по формулам (4), (5) получено, что в нормальной ситуации нагрузки  $\lambda_{01} = 0,34916$  и  $\lambda_{03} = 0,14716$ , а в ситуации перегрузки  $\lambda_{01} = 0,8729$  и  $\lambda_{03} = 0,3679$ .

Чтобы проверить эффективность нашего метода, дополнительный анализ предложенной модели выполнен с помощью статистического моделирования. Также статистический анализ основан на исходных данных, полученных специалистами British Telecom [5]. Аналитические результаты справедливы, когда время стремится к бесконечности, в то время как продолжительность моделирования ограничивается 2500 единицами времени.

В таблице 1 представлены результаты, полученные в нормальной ситуации нагрузки. Однако при одинаковых нагрузочных параметрах наблюдаются различия в значениях вероятностно-временных характеристик (среднее время ожидания начала обслуживания и средняя длина очереди) при аналитическом и имитационном моделировании. Эти отличия вызваны аппроксимациями в интенсивности прибытия сообщений в этих серверах. Значения вероятностно-временных характеристик, полученных с помощью аналитической и имитационной моделей, могут значительно различаться для нагрузочных параметров, близких к режиму перегрузки.

**Таблица 1**

Сравнение результатов в нормальной ситуации нагрузки

Параметр	Узлы IMS	Аналитические результаты	Результаты моделирования
Загруженность функции CSCF, $U$	P-CSCF	0,3397	0,3853
	I-CSCF	0,048	0,0448
	S-CSCF	0,563	0,5712
Среднее время ожидания, $\bar{w}$	P-CSCF	0,256	0,879
	I-CSCF	0,050	0,0612
	S-CSCF	0,620	3,4938
Средняя длина очереди	P-CSCF	0,174	0,6972
	I-CSCF	0,0024	0,0024
	S-CSCF	0,420	2,55

Рисунок 2 показывает загруженности функций управления CSCF в нормальном режиме и в условиях перегрузки. В ситуации перегрузки показано обслуживание сигнальных сообщений сервером S-CSCF по FIFO и при включении режима обслуживания по приоритету. Функция S-CSCF характеризуется самой высокой нагрузкой. Как следует из рисунка 2, ситуация перегрузки обнаружена при времени ( $T=275$ ). Активация режима обслуживания по приоритету уменьшает нагрузку на S-CSCF.

Среднее время ожидания заявок в очереди иллюстрируется на рисунке 3. Видно, что среднее время ожидания в P-CSCF и I-CSCF ниже, чем среднее время ожидания в S-CSCF, и

примерно постоянно. В режиме перегрузки среднее время ожидания увеличивается линейно в S-CSCF при использовании дисциплины обслуживания FIFO, в то же время при реализации режима обслуживания по приоритету обеспечивается снижение времени ожидания, и оно приблизительно стремится к постоянной величине.

