# К анализу эффектов группового поступления сигнальных сообщений на время ожидания начала обслуживания[1]

Ю.В. Гайдамака, Э.Р. Зарипова, Ю.Н. Орлов

**Аннотация.** В работе формулируется подход к анализу зависимости параметров вероятностно-временных характеристик математической модели SIP сервера с групповым поступлением сообщений от распределения длины группы. Развиваемый подход использует непараметрические методы статистического анализа. На основе проведенного анализа были найдены коэффициенты эластичности вероятностно-временных характеристик модели в зависимости от расстояния между распределениями длины группы заявок, что позволило получить поправочные коэффициенты для оценки этих характеристик при законах распределения длины группы заявок, отличных от равномерного.

**Ключевые слова:** оптимизация, математическая модель, вероятностно-временные характеристики, система массового обслуживания, социальные сети, непараметрические статистики.

## Введение

Для предоставления телекоммуникационных услуг широко используется платформа IP-based Multimedia Subsystem - подсистема передачи мультимедийных сообщений на базе протоколов Интернет, основным сигнальным протоколом которой является протокол установления сессий SIP. Сигнальные сообщения протокола SIP создают на узлы сети дополнительную нагрузку, которая влияет на качество предоставления услуг, поэтому исследование структуры сигнального трафика и его особенностей является актуальной задачей. В статье исследуется модель обслуживания сигнального трафика сервером протокола установления сессий. Одной из особенностей трафика, учтенной в данной работе, является ограниченное «время жизни» сигнального сообщения, по истечении которого информация устаревает. Это приводит к необходимости ретрансляции сообщения, что дополнительно нагружает сервер протокола установления сессий. Второй особенностью является групповой характер поступления сообщений на сервер протокола установления сессий. В качестве исследуемых характеристик выбраны средняя длина очереди и среднее время ожидания начала обслуживания. Первая характеристика позволяет проектировщику рассчитать объем буфера сервера протокола установления сессий для минимизации потерь сообщений и снижения числа ретрансляций. Вторая характеристика в сумме со временем обслуживания позволяет оценить время пребывания сообщения в SIP сервере и должна соответствовать времени жизни сообщения.

Структура сигнального трафика, генерируемого в социальных сетях, показывает [1, 2], что

[1] Исследование выполнено при частичной финансовой поддержке РФФИ в рамках научных проектов № 15-07-03051, 15-07-03608, 14-01-00145, 15-01-07944.

обычно более 50 процентов объема сигнального трафика составляют сообщения присутствия (Presence Notify), которые уведомляют подписчиков абонента об изменении его статуса присутствия. При изменении статуса на терминалы всех подписчиков абонента, находящихся online, поступает информация о новом статусе, которая становится доступна на экране абонентского устройства. Для анализа характеристик процесса обслуживания сигнальных сообщений присутствия сервером протокола установления сессий построена математическая модель и получены формулы для расчета вероятностно-временных характеристик процесса, после чего проведена оценка параметров зависимости характеристик от функции распределения длины группы заявок.

Процесс поступления сигнальных сообщений присутствия на сервер протокола установления сессий можно исследовать с помощью однолинейной системы массового обслуживания (СМО) с групповым входящим потоком и прогулками прибора в периоды простоя. Под прогулкой прибора будем подразумевать обслуживание сервером других сигнальных сообщений, отличных от сообщений присутствия. Группа заявок, поступающих на обслуживание, в физической модели соответствует сообщениям Presence Notify для подписчиков абонента, находящихся online.

В [2-4] получены формулы для расчета среднего значения длины очереди и среднего времени ожидания начала обслуживания для произвольных функций распределения длины группы заявок. В [4] проведен анализ характеристик для четырех законов распределения длины группы заявок: Ципфа, геометрического, логарифмического и равномерного. Анализ данных в [4] показал статистическую зависимость исследуемых характеристик от вариации функции распределения длины группы заявок для случая пуассоновского потока заявок. В отличие от работ, в которых исследовалось только геометрическое распределение длины группы заявок, в [4] в качестве предварительной оценки характеристик дана рекомендация по применению равномерного распределения длины группы заявок, что существенно упрощает формулы для инженерных расчетов. В данной статье на основе подхода [4] к анализу вероятностно-временных характеристик используется определенная аппроксимация чувствительности параметров модели процесса обслуживания сигнальных сообщений к виду распределения длины группы заявок.

