

Kuranov A.S., Fedorov A.S. Research of the IoT Devices load impact on the Mobile Network Core

Исследование влияния нагрузки устройств Интернета Вещей на ядро сети мобильной связи

Kuranov Alexander Sergeevich,

Graduate Student, Saint-Petersburg National Research University of Information Technologies, Mechanics and Optics

Fedorov Andrey Sergeevich,

Engineer of Research and Education Center «Wireless Infotelecommunication Networks», SPbSUT

Куранов Александр Сергеевич,

Магистрант, Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики

Федоров Андрей Сергеевич,

Инженер научно-образовательного центра «Беспроводные инфотелекоммуникационные сети», СПбГУТ им. проф. М.А. Бонч-Бруевича

***Abstract.** This article is devoted to research of the IoT Devices load impact on the Mobile Network Core. The developed and described model allows to evaluate the MPS indicator at various network nodes, the time of communication establishment, the time of data transmission and packet loss. A structural diagram of the model used, a description of its parameters and modeling conditions, results and conclusions are presented.*

***Keywords:** Internet of Things, IoT, mobile communications, core, load.*

***Аннотация.** Данная работа посвящена исследованию нагрузки, создаваемой устройствами Интернета Вещей, на сигнальные интерфейсы ядра сети четвертого поколения. Разработанная и описанная модель позволяет оценить показатель MPS (messages per second) на различных узлах сети, время установления связи и время передачи данных, а также потери пакетов по истечению времени ожидания ответа или переполнению буфера. Представлена структурная схема используемой модели, описание ее параметров и условий моделирования, результаты и выводы.*

***Ключевые слова:** Интернет Вещей, мобильная связь, ядро, нагрузка.*

DOI 10.54092/25421085_2022_4_90

Рецензент: Сагитов Рамиль Фаргатович, кандидат технических наук, доцент, заместитель директора по научной работе в ООО «Научно-исследовательский и проектный институт экологических проблем», г. Оренбург

Поскольку большую часть «абонентов» сетей Интернета Вещей являются датчики, проводящие мониторинг какого-либо параметра и периодически отправляющие данные, то для работы модели были введены следующие допущения:

- трафик с датчиков поступает с некоторым постоянным периодом;
- трафик передается только в восходящей линии, в нисходящей линии передаются только подтверждения о приеме (ACK);

- PGW и SGW объединены в единый узел UGW (Universal Gateway), часто использующийся в домашней сети, сигнальная нагрузка на PGW при этом минимальна;
- нагрузка, создаваемая пользовательскими пакетами на пользовательскую плоскость, не рассматривается, поскольку является слишком малой по сравнению с обычным трафиком от смартфонов;
- в случае неуспешной обработки транзакции (например, при переполнении буфера или срабатывания таймаута ожидания в очереди) вся процедура для данного устройства считается неуспешной и переотправки не происходит;
- в рассматриваемых далее сценариях не учитываются процедуры Attach и Detach, поскольку данные процедуры происходят при включении устройства и при сбоях в сети или на устройстве, что происходит крайне редко, поэтому если данные процедуры происходят постоянно, то предположительно на каком-то узле имеется ошибка конфигурации;
- процедура PDN Connectivity также не рассматривается, так как предполагается, что датчикам достаточно одного сквозного канала для передачи данных;
- не рассматривается нагрузка на HSS, так как данный элемент не участвует в рассматриваемых сценариях.

В рассматриваемой модели используются следующие сценарии:

- передача данных через NAS по сценарию CP Clot оптимизации с дальнейшей передачей данных на SCEF по интерфейсу T6a;
- передача данных по NAS по сценарию CP Clot оптимизации с дальнейшей передачей данных в IP сеть по интерфейсу S11-U;
- передача данных с UP Clot оптимизацией через SGW.

Реализованные в модели алгоритмы передачи сообщений для каждого из приведенных выше сценариев представлены на рисунках 1, 2 и 3 соответственно.