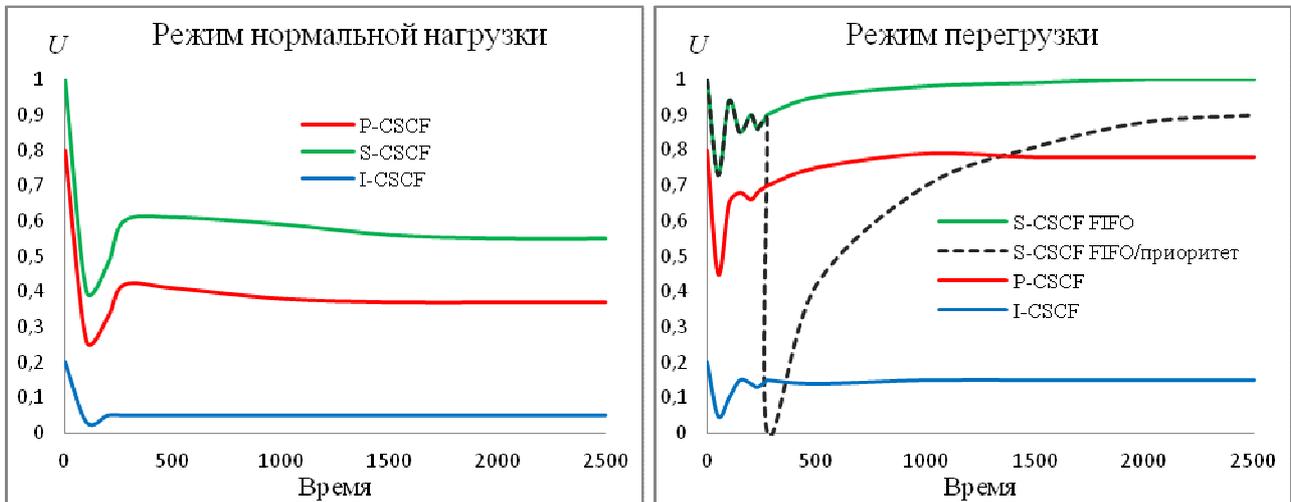


Рис. 2. Загруженности процессора функций управления CSCF

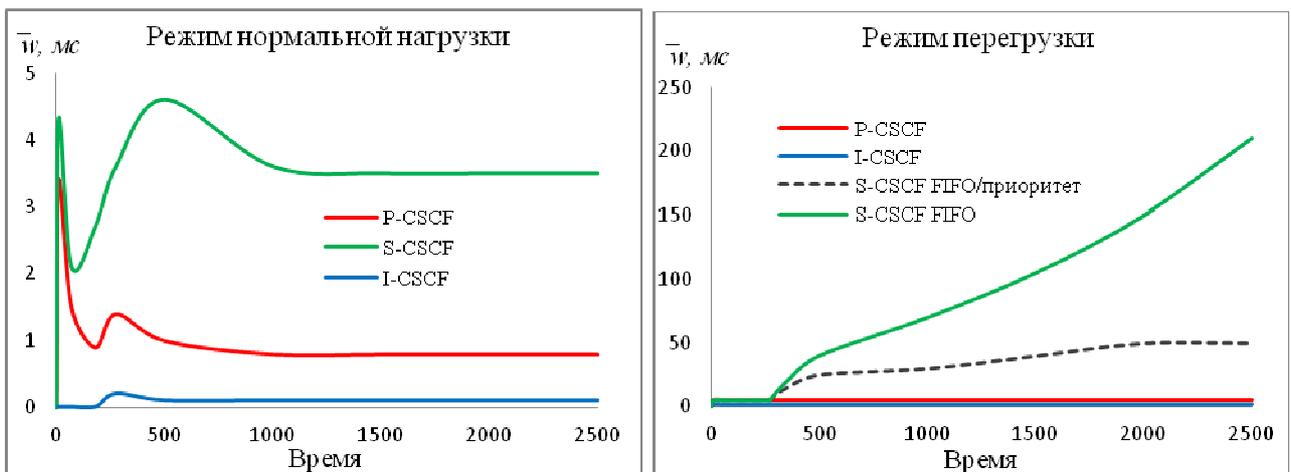


Рис. 3. Среднее время ожидания обслуживания сигнальных сообщений на CSCF

Таблица 2 показывает среднее время пребывания SIP-сообщений разных классов в S-CSCF. Среднее время пребывания почти в 100 раз ниже для сообщений первого класса, которым назначен самый высокий приоритет, в режиме обслуживания по приоритету.

Таблица 2

Среднее время пребывания сообщений в случае перегрузки

SIP сообщений		Класс 1	Класс 2	Класс 3
Среднее время пребывания в S-CSCF	FIFO	195,63	202,29	188,40
	FIFO/Приоритет	2,518	144,14	Не определено

Метод классификации и приоритетов очереди SIP-сообщений увеличивает количество

SIP-сообщений, обслуживаемых в условиях перегрузки. Наименьшее количество таких SIP-сообщений относится к классу 3, который предотвращает создание новой сессии, если не хватает ресурсов. Это обеспечивает гарантированные QoS для уже установленных сессий. Поскольку эти сообщения прекращают существующие сеансы, это позволяет быстрее освободить оккупированные ресурсы и повышает общую QoS. Следует подчеркнуть, что применение предложенного метода влияет на долю успешно установленных сеансов. Можно предположить снижение этого показателя, поскольку отнесение сообщения INVITE к нижнему по приоритету классу сообщений приводит к повторному инициированию сеансов.

### **Заключение**

В статье была построена модель обслуживания сигнальных сообщений сервером S-CSCF с применением метода классификации и приоритизации SIP-сообщений, предназначенного для исследования поведения сервера в моменты перегрузок. Данные, полученные в результате предложенного метода, позволяют сделать вывод о возможности использования предложенного метода управления очередью в случае перегрузок как надежного механизма, предотвращающего переход сервера S-CSCF в перегруженное состояние.

### **Список литературы**

1. Гамиль А.А. Исследование анализа трафика сигнализаций SIP в подсистеме IMS на основе OpenIMSCore. // IV Международная научно-техническая и научно-методическая конференция «Актуальные проблемы инфотелекоммуникаций в науке и образовании»: материалы конференции. – СПб.: СПбГУТ, 2015. – С. 290–294.
2. Гамиль А.А., Куликов Н.А. Построение модели задержки сигнального трафика в сети связи на базе подсистемы IMS // Электросвязь – 2014. – № 9. – С. 8–13.
3. Клейнрок Л. Вычислительные системы с очередями. – М.: Мир, 1979.
4. Яновский Г.Г. IP Multimedia Subsystem: принципы, стандарты и архитектура // М.: Вестник связи, 2006. – № 3. – С. 71–76.
5. Abhayawardhana V.S., Babbage R. A traffic model for the IP Multimedia Subsystem (IMS) // IEEE Magazine – 2007 – PP. 783–787.
6. Juan M., Dirk T., Validation of the Signaling Procedures of a Delivery Platform for IMS Services // International Federation for Information Processing. – 2008. – Vol. 284. – P. 259–270.

### **Рецензенты:**

Гольдштейн Б.С., д.т.н., профессор, заведующий кафедрой инфокоммуникационных систем, Санкт-Петербургский государственный университет телекоммуникаций, г. Санкт-

Петербург;

Кучерявый А.Е., д.т.н., профессор, заведующий кафедрой сетей связи и передачи данных,  
Санкт-Петербургский государственный университет телекоммуникаций, г. Санкт-  
Петербург.