Вариация функции распределения определялась в [4] в различных нормах, из которых затем были отобраны те, в которых достоверность анализируемой статистической связи была наибольшей. Выявлено, что анализ чувствительности численных оценок параметров модели SIP-сервера к вариации распределения длины группы заявок может быть проведен с высокой степенью достоверности (0,99).

Поскольку для непосредственного анализа доступны лишь выборочные распределения изучаемых параметров, возникают следующие вопросы.

1. Если распределение длины группы заявок стационарно, принимая за таковое некоторое выборочное распределение, можно ли в рамках непараметрического анализа оценить вариацию параметров модели сервера, обусловленную статистическими флуктуациями выборочного распределения?

2. Если распределение нестационарно, то каков его уровень нестационарности и как в таком случае оценить вариацию параметров модели сервера?

Используя результаты [4] анализа чувствительности и методы анализа нестационарных распределений, развитые в [5, 6], можно дать ответы на поставленные вопросы, что и является целью настоящей работы.

Подчеркнем, что на практике измеримой величиной является длина группы заявок, тогда как время ожидания начала обслуживания и длина очереди зависят как от этой величины, так и от времени обслуживания одной заявки и длительности прогулки прибора. Распределения последних двух параметров считаются известными, так как они представляют собой технические характеристики устройства. Однако оценка доверительного интервала для среднего значения длины очереди не может быть получена непосредственно из соответствующего эмпирического распределения, поскольку, как уже говорилось, доступны лишь косвенные измерения, результаты которых могут быть преобразованы к требуемым величинам только в рамках определенных моделей. Для одной из таких моделей в работе [4] и был проведен анализ чувствительности, применяемый далее к эмпирическим распределениям групп заявок по длине.

## 1. Модель SIP сервера в виде системы массового обслуживания с групповым поступлением заявок и прогулкой прибора

Математическая модель процесса обслуживания сигнальных сообщений присутствия исследуется в виде СМО с групповым характером поступления заявок и прогулками прибора на периодах простоя, по классификации Башарина-Кендалла соответствует СМО типа $M^{[X]}|G|1|\infty$. Предположим, что группы заявок поступают в соответствии с пуассоновским потоком с интенсивностью $\lambda$, а время обслуживания заявки - случайная величина с функцией распределения (ФР) $B(t)$, имеющая среднее значение $b_1$ и конечный второй момент $b_2$. В периоды простоя прибор уходит на прогулку, длительность которой есть случайная величина с ФР $V(t)$, для которой существуют первый и второй моменты $v_1$ и $v_2$.

Пусть $f(k)$ - вероятность того, что длина группы заявок равна $k \geq 1$. Соответствующую ФР длины группы заявок обозначим $F(k)$. Для СМО типа $M^{[X]}|G|1|\infty$ с прогулками исследуется средняя длина очереди и среднее время ожидания начала обслуживания в зависимости от нагрузки, поступающей на систему. Нагрузкой $\rho$ сервера называется величина

$$\rho = \lambda b_1 l^{(1)}, \quad (1)$$

где величина $l^{(1)}$ - среднее значение длины группы заявок, соответствующее ФР $F(k)$.

В работе [2] была получена производящая функция случайной величины числа заявок в очереди однолинейного сервера с прогулками прибора. Она имеет вид:

$$P(z) = \frac{1-\rho}{\lambda v_1} \cdot \frac{1-z}{1-L(z)} \cdot \frac{1-\varphi(\lambda, z)}{\beta(\lambda, z) - z}, \quad (2)$$