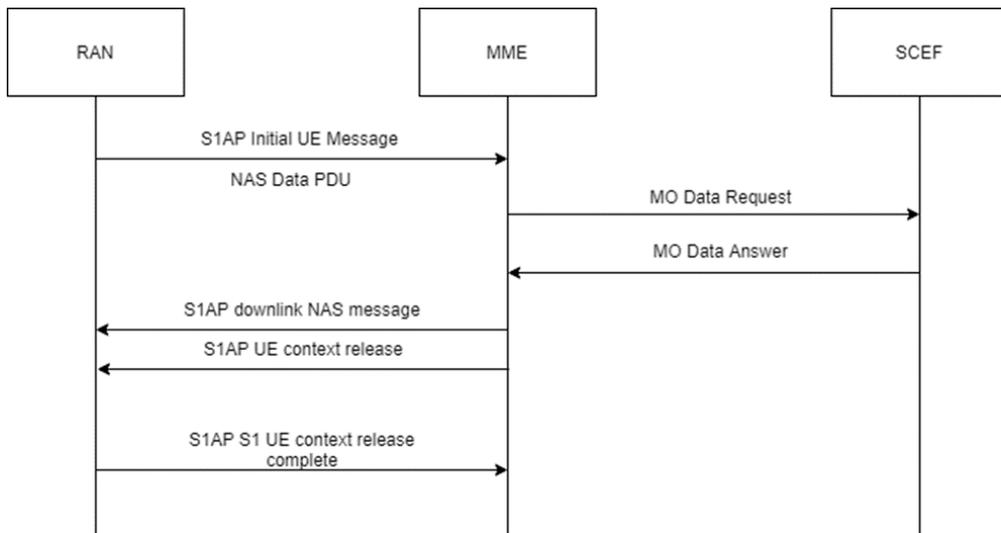


Рисунок 1. Рассматриваемый сценарий №1

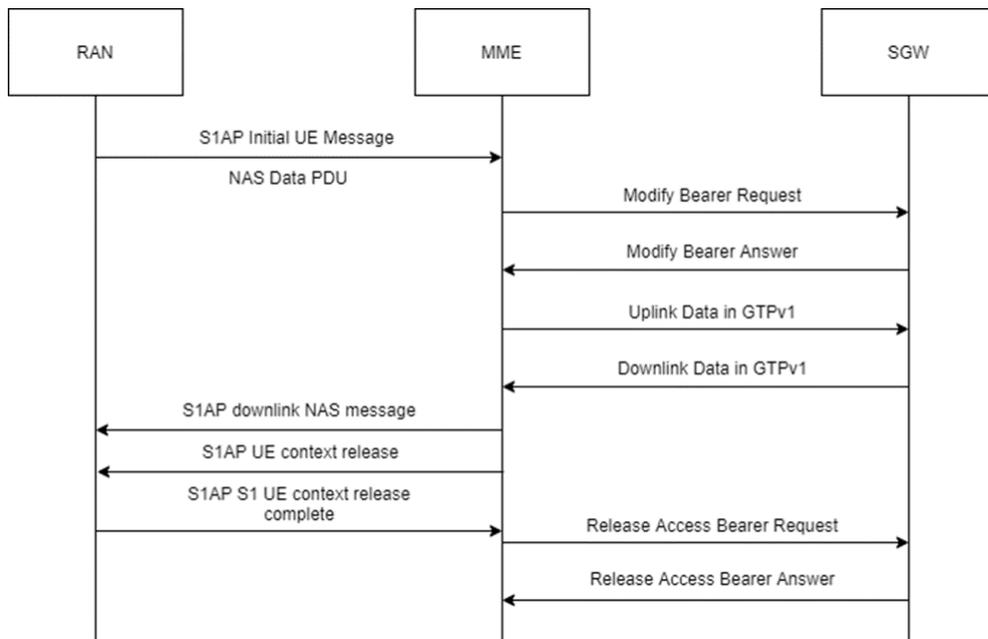


Рисунок 2. Рассматриваемый сценарий №2

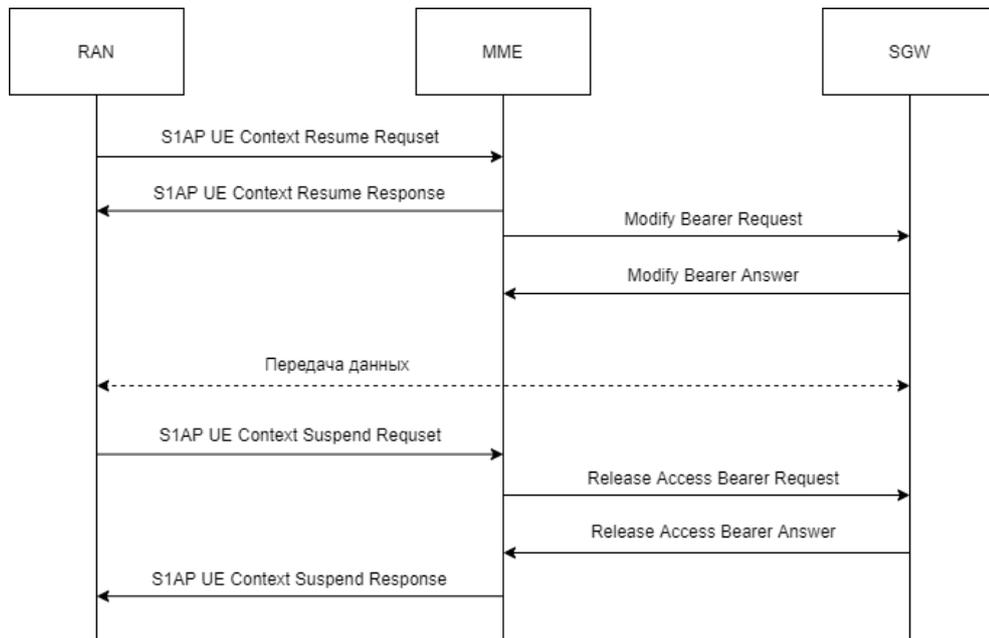


Рисунок 3. Рассматриваемый сценарий №3

Консорциум 3GPP предлагает в качестве модели генерируемого трафика использовать процесс Пуассона. В основе их заключения лежит теорема Палма-Хинчина. Она гласит, что суперпозиция большого числа независимых процессов будет обладать пуассоновскими свойствами и, следовательно, может быть описана пуассоновским процессом [1]. Таким образом, время между прибытиями заявок будет подчиняться Пуассоновскому процессу и будет описываться следующим выражением [2]:

$$A = e^{\lambda} = e^{n/T} \quad (1)$$

В выражении (1):

- $\lambda = n/T$ – интенсивность прибытия заявок;
- n – количество устройств;
- T – период отправки данных устройствами.

