

УДК 004.725.5

<https://doi.org/10.31854/2307-1303-2026-14-2-53-69>

EDN: CBNOAJ

Исследование характеристик иерархической Edge-Fog-Cloud системы обработки IoT-трафика методом имитационного моделирования

 Комаров И. И. ,  Парамонов А. И.

Санкт-Петербургский государственный университет телекоммуникаций им. проф. М. А. Бонч-Бруевича,
Санкт-Петербург, 193232, Российская Федерация

Постановка задачи. Стремительный рост нагрузки на каналы связи и вычислительные узлы происходит из-за увеличения количества подключенных устройств и интенсивности информационного обмена в системах Интернета вещей. Централизованная передача всего объема данных в облачную инфраструктуру при ограниченной пропускной способности каналов и конечной емкости буферов может сопровождаться увеличением задержки доставки и вероятности потери сообщений. Распределение обработки между граничными, туманными и облачными узлами позволяет снизить объем трафика, передаваемого на верхние уровни сети, однако эффективность различных вариантов архитектуры зависит от интенсивности, структуры и вариативности входного потока. **Целью работы** является оценка влияния распределения обработки гетерогенного IoT-трафика между граничным, туманным и облачным уровнями на среднюю задержку доставки и вероятность потери сообщений. **Используемые методы.** Решение поставленной задачи основано на дискретно-событийном имитационном моделировании в среде AnyLogic с использованием элементов библиотеки моделирования процессов. Исследуются три варианта построения системы: централизованная облачная обработка, совместное использование граничных и облачных вычислений и трехуровневая архитектура Edge-Fog-Cloud. В модели учитываются три класса сообщений, конечная емкость буферов, производительность вычислительных узлов и каналов связи, агрегация части обработанного трафика, а также вариативность процессов поступления и обслуживания сообщений. Элементами **новизны** представленной работы являются имитационная реализация многоуровневой архитектуры обработки гетерогенного IoT-трафика с раздельным учетом граничного, туманного и облачного уровней, а также исследование влияния вариативности процессов поступления и обслуживания на задержку доставки сообщений при изменении интенсивности входного потока. **Результат.** Разработана имитационная модель, позволяющая выполнять параметрическое исследование трех вариантов архитектуры при изменении интенсивности входного потока от 20 до 400 сообщений в секунду. Получены зависимости вероятности потери сообщений и средней задержки доставки от интенсивности трафика для трех сценариев обработки. Для многоуровневой архитектуры дополнительно получены зависимости средней задержки от степени вариативности процессов поступления и обслуживания. **Теоретическая / Практическая значимость:** разработанная модель позволяет проводить сравнительную оценку вариантов построения распределенных IoT-систем, выявлять участки возникновения перегрузок и оценивать влияние характеристик трафика, вычислительных узлов, каналов связи и конечных буферов на качество обслуживания сообщений.

Ключевые слова: Интернет вещей, граничные вычисления, туманные вычисления, системы массового обслуживания, задержка, потеря сообщения

Библиографическая ссылка на статью:

Комаров И. И., Парамонов А. И. Исследование характеристик иерархической Edge-Fog-Cloud системы обработки IoT-трафика методом имитационного моделирования // Информационные технологии и телекоммуникации. 2026. Т. 14. № 2. С. 53–69. DOI: 10.31854/2307-1303-2026-14-2-53-69. EDN: CBNOAJ

Reference for citation:

Komarov I. I., Paramonov A. I. Research of the Characteristics of the Hierarchical Edge-Fog-Cloud System for Processing IoT Traffic Using Simulation Modeling // Telecom IT. 2026. Vol. 14. Iss. 2. PP. 53–69. (in Russian). DOI: 10.31854/2307-1303-2026-14-2-53-69. EDN: CBNOAJ

Актуальность

Развитие технологий Интернета вещей (IoT, *аббр. от англ.* Internet of Things) сопровождается увеличением количества подключенных устройств и объема формируемого ими информационного трафика. В распределенных системах мониторинга, промышленной автоматизации и управления большое число конечных устройств непрерывно формирует сообщения, различающиеся по назначению, требованиям к оперативности обработки и допустимому времени доставки. При увеличении интенсивности информационного обмена возрастают требования не только к вычислительным ресурсам системы, но и к пропускной способности каналов связи и емкости промежуточных буферов [1].

Традиционный подход к организации обработки данных Интернета вещей предполагает передачу основной части поступающих сообщений в облачный центр обработки данных. Централизация вычислительных ресурсов упрощает организацию обработки и хранения информации, однако приводит к необходимости передачи значительного объема трафика через сетевую инфраструктуру [2]. При увеличении интенсивности входного потока ограниченная производительность отдельных элементов тракта передачи и обработки может приводить к накоплению сообщений в очередях. Следствием этого является увеличение времени доставки данных, а при заполнении конечных буферов – потеря части поступающих сообщений.

Одним из способов снижения нагрузки на облачную инфраструктуру является перенос части вычислительных операций ближе к источникам данных. В рассматриваемой иерархии под термином Edge понимается граничный уровень обработки данных непосредственно вблизи источников их формирования, под Fog – промежуточный уровень распределенной обработки между граничными узлами и облачной инфраструктурой, а под Cloud – централизованный облачный уровень обработки и хранения данных. В архитектуре граничных вычислений обработка выполняется непосредственно на Edge-узлах, что позволяет сократить объем информации, передаваемой в облачный центр [3]. Дальнейшим развитием данного подхода является использование промежуточного Fog-уровня, объединяющего группы граничных узлов и обеспечивающего дополнительную обработку и агрегацию поступающей информации перед ее передачей в облачную инфраструктуру [4].

Таким образом, возможны различные варианты организации вычислительного процесса: от централизованной облачной обработки до многоуровневого распределения вычислений между Edge-, Fog- и Cloud-узлами [5]. При этом введение дополнительных уровней само по себе не гарантирует улучшения характеристик системы. Результат определяется соотношением интенсивности входного потока, производительности вычислительных узлов, пропускной способности каналов, емкости буферов и правил распределения сообщений между уровнями.

Дополнительное влияние на функционирование системы оказывает неоднородность информационного потока. Сообщения разных классов могут иметь различные требования к способу обработки и маршруту прохождения через систему [6]. Кроме того, реальные процессы поступления и обслуживания сообщений характеризуются вариабельностью временных интервалов. Поэтому

оценка архитектуры только по средним значениям интенсивности поступления и производительности обслуживающих устройств не позволяет в полной мере определить ее поведение при изменении характера нагрузки.