где $L(z)$ есть производящая функция случайной длины группы заявок для ФР $F(k)$:

$$L(z) = \sum_{k=0}^{\infty} f(k) z^k .$$

Остальные функции, входящие в (2), определены равенствами:

$$v_1 = \int_0^\infty t dV(t), \quad \varphi(\lambda, z) = \int_0^\infty e^{-\lambda t(1-L(z))} dV(t),$$

$$\beta(\lambda, z) = \int_0^\infty e^{-\lambda t(1-L(z))} dB(t).$$

Согласно [4] средняя длина очереди в зависимости от нагрузки (1) определяется через производящую функцию (2) по формуле:

$$N = \lim_{z\to 1} P'(z) = \frac{v_2}{2v_1 b_1}\rho + \frac{l^{(2)} - l^{(1)}}{2l^{(1)}} \frac{\rho}{1-\rho} + \frac{b_2}{2b_1^2}\frac{\rho^2}{1-\rho}. \quad (3)$$

Входящие в (3) моменты распределений определяются формулами:

$$v_s = \int_0^\infty t^s dV(t), \quad b_s = \int_0^\infty t^s dB(t),$$

$$l^{(s)} = \sum_k k^s f(k); \quad s = 1; 2.$$

Среднее время ожидания начала обслуживания заявки определяется формулой:

$$\tau = \frac{Nb_1}{\rho} = \frac{v_2}{2v_1} + \frac{l^{(2)} - l^{(1)}}{2l^{(1)}} \frac{b_1}{1-\rho} + \frac{b_2}{2b_1}\frac{\rho}{1-\rho}. \quad (4)$$

## 2. Чувствительность параметров модели обслуживания сообщений к вариации вероятностей $f(k)$

Рассмотрим, как изменятся величины $N$ и $\tau$, определяемые в (3-4), при малой вариации вероятностей $f(k)$. Пусть новое распределение имеет вид $\tilde{f}(k) = f(k) + \varepsilon(k) f(k)$, причем в силу сохранения нормировки выполняется условие $\sum_k \varepsilon(k) f(k) = 0$. Введем величину $E = \sup_k |\varepsilon(k)|$. Тогда получаем оценку расстояния между распределениями в норме L1:

$$\varepsilon = \left\| f - \tilde{f} \right\|_{L1} = \sum_k \left| f(k) - \tilde{f}(k) \right| = \\ = \sum_k |\varepsilon(k)| \, f(k) \le E. \tag{5}$$

Используя (5), получаем оценку вариации некоторой (дифференцируемой) функции от среднего значения длины группы заявок, если $E << 1$:

$$\left|\delta l^{(1)}\right| = \left|\sum_k k\left( f(k) - \tilde{f}(k)\right)\right| < \sum_k k \left| f(k) - \tilde{f}(k) \right| = \\ \sum_k k \left|\varepsilon(k)\right| f(k) \le E l^{(1)};$$

$$\left|\delta u\left(l^{(1)}\right)\right| = \left|\frac{\partial u\left(l^{(1)}\right)}{\partial\left(l^{(1)}\right)}\right| \cdot \left|\delta l^{(1)}\right| \le Eu \left|\frac{\partial \ln u}{\partial \ln l}\right|\Bigg|_{l=l^{(1)}} + o\left(E\right). \tag{6}$$

Чувствительностью функции к вариации параметра называется логарифмическая производная функции по логарифму параметра. Тогда из (6) следует, что вариация некоторой функции от среднего значения, получаемая при вариации распределения, в линейном приближении по $E$ не превосходит произведения этой функции на супремум вариации плотности распределения и на модуль указанной чувствительности. В частности, линейные по $E$ оценки вариаций средней длины очереди (3) и среднего времени ожидания (4) имеют вид:

$$|\delta N| \le \frac{E\rho}{2} \cdot \\ \cdot \left| \frac{v_2}{v_1 b_1} + \frac{l^{(2)} - l^{(1)}}{l^{(1)}} \frac{b_1}{\left(1-\rho\right)^2} + \frac{b_2}{b_1^2}\left(2-\rho\right)\frac{\rho}{\left(1-\rho\right)^2} \right|; \\ |\delta\tau| \le \frac{E\rho}{2\left(1-\rho\right)^2} \cdot \left| \frac{l^{(2)} - l^{(1)}}{l^{(1)}} b_1 + \frac{b_2}{b_1} \right|. \tag{7}$$