Для проведения имитационного моделирования использовалось ПО AnyLogic [3]. Разработанная модель представлена на рисунке 4, параметры модели представлены в таблице 1. Время задержки в сети радиодоступа (RAN) представлено треугольным распределением.

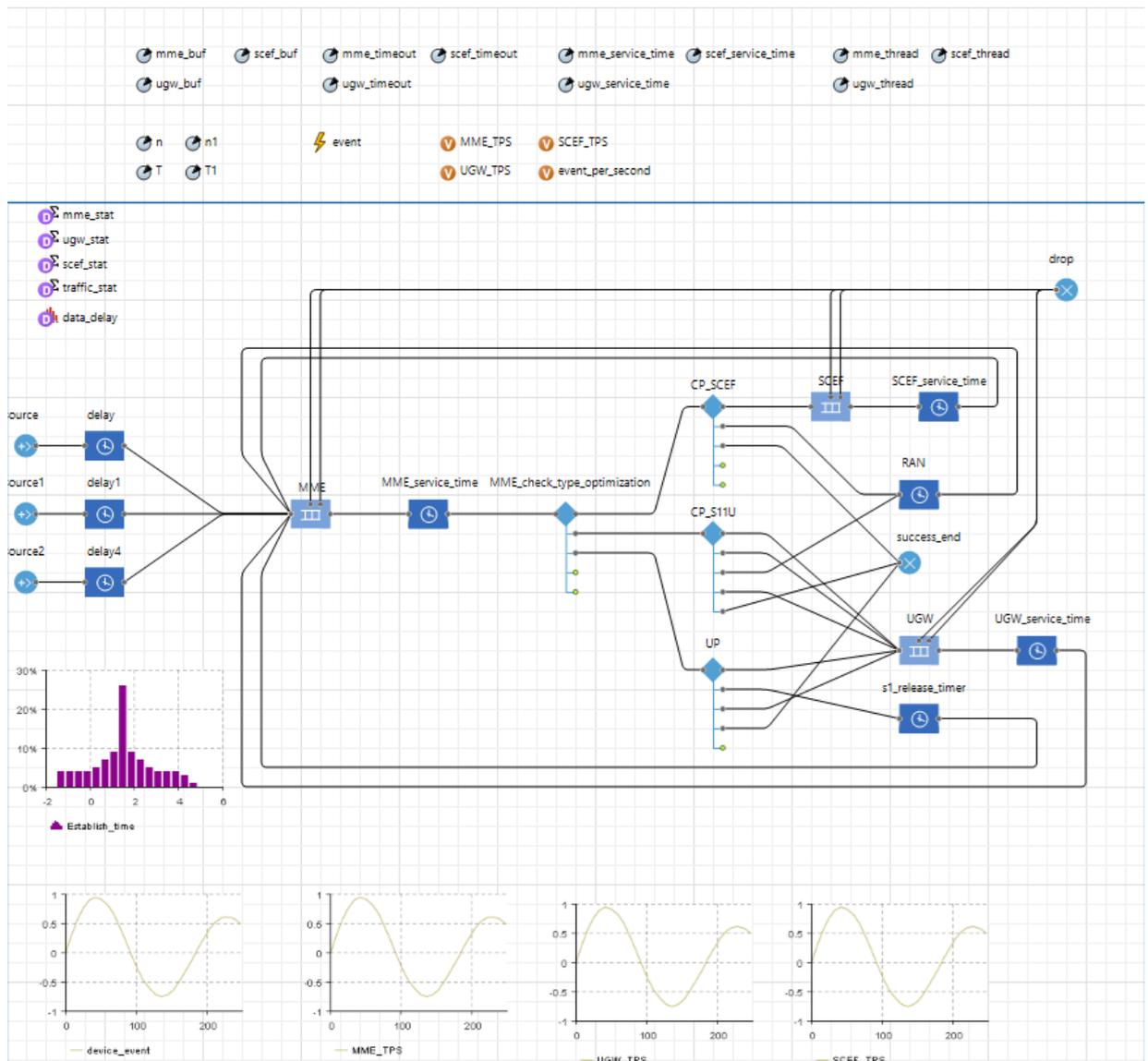


Рисунок 4. Разработанная модель

Параметры, использующиеся в модели

Параметр	Описание
mme_service_time	Время обработки транзакции на MME
ugw_service_time	Время обработки транзакции на UGW
scef_service_time	Время обработки транзакции на SCEF
mme_buf	Размер буфера сообщений на MME
ugw_buf	Размер буфера сообщений на UGW
scef_buf	Размер буфера сообщений на SCEF
mme_timeout	Время ожидания в буфере на MME
ugw_timeout	Время ожидания в буфере на UGW
scef_timeout	Время ожидания в буфере на SCEF
mme_thread	Количество одновременно обрабатываемых сообщений на MME
ugw_thread	Количество одновременно обрабатываемых сообщений на UGW
scef_thread	Количество одновременно обрабатываемых сообщений на SCEF
s1_release_time	Время вызова процедуры s1 release при отсутствии трафика
n	Количество устройств
T	Период отправки сообщений

Все эксперименты, описываемые в работе, проведены с параметрами, представленными на рисунке 5, если явно не указано иное.