Для исследования подобных систем целесообразно применение имитационного моделирования, позволяющего воспроизводить процессы поступления, маршрутизации, ожидания, обработки и передачи отдельных сообщений во времени [7]. Имитационная модель дает возможность исследовать работу одной и той же системы при различных значениях входной нагрузки и параметров ее элементов, а также непосредственно оценивать вероятность потери сообщений и среднее время их доставки.

В связи с этим актуальной является задача сравнительного исследования вариантов организации обработки гетерогенного IoT-трафика в иерархической Edge-Fog-Cloud системе с учетом ограниченной емкости буферов, конечной производительности узлов и каналов связи, а также вариативности процессов поступления и обслуживания сообщений.

В работе для решения данной задачи разрабатывается дискретно-событийная имитационная модель, и исследуются три варианта организации системы: централизованная облачная обработка, двухуровневая обработка с использованием Edge- и Cloud-узлов и иерархическая обработка с использованием Edge-, Fog- и Cloud-уровней. В качестве основных показателей эффективности рассматриваются вероятность потери сообщений и средняя задержка их доставки.

Постановка задачи

Для формального описания исследуемой системы используются обозначения, представленные в таблице 1.

Таблица 1. Принятые обозначения

Обозначение	Содержание
λ	интенсивность общего входного потока сообщений
$\alpha A, \alpha B, \alpha C$	доли сообщений классов A, B и C
ME	количество граничных вычислительных узлов
MF	количество туманных вычислительных узлов
G	количество Edge-узлов, обслуживаемых одним Fog-узлом
μE	интенсивность обслуживания Edge-узла
μF	интенсивность обслуживания Fog-узла
μC	интенсивность обслуживания Cloud-узла
μ_{route}	интенсивность маршрутизации
K	емкость соответствующего буфера
M	количество сообщений, используемых при агрегации
C_a	коэффициент вариации интервалов поступления сообщений
C_s	коэффициент вариации времени обслуживания
P_{loss}	вероятность потери сообщений
T_{avg}	средняя задержка доставки сообщений
s	номер исследуемого сценария

На вербальном уровне задача может быть декомпозирована в виде:

- построение имитационных моделей трех вариантов организации обработки IoT-трафика: Cloud-Only, Edge-Cloud и Edge-Fog-Cloud;
- определение средней задержки доставки и вероятности потери сообщений при изменении интенсивности входного потока;
- сравнительная оценка характеристик рассматриваемых вариантов архитектуры;
- исследование влияния вариабельности процессов поступления и обслуживания сообщений на характеристики трехуровневой Edge-Fog-Cloud архитектуры.

На формальном уровне постановка задачи исследования имеет следующий вид.

Дано: входной поток сообщений интенсивностью λ , разделенный на три класса A , B и C с долями $\alpha A = 0,1$; $\alpha B = 0,3$; $\alpha C = 0,6$; множество из $ME = 6$ граничных и $MF = 2$ туманных вычислительных узлов; заданные характеристики производительности вычислительных узлов, каналов связи и емкости буферов; три варианта организации обработки трафика. Интенсивность входного потока при проведении имитационного эксперимента изменяется от 20 до 400 сообщ./с с шагом 20 сообщ./с. Для оценки вариантов архитектуры используются средняя задержка доставки T_{avg} и вероятность потери сообщений P_{loss} . Дополнительно для трехуровневой архитектуры исследуется влияние коэффициентов вариации C_a и C_s при условии $C_a = C_s$.

Имитационная модель системы обработки IoT-трафика

Для проведения вычислительных экспериментов разработана дискретно-событийная имитационная модель иерархической системы обработки IoT-трафика в среде AnyLogic. Модель предназначена для исследования процессов поступления, маршрутизации, передачи и обработки сообщений, а также для регистрации задержек и потерь, возникающих при ограниченной производительности элементов системы [8].

Генерация сообщений для всех исследуемых сценариев осуществляется общей входной подсистемой. Сформированному сообщению присваивается один из трех классов: A , B или C . Выбор класса выполняется в соответствии с заданными долями трафика $\alpha A = 0,1$; $\alpha B = 0,3$; $\alpha C = 0,6$. После классификации сообщение направляется в одну из трех топологий в соответствии с номером текущего сценария. Такой способ построения позволяет использовать единый входной поток при последовательном исследовании различных вариантов организации обработки.

Для реализации модели использованы элементы библиотеки моделирования процессов AnyLogic. Формирование входного потока выполняется блоком Source, а элементы SelectOutput используются для классификации сообщений, выбора сценария и дальнейшей маршрутизации. Блоки Queue моделируют конечные буферы каналов и вычислительных узлов [9]. С помощью блоков Delay воспроизводятся задержки передачи и процессы обработки сообщений. Блоки

Sink используются для завершения движения сообщений по соответствующему маршруту и регистрации результатов моделирования.

В связи со значительным размером разработанной модели ее структура далее представлена отдельно для каждого исследуемого сценария.

Сценарий Cloud-Only

Первый сценарий соответствует централизованной организации обработки IoT-трафика. После формирования и классификации сообщения направляются в ветвь Cloud-Only, структура которой представлена на рисунке 1.

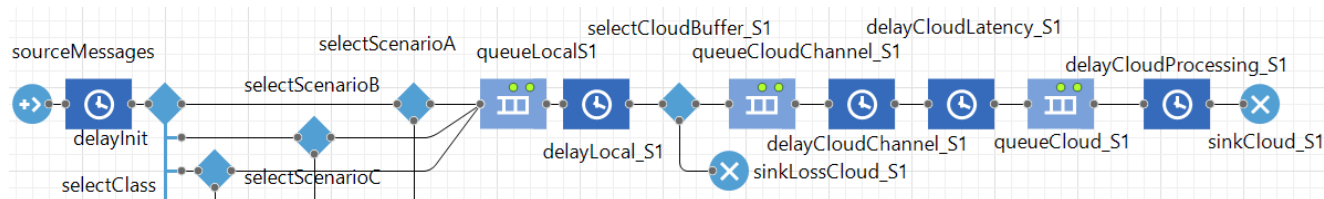


Рис. 1. Имитационная модель сценария Cloud-Only

В начале маршрута учитывается локальная составляющая задержки, после чего сообщение поступает к подсистеме облачного канала. Перед передачей выполняется проверка доступности его буфера. Если суммарная занятость элементов канала не превышает установленную емкость, сообщение принимается для дальнейшей передачи. При отсутствии свободной емкости оно направляется в блок регистрации потерь.