Однако теоретические оценки (7) не дают достаточной точности, поскольку оказываются слишком завышенными, хотя и не улучшаемыми в классе функций, которые выбираются на роль эмпирических функций распределения. Неулучшаемость следует из существования вариации, при которой $\varepsilon(k) = \{0; \pm E\}$, а недостаточная точность вытекает из неравенства (5): так как $\varepsilon \le E$, то для сильно неравномерных распределений норма вариации в виде $\sup_k |\varepsilon(k)|$ слишком груба, ибо расстояние между распределениями может быть тогда заметно меньше, чем $E$. В [4] был проведен анализ чувствительности вышеуказанных параметров на основе численных результатов, полученных для различных зависимостей $f(k)$. Было выяснено, что четыре типа норм – L1 и C для $F(k)$, L1 для $f(k)$ и близкая к ним норма в виде дополнения до единицы общей площади $S$ графиков для двух плотностей – весьма точно определяют вариацию величин (3) и (4) при вариации плотностей.

Обозначим через $\delta f$ вариацию вероятностей $f(k)$, понимаемую в смысле определенной нормы, $\delta N(\rho)$ - соответствующую вариацию средней длины очереди при заданном значении нагрузки $\rho$, $\delta\tau(\rho)$ - вариацию среднего времени ожидания начала обслуживания заявки. Обработка данных показала (см. также [4]), что при вариации равномерного распределения, которое определялось на промежутке от 1 до максимальной длины группы, равной в рассматриваемом примере числу 8, имеет место связь между $\delta N(\rho)$ и $\delta f$, а также между $\delta\tau(\rho)$ и $\delta f$ с детерминацией 0,99:

$$\left|\frac{\delta N(\rho)}{N(\rho)}\right| = 0,2\left(1 - \ln\rho\right)\left(\frac{\delta f}{A}\right)^{0,27}; \\ \left|\frac{\delta\tau(\rho)}{\tau(\rho)}\right| = \frac{0,106 + 0,041\ln\rho}{\rho}\left(\frac{\delta f}{A}\right)^{0,27}; \ \rho \in [0,1; 0,9]. \tag{8}$$

Здесь коэффициент $A$ в зависимости от выбранной нормы равен:

$$A_S = A_{C-F} = 1,0; \ A_{L1-f} = 2,0; \ A_{L1-F} = 5,7.$$

## 3. Оценка вариации параметров модели обслуживания сообщений для эмпирического распределения

Применим формулу (8) для оценки неопределенности в позиционировании среднего значения длины очереди, если эмпирическое распределение групп заявок по длинам строится по выборке длины $n$ в момент времени $t$. Выборка отсчитывается назад от указанного момента времени. Эту, возможно, нестационарную выборочную вероятность того, что длина группы заявок равна $k \geq 1$, обозначим $f_n(k,t)$. Соответствующую функцию распределения обозначим $F_n(k,t)$.

Прежде всего, исследуем ряд на стационарность его выборочной функции распределения. Согласно методике, развитой в [5, 6], разбиваем весь имеющийся массив данных на попарно непересекающиеся выборки длины $n$. Максимальная длина $n_{\max}$ выбирается так, чтобы можно было построить не менее 10 непересекающихся пар выборок. По каждой выборке строятся распределения $F_n(k,t)$, после чего для пар выборок равных длин находятся расстояния между ними в норме C. Как известно (см., напр., [7]), для стационарных распределений имеет место сходимость по вероятности указанных расстояний к функции Колмогорова:

$$\lim_{n\to\infty} P\left\{0 < \sqrt{n/2}\left\|F_n(k,t_1) - F_n(k,t_2)\right\| < z\right\} = K(z).$$

Рассмотрим так называемый согласованный уровень стационарности $\varepsilon^*(n)$, определяемый условием: вероятность превышения расстояния между распределениями, равного $\varepsilon$, равна значимости используемого для этой цели критерия, т.е.