Параметр	Тип	Значение		
		Мин.	Макс.	Шаг
n	Диапазон	1	100	1
T*	Фиксированный	1		
n1	Фиксированный	0		
T1	Фиксированный	1000		
ugw_buf	Фиксированный	1000		
mme_buf	Фиксированный	1000		
scef_buf	Фиксированный	1000		
mme_timeout*	Фиксированный	3		
ugw_timeout	Фиксированный	3		
scef_timeout	Фиксированный	3		
mme_service_time	Фиксированный	100		
ugw_service_time	Фиксированный	100		
scef_service_time	Фиксированный	100		
ugw_thread	Фиксированный	10		
mme_thread	Фиксированный	10		
scef_thread	Фиксированный	10		
s1_release_time*	Фиксированный	10		
ran_mode_transmission_time	Фиксированный	30		
ran_min_transmission_time	Фиксированный	20		
ran_max_transmission_time	Фиксированный	70		
n2	Фиксированный	0		
T2	Фиксированный	1000		

Рисунок 5. Значения параметров модели

На рисунках 6, 7, 8 представлено генерируемое количество запросов от IoT абонентов, количество сообщений в секунду на MME и количество отброшенных пакетов соответственно.

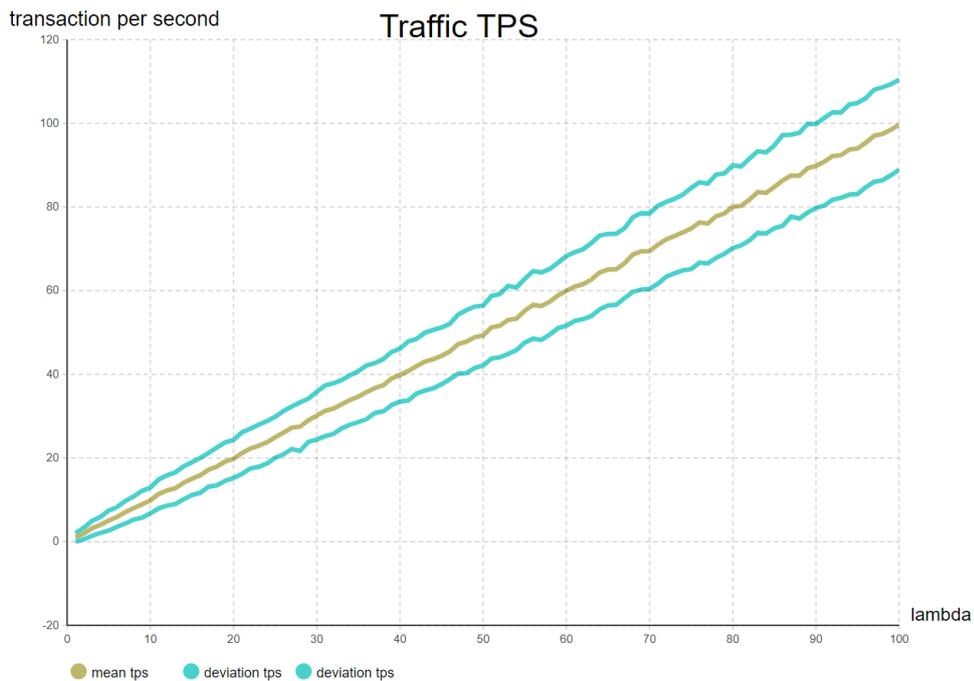


Рисунок 6. Генерируемое количество запросов от IoT абонентов

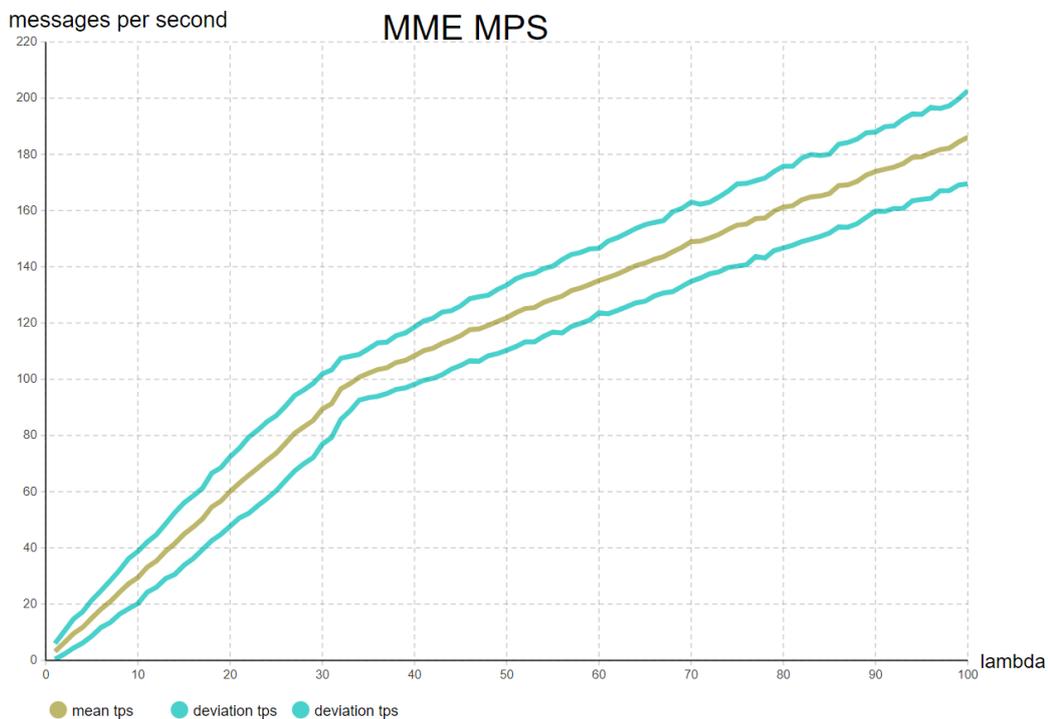


Рисунок 7. Количество сообщений в секунду на MME

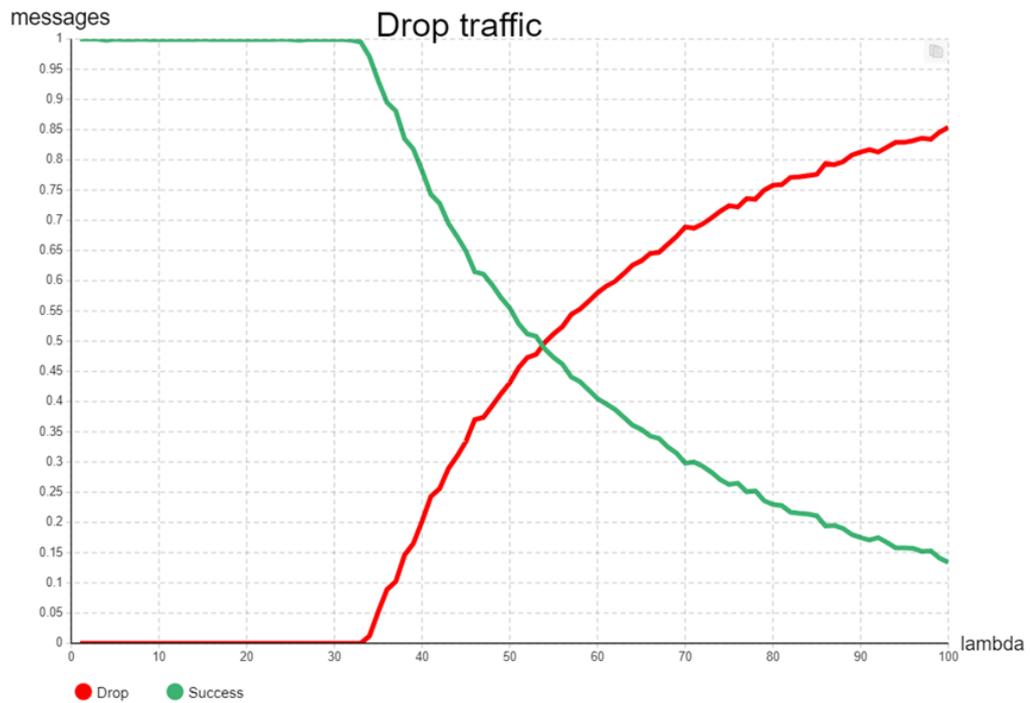


Рисунок 8. Количество отброшенных пакетов

По графикам на приведенных выше рисунках видно, что с ростом количества устройств линейно растет и количество поступающих транзакций в секунду. В момент, когда средняя нагрузка на ММЕ приближается к 100 сообщениям в секунду, возникают транзакции, которые неуспешно завершаются по истечению таймаута ожидания в буфере. При этом коэффициент роста меняется, но продолжает расти также линейно.

Подобная ситуация возникает и с SCEF, который возвращает обработанные сообщения обратно на ММЕ, что представлено на рисунке 9.