Принятые сообщения последовательно проходят очередь канала, моделирование времени передачи и далее поступают в облачную подсистему. На облачном уровне также учитываются ожидание в очереди и время обработки. После успешного прохождения маршрута сообщение поступает в конечный блок, где регистрируется факт успешной доставки и определяется время нахождения сообщения в системе.

Таким образом, первый сценарий используется в качестве базового варианта, в котором отсутствует распределение вычислительной нагрузки между промежуточными уровнями.

Сценарий Edge-Cloud

Во втором сценарии в систему вводится уровень граничных вычислений. В отличие от централизованного варианта, поступающий поток распределяется между шестью параллельными Edge-ветвями. Фрагмент имитационной модели данного сценария представлен на рисунке 2 (общая подсистема формирования и классификации входного потока повторно не приводится).

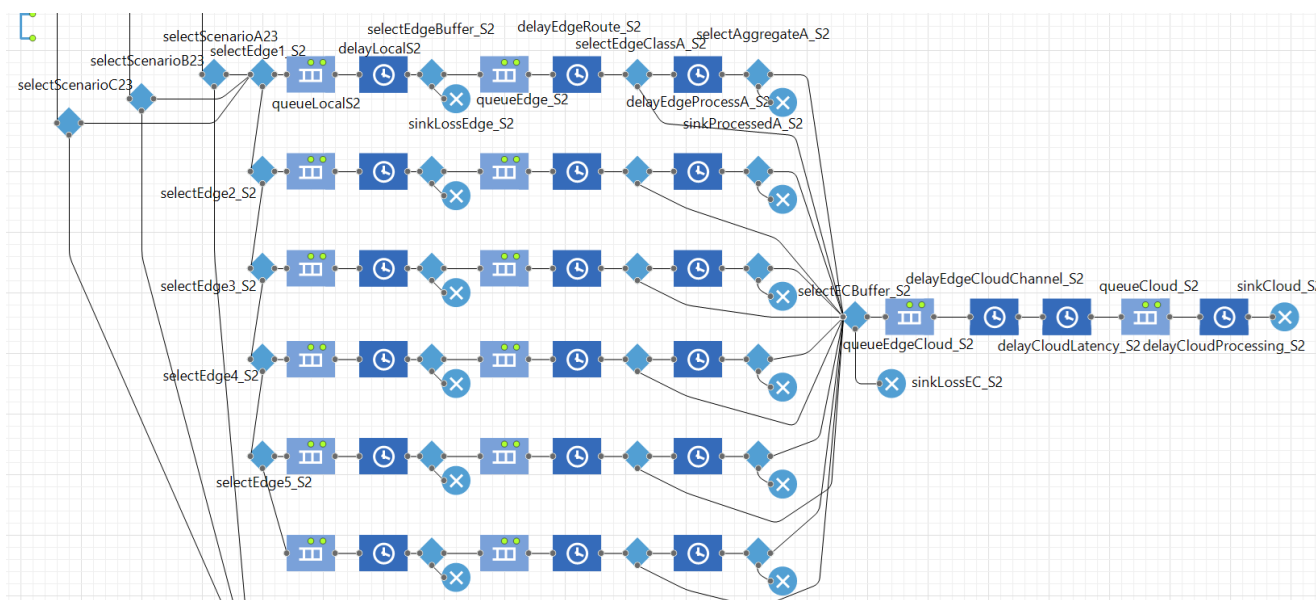


Рис. 2. Имитационная модель сценария Edge-Cloud

Каждая из шести ветвей содержит собственные элементы очереди и обработки, что позволяет моделировать независимую загрузку отдельных Edge-узлов. Перед поступлением сообщения в соответствующую очередь производится проверка наличия свободной емкости. При переполнении буфера сообщение регистрируется как потерянное на граничном уровне.

После прохождения локальной обработки определяется дальнейший маршрут сообщения. Сообщения, для которых обработка на Edge-уровне является достаточной, учитываются как успешно обработанные. Сообщения, требующие передачи результатов на верхний уровень, направляются к подсистеме Edge-Cloud. Потоки от шести граничных ветвей далее поступают к облачному каналу, после чего проходят соответствующие этапы передачи, ожидания и обработки в облачной инфраструктуре.

Использование шести параллельных ветвей позволяет распределить входную нагрузку между несколькими граничными узлами и исследовать влияние такого распределения на задержку и вероятность потерь по сравнению с централизованной обработкой.

Сценарий Edge-Fog-Cloud

Третий сценарий представляет трехуровневую архитектуру, включающую граничный, туманный и облачный уровни обработки. Как и во втором сценарии, входной поток распределяется между шестью параллельными Edge-ветвями. После выполнения локальной обработки дальнейший маршрут сообщения определяется его классом и необходимостью передачи на последующие уровни системы. Структура имитационной модели данного сценария представлена на рисунке 3.