$$1-\varepsilon^* = K\left(\sqrt{n/2}\varepsilon^*\right). \tag{9}$$

Затем строим функцию распределения $G_n(z)$ попарных расстояний $z$ между выборками длины $n$ для конкретного временного ряда и численно решаем уравнение относительно эмпирического согласованного уровня стационарности $z_0(n)$:

$$G_n(z_0) = 1 - z_0. \tag{10}$$

Для каждой длины $n$ находится индекс нестационарности $J(n)$, равный отношению доли расстояний, не превосходящих $z_0(n)$ по имеющимся эмпирическим данным, к доле расстояний, не превосходящих стационарный уровень шума $\varepsilon^*(n)$:

$$J(n) = \frac{1 - z_0(n)}{1 - \varepsilon^*(n)}. \tag{11}$$

Ряд считается стационарным, если $J(n) \leq 1$, и нестационарным в противном случае. Если эмпирическое распределение $F_n(k)$ оказалось стационарным, то для определения средних значений параметров модели сервера поступаем следующим образом. Сначала строим функцию распределения $W_n(q)$ расстояний между эмпирическим выборочным распределением для определенной длины выборки $n$ и равномерным распределением в одной из норм, указанных в (8). Затем находим его медиану $q_{n,1/2}$ и содержащий ее интервал, определяемый квантилями $q_{n,1/2-\alpha}$ и $q_{n,1/2+\alpha}$. Подстановка величины $q_{n,1/2}$ в формулы (8) вместо $\delta f$ (в соответствующей норме) дает изменение средней длины очереди и среднего времени ожидания по сравнению с равномерным распределением для определенного значения нагрузки $\rho$. Доверительный интервал на уровне доверия $2\alpha$ определяется подстановкой в (8) квантилей $q_{n,1/2\pm\alpha}$. Если же распределение $F_n(k)$ на некоторых длинах $n$ нестационарно ($J_n > 1$), то та же процедура приводит к построению доверительного интервала, содержащего среднее значение параметра, на уровне доверия $2\alpha / J_n$.

Практическое применение описанной методики состоит в следующем. Возьмем в качестве опорного теоретического распределения $f(k)$ равномерное на промежутке $1 \leq k \leq M$. Для этого распределения формула (3) средней длины очереди преобразуется к виду

$$N_0 = \frac{v_2}{2v_1 b_1}\rho + \frac{M-1}{3}\frac{\rho}{1-\rho} + \frac{b_2}{2b_1^2}\frac{\rho^2}{1-\rho}, \quad (12)$$

рассматриваемому как нулевое приближение.

Предположим, что наблюдаемое эмпирическое распределение стационарно и имеет выборочную ФР $F_n(k)$. Тогда расстояние между этим распределением и равномерным может быть вычислено явно. Например, в норме C-F оно равно

$$\delta f = \sup_k \left| F_n(k) - k/M \right|. \quad (13)$$

Подставляя (13) в (8), получаем оценку средней длины очереди, отвечающей этому эмпирическому распределению:

$$N/N_0 \approx 1 + 0,2(1 - \ln\rho)(\delta f)^{0,27};\ \rho \in [0,1;\,0,9]. \quad (14)$$

Формула (14) вычислительно гораздо более проста, чем нахождение производящей функции по формуле (2), в которой $L(z)$ вычисляется по эмпирическому распределению $F_n(k)$, что, во-первых, технически весьма трудоемко и, во-вторых, приводит к вычислительным погрешностям, которые в относительных величинах могут превосходить неточность аппроксимации (14).

В частности, предположим, что эмпирическое распределение $f_n(k)$ взято из генеральной совокупности $f(k)$, имеющей на промежутке $1 \le k \le M$ логарифмическое распределение с параметром $q$, т.е.

$$f(k) = \frac{1}{Z}\frac{q^k}{k}, \quad Z = \sum_{k=1}^{M}\frac{q^k}{k}.$$

В работе [4] вычисления были проведены для случая $q = 0{,}85;\ M = 8$. По формуле (13) находим, что для этих параметров $\delta f = 0{,}43$. Следовательно, из (14) получаем, что средняя длина очереди для этого распределения будет превосходить аналогичную величину для равномерного распределения приблизительно на $0{,}16(1 - \ln\rho)N_0(\rho)$.