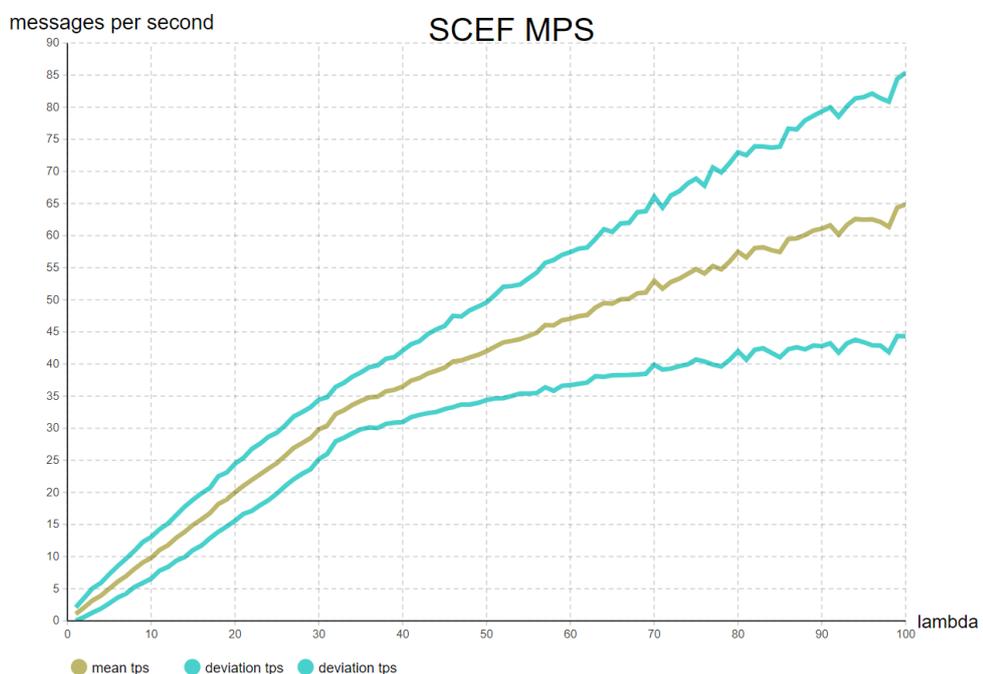


Рисунок 9. Количество сообщений в секунду на SCEF

Нагрузка на SCEF продолжает расти. Хотя на выходе MME должно быть 100 транзакций в секунду, на SCEF продолжается рост поступающих заявок, так как с ростом числа входящих заявок на MME увеличивается и соотношение новых заявок и ответов от SCEF. Аналогичная ситуация происходит и с UGW в других сценариях. Логично предположить, что при заданных характеристиках системы количество заявок на SCEF не превысит производительность MME. Это можно проверить, если понизить производительность MME до 20 MPS. Тогда видно, что рост поступающих на SCEF транзакций не превысит 20 MPS, так как MME может пропустить больше 20 MPS, что представлено на рисунке 10.

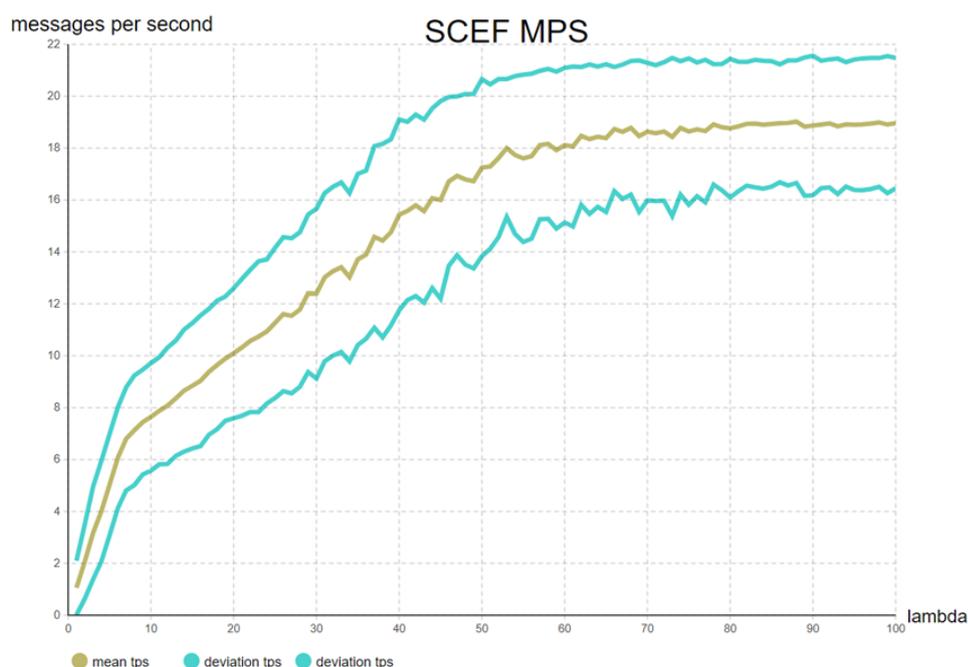


Рисунок 10. Количество сообщений в секунду на SCEF (производительность MME в 20 MPS)

На рисунках 11, 12 и 13 представлены зависимости показателей MPS и TPS при сценариях, описанных выше. В сценарии 1 при отсутствии неуспешно завершённых транзакций каждая транзакция генерирует на MME 3 сообщения и 1 на SCEF. В сценарии 2 каждая транзакция генерирует на MME 5 сообщений, а на SGW при этом генерируется 3 сообщения. В сценарии 3 каждая транзакция генерирует нагрузку на MME в 4 сообщения и 2 на SGW.