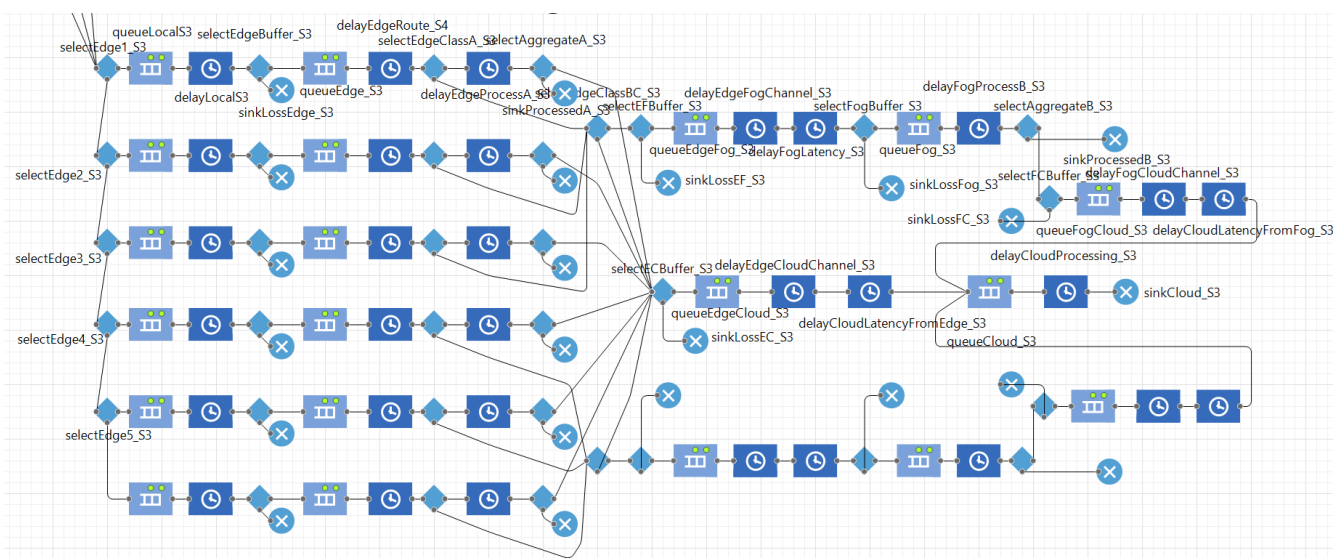


Рис. 3. Имитационная модель сценария Edge-Fog-Cloud

Граничный уровень включает шесть Edge-узлов, каждый из которых содержит собственные элементы буферизации и обработки. При отсутствии свободной емкости соответствующего буфера сообщение регистрируется как потерянное. После обработки сообщения разделяются в зависимости от класса и направляются по соответствующему маршруту.

Шесть Edge-узлов объединены в две группы по три узла, каждая из которых связана с соответствующим Fog-узлом. На туманном уровне моделируются передача сообщений от Edge-узлов, ожидание в конечных буферах и последующая обработка. Для части трафика предусмотрена агрегация: после обработки группы из $m = 10$ сообщений формируется одно агрегированное сообщение для передачи на облачный уровень [10].

Сообщения, требующие дальнейшей обработки, поступают через соответствующие каналы в облачную подсистему, где учитываются задержка передачи, ожидание в очереди и время обработки. После завершения маршрута сообщения направляются в блоки регистрации успешно обработанного трафика.

Для каждого из трех сценариев ведется отдельный учет количества поступивших, успешно обработанных и потерянных сообщений классов A, B и C. Для успешно завершивших обработку сообщений определяется полное время нахождения в системе. На основании накопленной статистики рассчитываются основные показатели исследования – вероятность потери сообщений и средняя задержка доставки.

Разработанная имитационная модель обеспечивает исследование трех вариантов организации вычислений при одинаковых параметрах входного трафика и позволяет сопоставить их характеристики при последовательном увеличении нагрузки. Дополнительно для сценария Edge-Fog-Cloud предусмотрено изменение вариативности процессов поступления и обслуживания сообщений, что позволяет оценить ее влияние на среднюю задержку доставки.

Организация вычислительного эксперимента

Вычислительный эксперимент проводился на разработанной имитационной модели для трех вариантов организации обработки IoT-трафика: Cloud-Only, Edge-Cloud и Edge-Fog-Cloud. Для обеспечения сопоставимости результатов во всех сценариях использовался единый входной поток и одинаковое соотношение сообщений различных классов.

Основным варьируемым параметром являлась интенсивность входного потока λ . Ее значение последовательно изменялось от 20 до 400 сообщ./с с шагом 20 сообщ./с. Для каждого значения интенсивности выполнялось моделирование трех исследуемых сценариев. Основные параметры модели, используемые при проведении вычислительного эксперимента, представлены в таблице 2.

Таблица 2. Параметры вычислительного эксперимента

Параметр	Значение
Диапазон интенсивности входного потока, сообщ./с	20–400
Шаг изменения, сообщ./с	20
Доля сообщений класса A, %	10
Доля сообщений класса B, %	30
Доля сообщений класса C, %	60
Количество Edge-узлов	6
Количество Fog-узлов	2
Количество Edge-узлов на один Fog-узел	3
Интенсивность обработки на Edge-узле, сообщ./с	130
Интенсивность маршрутизации и распознавания класса трафика, сообщ./с	900
Интенсивность обработки на Fog-узле, сообщ./с	400
Интенсивность обработки на Cloud-уровне, сообщ./с	750
Коэффициент агрегации сообщений	10

Параметр μ_{route} характеризует интенсивность выполнения операций распознавания класса поступившего сообщения и выбора его дальнейшего маршрута. Он не определяет производительность каналов связи. Процессы передачи сообщений между уровнями Edge, Fog и Cloud моделируются отдельно с учетом характеристик соответствующих каналов и задаваемых для них задержек передачи [11].

В ходе каждой итерации модели отдельно учитывалось количество поступивших сообщений классов A, B и C, количество сообщений, успешно завершивших обработку, а также количество сообщений, потерянных вследствие заполнения конечных буферов. Исходя из накопленных данных определялась вероятность потери сообщений. Для успешно обработанных сообщений дополнительно фиксировалось полное время нахождения в системе, на основании которого рассчитывалась средняя задержка доставки [12].

Основной вычислительный эксперимент предназначен для сравнительной оценки трех вариантов организации обработки трафика. Для каждого значения λ определялись суммарная вероятность потери сообщений и средняя задержка

доставки. Полученные значения использовались для построения зависимостей исследуемых показателей от интенсивности входного потока и последующего сопоставления сценариев Cloud-Only, Edge-Cloud и Edge-Fog-Cloud.

Дополнительно проведен эксперимент по оценке влияния вариабельности процессов поступления и обслуживания сообщений на трехуровневую архитектуру Edge-Fog-Cloud. Для этого совместно изменялись коэффициент вариации интервалов между поступлениями сообщений C_a и коэффициент вариации времени их обслуживания C_s . Исследование проводилось при условии $C_a = C_s$ для трех значений коэффициента вариации: 0,5; 1,0 и 2,0. Для каждого значения интенсивность входного потока изменялась от 20 до 400 сообщ./с. В качестве исследуемого показателя использовалась средняя задержка доставки сообщений.

Проведение двух вычислительных экспериментов позволяет, с одной стороны, сопоставить характеристики трех вариантов организации обработки при увеличении интенсивности IoT-трафика, а с другой – оценить влияние степени вариабельности процессов поступления и обслуживания на задержку в трехуровневой архитектуре Edge-Fog-Cloud.

Результаты имитационного моделирования

В результате вычислительного эксперимента получены зависимости вероятности потери сообщений и средней задержки доставки от интенсивности входного потока для трех вариантов организации обработки IoT-трафика. Сравнение выполнялось для сценариев Cloud-Only, Edge-Cloud и Edge-Fog-Cloud.