## Заключение

В работе предложен подход к оценке параметров модели сервера протокола установления сессий с групповым поступлением сообщений в зависимости от длины группы сообщений. Изучение именно этой зависимости стимулировано тем обстоятельством, что распределение длины группы не известно в виде генеральной совокупности, и, более того, не может быть известно, ибо эмпирические оценки такой совокупности нестационарны. Поэтому большое значение приобретают приближенные методы оценки параметров, не «привязанные» к определенному функциональному классу указанных распределений. Среди непараметрических методов в этой связи представляется продуктивным метод, использующий коэффициенты чувствительности параметров модели к изменению расстояний между функциями распределения. Возможность применения этого метода к нестационарным распределениям путем интерпретации нестационарного поведения как характерной вариации определенного базового распределения (например, равномерного) позволяет обойти формальную трудность, связанную с отсутствием теорем сходимости как по вероятности, так и по норме для рассматриваемых случайных величин.

Таким образом, изложенный подход к оценке вероятностно-временных характеристик модели сервера протокола установления сессий и проведенный анализ чувствительности характеристик позволяет дать рекомендацию инженерам о применении более простых формул для первичной оценки обслуживания сигнальных сообщений присутствия.

## Литература


1. Abhayawardhana V.S., Babbage R. A traffic model for the IP multimedia subsystem (IMS). // Proceedings of 65th vehicular technology conference. - 2007. – P. 783-787.
2. Самуйлов К.Е., Сопин Э.С. К анализу системы M[X]|G|1|r с прогулками прибора // Вестник РУДН. Серия: Математика. Информатика. Физика. – 2011. - №1. – С. 91-97.
3. Gaidamaka Yu., Pechinkin A., Razumchik R., Samouylov K., Sopin E. Analysis of M|G|1|R queue with batch arrivals and two hysteretic overload control policies // International Journal of Applied Mathematics and Computer Science, 2014, Vol. 24, No. 3, P. 519–534.
4. Гайдамака Ю.В., Зарипова Э.Р., Орлов Ю.Н. Анализ зависимости параметров модели сервера протокола установления сессий с групповым поступлением сообщений от распределения длины группы сообщений / Препринт ИПМ

им. М.В. Келдыша РАН, № 27, 2015. – 16 с. / URL: http://library.keldysh.ru/preprint.asp?id=2015-27
5. Орлов Ю.Н., Шагов Д.О. Индикативные статистики для нестационарных временных рядов / Препринт ИПМ им. М.В. Келдыша РАН, № 53, 2011. – 20 с. / URL: http://library.keldysh.ru/preprint.asp?id=2011-53
6. Орлов Ю.Н. Кинетические методы исследования нестационарных временных рядов. – М.: МФТИ, 2014. – 276 с.
7. Королюк В.С., Портенко Н.И., Скороход А.В., Турбин А.Ф. Справочник по теории вероятностей и математической статистике. – М.: Наука, 1985. – 640 с.

**Гайдамака Юлия Васильевна.** Доцент кафедры прикладной информатики и теории вероятностей Российского университета дружбы народов. Окончила РУДН в 1995 году. Кандидат физико-математических наук, доцент. Автор более 50 печатных работ. Область научных интересов: теория массового обслуживания, математическая теория телетрафика мультисервисных сетей, одноранговые сети связи (P2P-сети), сети 4G/5G с учетом трафика межмашинного взаимодействия (M2M, D2D), облачные вычисления (cloud computing). E-mail: ygaidamaka@sci.pfu.edu.ru

**Зарипова Эльвира Ринатовна.** Старший преподаватель кафедры прикладной информатики и теории вероятностей Российского университета дружбы народов. Окончила РУДН в 2003 году. Кандидат физико-математических наук. Автор 14 печатных работ. Область научных интересов: теория конечных графов, телекоммуникационные сети, анализ характеристик сети связи при внедрении межсетевых экранов. E-mail: ezarip@sci.pfu.edu.ru

**Орлов Юрий Николаевич**. Заведующий сектором кинетических уравнений Института прикладной математики им. М.В. Келдыша Российской академии наук. Окончил Московский физико-технический институт в 1987 г. Доктор физико-математических наук. Автор более 100 печатных работ, в том числе 7 монографий. Область научных интересов: релятивистская кинетическая теория, квантование динамических систем, прикладная математическая статистика, энергетика. E-mail: ov3159f@yandex.ru, yuno@kiam.ru