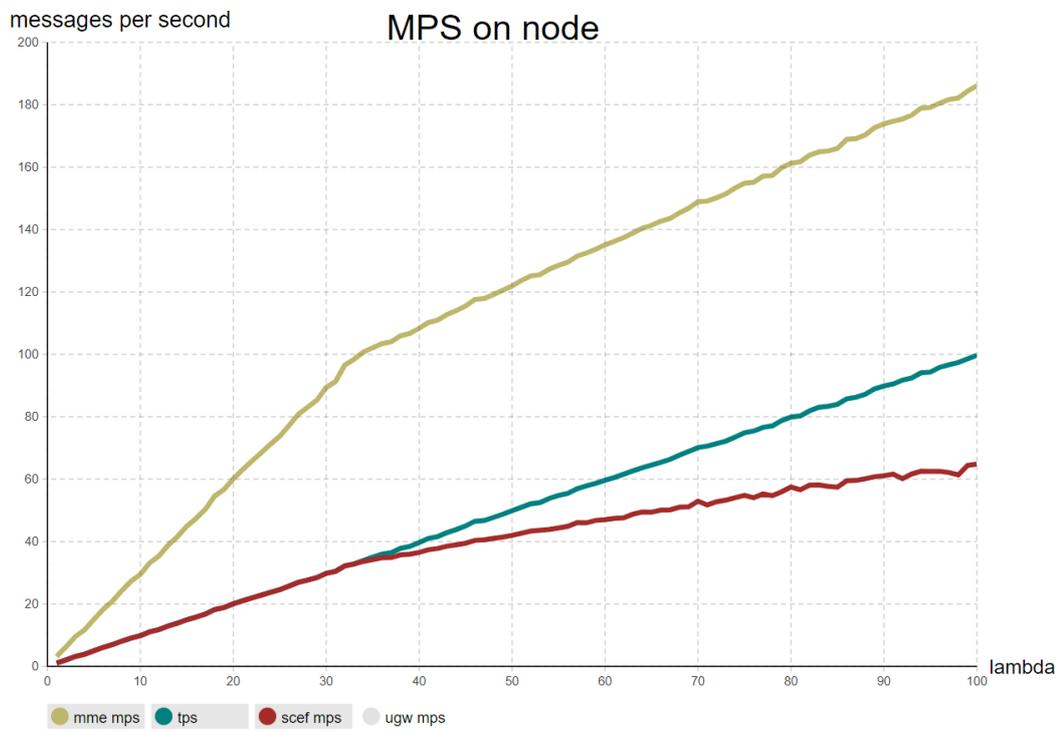


Рисунок 11. MPS при сценарии №1

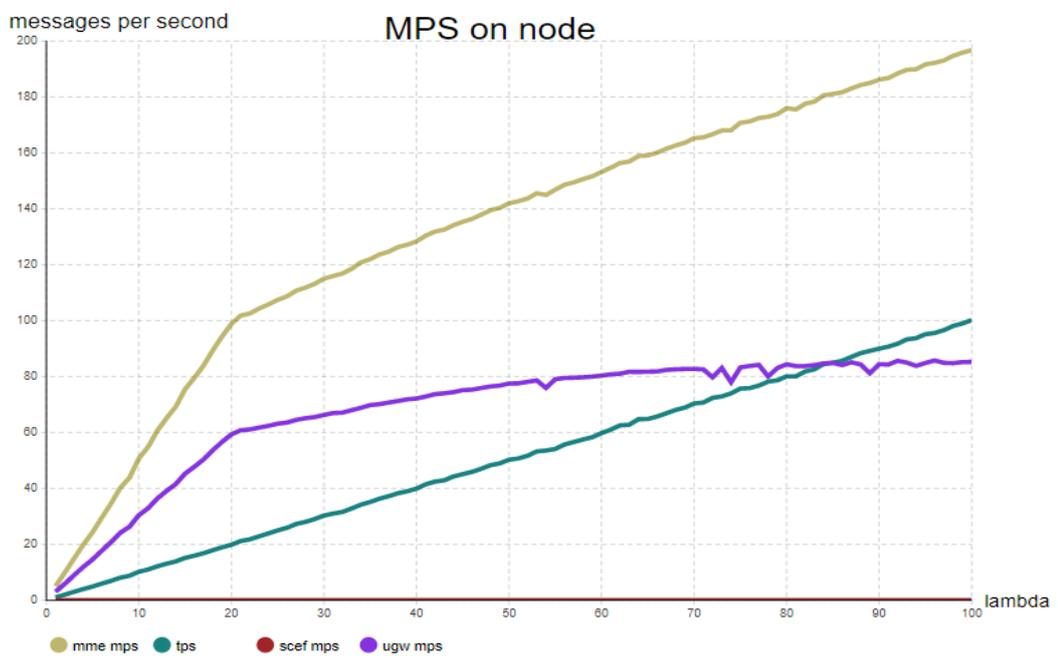


Рисунок 12. MPS при сценарии №2

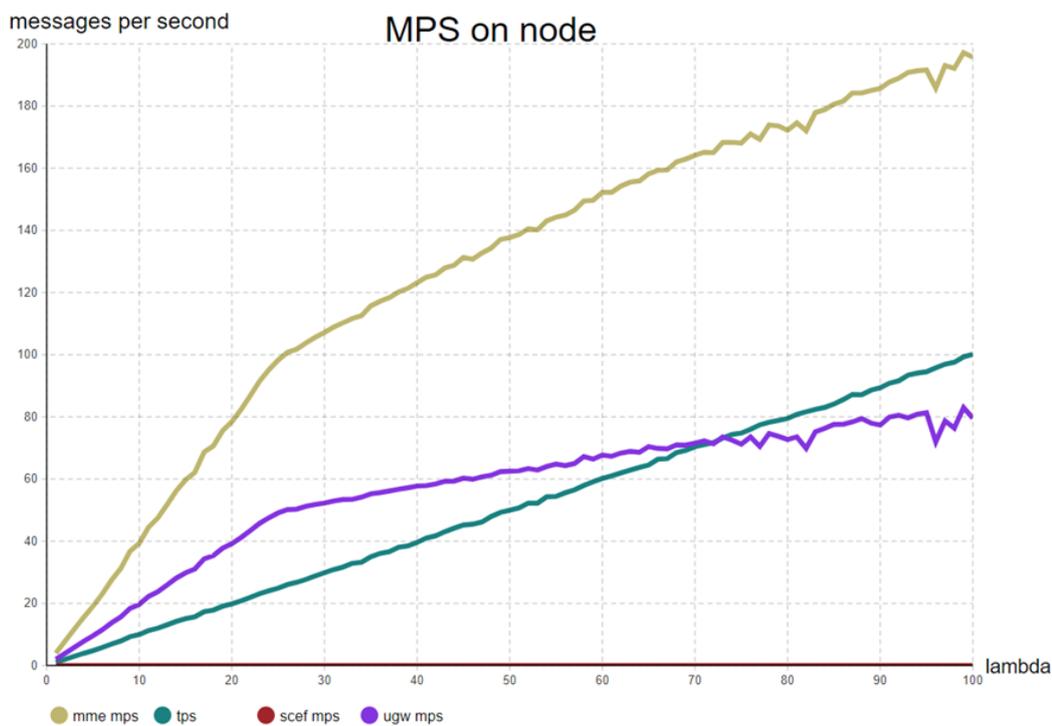


Рисунок 13. MPS при сценарии №3

На рисунках 14 и 15 показана зависимость времени доставки сообщения. Время доставки сообщения до сервера резко возрастает незадолго до начала отбрасывания сообщений по таймауту. С увеличением нагрузки время доставки стремится к значению 3.2 с (MME timeout 3 с + MME service time 100 мс + SCEF service timeout 