При увеличении интенсивности входного потока вероятность потери сообщений возрастает во всех исследуемых сценариях, однако характер роста существенно зависит от организации обработки. Наиболее выраженное увеличение потерь наблюдается при централизованной обработке Cloud-Only. При интенсивности 300 сообщ./с вероятность потери составляет 1,320 %, а при увеличении интенсивности до 400 сообщ./с она возрастает до 5,526 %.

Использование граничных вычислений позволяет уменьшить потери. Для сценария Edge-Cloud при интенсивности 300 сообщ./с вероятность потери составляет 0,435 %, а при 400 сообщ./с – 3,236 %. Таким образом, при высокой нагрузке распределение части обработки между Edge-узлами снижает вероятность потери по сравнению с централизованной архитектурой.

Наименьшая вероятность потери получена для трехуровневой архитектуры Edge-Fog-Cloud. При интенсивности 300 сообщ./с она составляет 0,037 %, а при 400 сообщ./с – 0,112 %. Следовательно, даже при максимальной исследуемой интенсивности вероятность потери в данном сценарии остается близкой к нулю.

На рисунке 4 показаны зависимости вероятности потери сообщений и средней задержки доставки от интенсивности входного потока. Для наглядного сравнения результаты при интенсивностях 300 и 400 сообщ./с представлены в таблице 3.

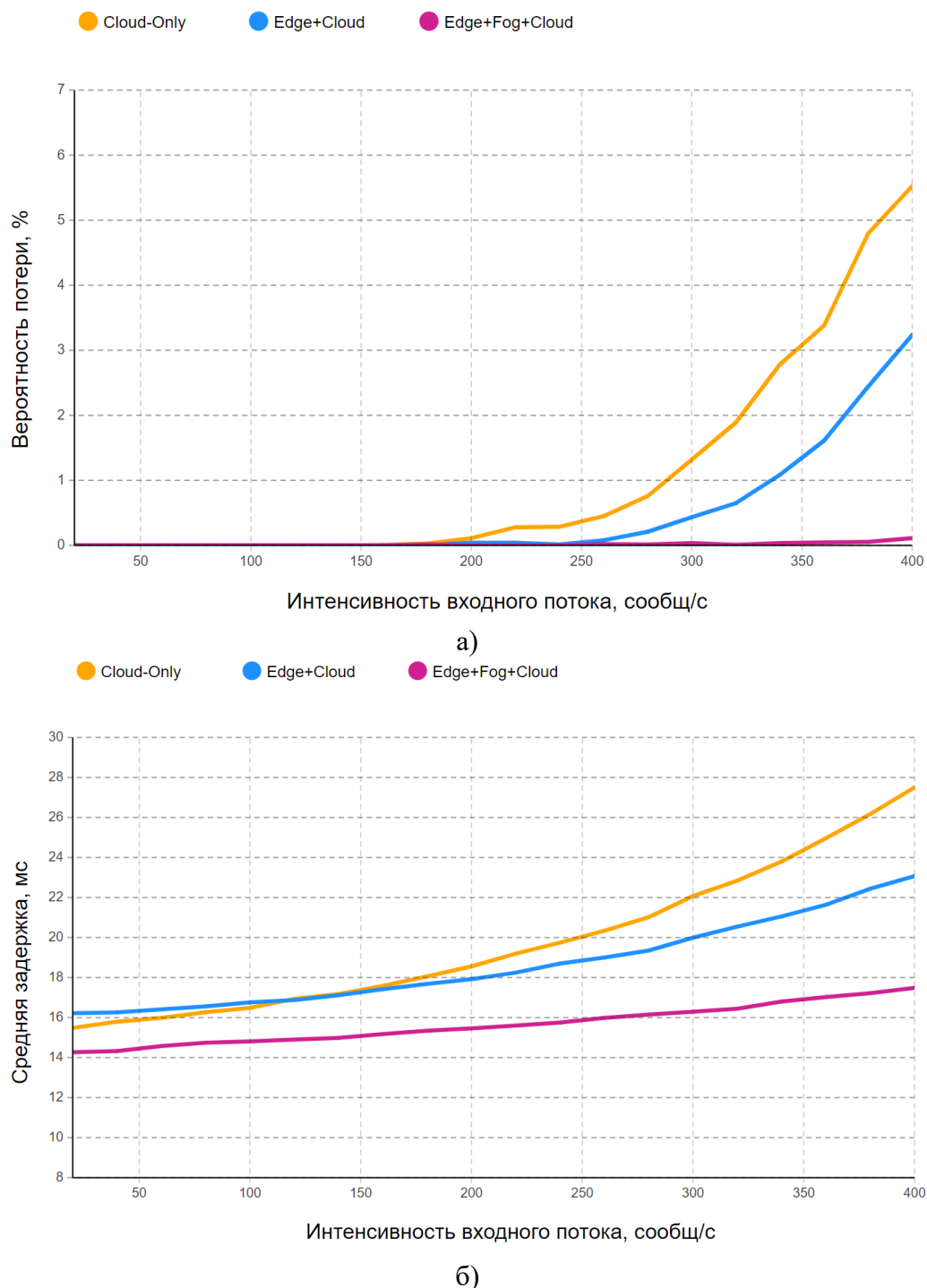


Рис. 4. Зависимость вероятности потери сообщений (а) и средней задержки доставки (б) от интенсивности входного потока

Таблица 3. Результаты моделирования при высокой интенсивности потока

λ, сообщ./с	Сценарий	Вероятность потери, %	Средняя задержка, мс
300	Cloud-Only	1,320	22,071
	Edge-Cloud	0,435	19,991
	Edge-Fog-Cloud	0,037	16,293
400	Cloud-Only	5,526	27,053
	Edge-Cloud	3,236	23,072
	Edge-Fog-Cloud	0,112	17,488

Результаты моделирования показывают аналогичное преимущество распределенной обработки по показателю задержки. При $\lambda = 300$ сообщ./с средняя задержка в сценарии Cloud-Only составляет 22,071 мс, при использовании Edge-Cloud она снижается до 19,991 мс, а в трехуровневой архитектуре – до 16,293 мс.

При увеличении интенсивности до 400 сообщ./с различия становятся более выраженными. Для централизованного сценария средняя задержка составляет 27,503 мс, для Edge-Cloud – 23,072 мс, для Edge-Fog-Cloud – 17,488 мс. Таким образом, при $\lambda = 400$ сообщ./с переход от Cloud-Only к Edge-Fog-Cloud обеспечивает снижение средней задержки примерно на 36,4 %, а вероятности потери – примерно на 98,0 %.

Дополнительный вычислительный эксперимент позволил оценить влияние вариабельности процессов поступления и обслуживания на среднюю задержку в сценарии Edge-Fog-Cloud. Коэффициенты вариации изменялись совместно при условии $C_a = C_s$ и принимали значения 0,5; 1,0; 2,0.

Результаты показывают, что увеличение вариабельности приводит к росту средней задержки даже при неизменной средней интенсивности входного потока [13]. При $\lambda = 300$ сообщ./с средняя задержка составляет 14,824 мс при $C_a = C_s = 0,5$; 16,293 мс при $C_a = C_s = 1$ и 21,687 мс при $C_a = C_s = 2$.

При $\lambda = 400$ сообщ./с соответствующие значения составляют 15,349; 17,488 и 24,085 мс. Таким образом, увеличение коэффициента вариации с 0,5 до 2 при максимальной исследуемой интенсивности приводит к росту средней задержки примерно на 56,9 %. Полученные зависимости представлены на рисунке 5 и в таблице 4.

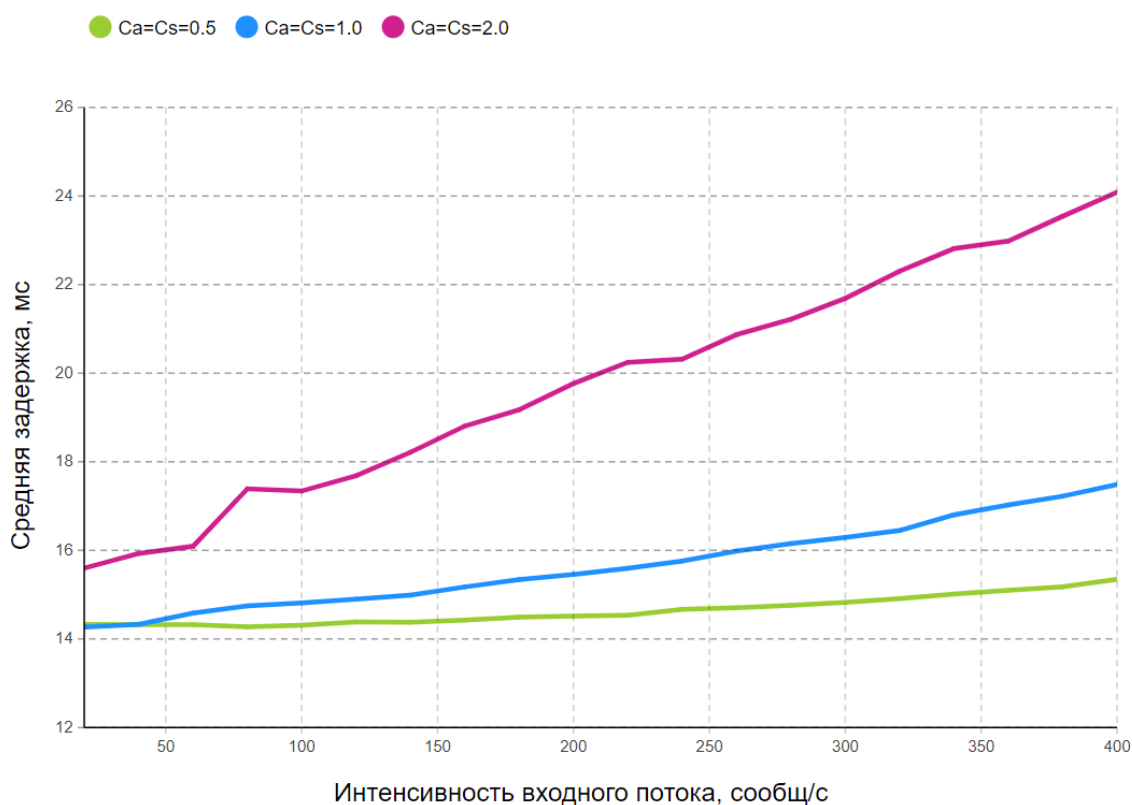


Рис. 5. Влияние коэффициентов вариации процессов поступления и обслуживания на среднюю задержку для архитектуры Edge-Fog-Cloud

Таблица 4. Результаты моделирования при высокой интенсивности потока

$C_a = C_s$	Средняя задержка	
	при $\lambda = 300$, мс	при $\lambda = 400$, мс
0,5	14,824	15,349
1,0	16,293	17,488
2,0	21,687	24,085

Полученные результаты показывают, что при оценке производительности распределенной IoT-системы необходимо учитывать не только среднюю интенсивность входного потока, но и его вариабельность. Более неравномерное поступление и обслуживание сообщений способствует формированию кратковременных перегрузок и увеличению срока ожидания в очередях, что приводит к росту полной задержки [14–15].

В целом результаты вычислительного эксперимента показывают преимущество трехуровневой архитектуры Edge-Fog-Cloud в рассматриваемом диапазоне нагрузки. Распределение обработки между граничным, туманным и облачным уровнями, а также агрегация части трафика позволяют снизить нагрузку на последующие элементы системы и обеспечить меньшие значения вероятности потери и средней задержки по сравнению с Cloud-Only и Edge-Cloud.

Выводы

В работе проведено имитационное исследование трех вариантов организации обработки IoT-трафика: централизованной архитектуры Cloud-Only, двухуровневой Edge-Cloud и трехуровневой Edge-Fog-Cloud. Разработанная в среде AnyLogic модель позволила воспроизвести процессы формирования, маршрутизации, буферизации, передачи и обработки сообщений с учетом ограниченной емкости очередей и распределения трафика между вычислительными уровнями.

Результаты моделирования показали, что различия между исследуемыми архитектурами становятся наиболее выраженными при увеличении интенсивности входного потока. При интенсивности 400 сообщ./с для сценария Cloud-Only получены вероятность потери 5,526 % и средняя задержка 27,503 мс. Применение архитектуры Edge-Cloud позволило снизить данные показатели до 3,236 % и 23,072 мс соответственно. Наименьшие значения получены для трехуровневой архитектуры Edge-Fog-Cloud: вероятность потери составила 0,112 %, а средняя задержка – 17,488 мс.

При максимальной исследуемой интенсивности переход от централизованной обработки к архитектуре Edge-Fog-Cloud обеспечил снижение средней задержки на 36,4 %, а вероятности потери сообщений – примерно на 98,0 %. Полученный результат обусловлен распределением обработки между несколькими вычислительными уровнями и сокращением объема трафика, передаваемого на последующие уровни, за счет агрегации сообщений.

Дополнительное исследование показало существенное влияние вариабельности процессов поступления и обслуживания на характеристики системы. При интенсивности 400 сообщ./с увеличение коэффициентов вариации $C_a = C_s$ с 0,5

до 2,0 привело к росту средней задержки с 15,349 до 24,085 мс, т. е. приблизительно на 56,9 %. Следовательно, оценка распределенных систем обработки IoT-трафика только по средним значениям интенсивности поступления и обслуживания может быть недостаточной: увеличение неравномерности процессов способно заметно ухудшать временные характеристики системы.

Таким образом, в рассматриваемом диапазоне параметров применение промежуточных Edge- и Fog-уровней позволяет повысить устойчивость системы к увеличению входной нагрузки и уменьшить задержку обработки по сравнению с централизованной архитектурой. Полученные результаты могут быть использованы при выборе структуры многоуровневых систем обработки IoT-трафика и определении параметров их вычислительных узлов с учетом интенсивности и вариабельности поступающего потока.

Литература

1. Кучерявый А. Е., Окунева Д. В., Парамонов А. И., Хоанг Ф. Н. Методы распределения трафика в гетерогенной сети Интернета вещей высокой плотности // Труды учебных заведений связи. 2024. Т. 10. № 2. С. 67–74. DOI: 10.31854/1813-324X-2024-10-2-67-74. EDN: RTNVEU
2. Волков А. Н., Мутханна А. С. А., Кучерявый А. Е., Бородин А. С., Парамонов А. И. и др. Перспективные исследования сетей и услуг 2030 в лаборатории 6G MEGANETLAB СПбГУТ // Электросвязь. 2023. № 6. С. 5–14. DOI: 10.34832/ELSV.2023.43.6.001. EDN: CJSYLS
3. Мутханна А. С. А. Модель интеграции граничных вычислений в структуру сети «воздух – земля» и метод выгрузки трафика для сетей Интернета Вещей высокой и сверхвысокой плотности // Труды учебных заведений связи. 2023. Т. 9. № 3. С. 42–59. DOI: 10.31854/1813-324X-2023-9-3-42-59. EDN: SBAHAR
4. Профрансов Д. Ю., Сафонова И. Е. К вопросу о туманных вычислениях и Интернете вещей // Образовательные ресурсы и технологии. 2017. № 4 (21). С. 30–39. DOI: 10.21777/2500-2112-2017-4-30-39. EDN: YMQGJD
5. Глушак Е. В., Ключев Д. С. Разработка и исследование моделей функционирования облачных и туманных вычислений // Радиотехника. 2025. Т. 89. № 3. С. 157–168. DOI: 10.18127/j00338486-202503-14. EDN: IGUDLR
6. Аль-Кереа З. А. Х., Мутханна А. С. А., Кучерявый А. Е. Интеллектуальная бессерверная вычислительная система для услуг телеприсутствия // Труды учебных заведений связи. 2025. Т. 11. № 4. С. 18–27. DOI: 10.31854/1813-324X-2025-11-4-18-27. EDN: UOYTAY
7. Бабаева Б. Э., Лисовская Е. Ю. Модель сети массового обслуживания с тремя узлами для расчета требуемой вычислительной мощности сети Интернета вещей // Информационные технологии и математическое моделирование (ИТММ-2024): материалы XXIII Международной конференции имени А. Ф. Терпугова. Томск: Национальный исследовательский университет, 2024. С. 91–97. EDN: AZBPAK
8. Клейнрок Л. Теория массового обслуживания. М.: Машиностроение, 1979. 432 с.

9. Iversen V. B. Teletraffic Engineering and Network Planning. DTU Fotonik. 2015. 398 p.
10. Pourghebleh B., Navimipour N. J. Data Aggregation Mechanisms in the Internet of Things: A Systematic Review of the Literature and Recommendations for Future Research // Journal of Network and Computer Applications. 2017. Vol. 97. PP. 23–24. DOI: 10.1016/j.jnca.2017.08.006
11. Зелигер Н. Б., Чугреев О. С., Яновский Г. Г. Проектирование сетей и систем передачи дискретных сообщений: учебное пособие для вузов. М.: Радио и связь, 1984. 176 с.
12. Шнепс-Шнеппе М. А. Системы распределения информации. Методы расчета: справочное пособие. М.: Связь, 1979. 344 с.
13. Kingman J. F. C. The Single Server Queue in Heavy Traffic // Mathematical Proceedings of the Cambridge Philosophical Society. 1961. Vol 57. Iss. 4. PP. 902–904. DOI: 10.1017/S0305004100036094
14. Kingman J. F. C. On Queues in Heavy Traffic // Journal of the Royal Statistical Society. Series B (Methodological). 1962. Vol. 24. Iss. 2. PP. 383–392. DOI: 10.1111/j.2517-6161.1962.tb00465.x
15. Шнепс-Шнеппе М. А. Численные методы теории телетрафика. М.: Связь, 1974. 232 с.

Статья поступила 14 июля 2026 г.
Одобрена после рецензирования 24 августа 2026 г.
Принята к публикации 21 сентября 2026 г.

Информация об авторах

Комаров Иван Ильич – аспирант кафедры сетей связи и передачи данных Санкт-Петербургского государственного университета телекоммуникаций им. проф. М. А. Бонч-Бруевича. E-mail: komarov1.ii@sut.ru

Парамонов Александр Иванович – доктор технических наук, профессор, профессор кафедры сетей связи и передачи данных Санкт-Петербургского государственного университета телекоммуникаций им. проф. М. А. Бонч-Бруевича. E-mail: paramonov@sut.ru

<https://doi.org/10.31854/2307-1303-2026-14-2-53-69>
EDN: CBNOAJ

Research of the Characteristics of the Hierarchical Edge-Fog-Cloud System for Processing IoT Traffic Using Simulation Modeling

 I. Komarov ,  A. Paramonov 

The Bonch-Bruевич Saint Petersburg State University of Telecommunications,
St. Petersburg, 193232, Russian Federation

Purpose. The increasing number of connected devices and the intensity of information exchange in Internet of Things systems leads to an increased load on communication channels and computing nodes. Centralized transmission of the entire data volume to the cloud infrastructure with limited channel bandwidth and finite buffer capacity may be accompanied by an increase in delivery latency and the probability of message loss. Distributing processing between edge, fog, and cloud nodes allows for a reduction in the volume of traffic transmitted to the upper network layers; however, the effectiveness of various architecture options depends on the intensity, structure, and variability of the input flow. The aim of this work is to evaluate the impact of distributing the processing of heterogeneous IoT traffic between the edge, fog, and cloud layers on the average delivery latency and the probability of message loss. **Methods.** The solution to the problem is based on discrete-event simulation in the AnyLogic environment using elements of the process modeling library. Three system design options are investigated: centralized cloud processing, joint use of edge and cloud computing, and a three-tier Edge–Fog–Cloud architecture. The model takes into account three message classes, finite buffer capacity, performance of computing nodes and communication channels, aggregation of a portion of the processed traffic, as well as the variability of message arrival and servicing processes. **Novelty.** The novelty of the presented study is the simulation implementation of a multi-tier architecture for processing heterogeneous IoT traffic with separate consideration of the edge, fog, and cloud layers, as well as a study of the impact of variability of arrival and servicing processes on message delivery delay with changes in the input flow intensity. **Results.** A simulation model has been developed that enables a parametric study of three architecture options with changes in the input flow intensity from 20 to 400 messages per second. The dependences of the probability of message loss and the average delivery delay on the traffic intensity for three processing scenarios are obtained. For a multi-tier architecture, dependencies of average latency on the degree of variability of the arrival and service processes were additionally obtained. **Practical relevance.** The developed model enables a comparative evaluation of distributed IoT system design options, identification of congestion zones, and assessment of the impact of traffic characteristics, computing nodes, communication channels, and final buffers on message service quality.

Key words: IoT, edge computing, fog computing, queueing systems, delay, message loss

References

1. Kucheryavy A., Okuneva D., Paramonov A., Huang N. Methods of Traffic Distribution in a Heterogeneous High-Density Internet of Things Network // Proceedings of Telecommunication Universities. 2024. Vol. 10. Iss. 2. PP. 67–74. (in Russian) DOI: 10.31854/1813-324X-2024-10-2-67-74. EDN: RTNVEU
2. Volkov A. N., Muthanna A. S. A., Kucheryavy A. E., Borodin A. S., Paramonov A. I., et al. Perspective Research of Networks and Services 2030 in the Laboratory 6G MEGANETLAB SPbSUT // Electrosvyaz. 2023. Iss. 6. PP. 5–14. (in Russian) DOI: 10.34832/ELSV.2023.43.6.001. EDN: CJSYLS
3. Muthanna A. S. A. Model for Integrating Edge Computing into an Air-Ground Network Structure and Offloading Traffic Method for High and Ultra-High

Densities Internet of Things Networks // Proceedings of Telecommunication Universities. 2023. Vol. 9. Iss. 3. PP. 42–59. (in Russian) DOI: 10.31854/1813-324X-2023-9-3-42-59. EDN: SBAHAR

4. Proferansov D. Yu., Safonova I. E. To the Question of Fog Computing and the Internet of Things // Education Resources and Technologies. 2017. Iss. 4 (21). PP. 30–39. (in Russian) DOI: 10.21777/2500-2112-2017-4-30-39. EDN: YMQGJD

5. Glushak E. V., Kluyev D. S. Development and Research Cloud and Fog Computing Operation Models // Radioengineering. 2025. Vol. 89. Iss. 3. PP. 157–168. (in Russian) DOI: 10.18127/j00338486-202503-14. EDN: IGUDLR.

6. Al-Kerea Z. A. H., Muthanna A. S. A., Kucheryavy A. E. Intelligent Serverless Computing System for Telepresence Services // Proceedings of Telecommunication Universities. 2025. Vol. 11. Iss. 4. PP. 18–27. (in Russian) DOI: 10.31854/1813-324X-2025-11-4-18-27. EDN: UOYTAY

7. Babaeva B. E., Lisovskaya E. Y. Model of a Queueing Network with Three Nodes for Calculating the Required Computing Capacity of the Internet of Things Network // Information Technologies and Mathematical Modeling (ITMM 2024): Proceedings of the 23rd International Conference named after A. F. Terpugov. Tomsk: National Tomsk State University Publ., 2024. PP. 91–97 (in Russian). EDN: AZBPAK

8. Kleynrok L. Queueing Theory. Moscow: Mashinostroenie Publ., 1979. 432 p. (in Russian)

9. Iversen V.B. Teletraffic Engineering and Network Planning. DTU Fotonik, 2015. 398 p.

10. Pourghebleh B., Navimipour N. J. Data aggregation mechanisms in the Internet of Things: A systematic review of the literature and recommendations for future research // Journal of Network and Computer Applications. 2017. Vol. 97. PP. 23–24. DOI: 10.1016/j.jnca.2017.08.006

11. Zeliger N. B., Chugreev O. S., Yanovskiy G.G. Design of Networks and Digital Data Transmission Systems: Textbook for Higher Education Institutions. Moscow.: Radio & Svyaz' Publ., 1984. 176 p. (in Russian)

12. Shneps-Shneppe M. A. Information Distribution Systems. Methods of Calculation: Reference Manual. Moscow: Svyaz' Publ., 1979. 344 p. (in Russian)

13. Kingman J. F. C. The Single Server Queue in Heavy Traffic // Mathematical Proceedings of the Cambridge Philosophical Society. 1961. Vol. 57. Iss. 4. PP. 902–904. DOI: 10.1017/S0305004100036094

14. Kingman J. F. C. On Queues in Heavy Traffic // Journal of the Royal Statistical Society. Series B (Methodological). 1962. Vol. 24. Iss. 2. PP. 383–392. DOI: 10.1111/j.2517-6161.1962.tb00465.x

15. Shneps-Shneppe M. A. Numerical Methods of Teletraffic Theory. Moscow: Svyaz' Publ., 1974. 232 p. (in Russian)

Information about Authors

Komarov Ivan – Postgraduate student of the Department of Communication and Data Transmission Networks (The Bonch-Bruевич Saint Petersburg State University of Telecommunications). E-mail: komarov1.ii@sut.ru

Paramonov Alexander – D. Sc. of Engineering Sciences, Professor, Professor at the Department of Communication and Data Transmission Networks (The Bonch-Bruевич Saint Petersburg State University of Telecommunications).
E-mail: paramonov@sut.ru