100 мс). При этом, если уменьшить буфер так, чтобы пакеты стали отбрасываться не по таймауту, а по отсутствию места в буфере, то время, к которому будет стремиться время передачи пакетов, станет зависеть от объема буфера. Это время будет равно сумме отношению размера буфера к производительности узла и времени обработки транзакции. Например, на рисунке 15 $mme_buf = 150$ мс, $mme_thread = 10$ мс и $mme_service_time = 100$ мс. А время стремится к 1.7 с.

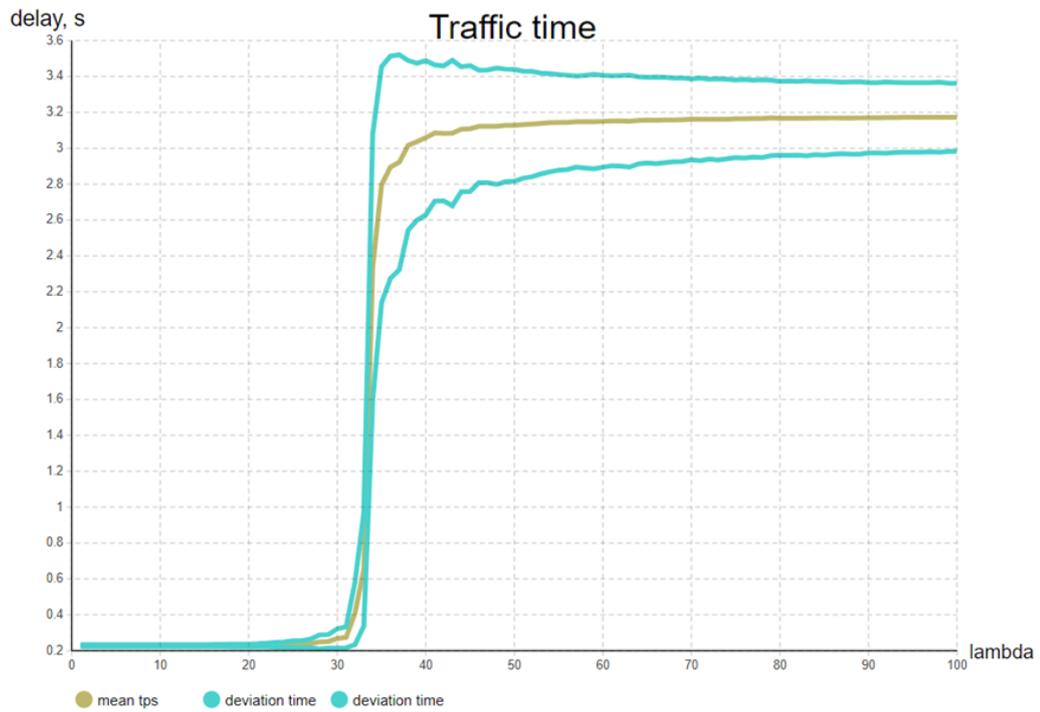


Рисунок 14. Время доставки сообщения

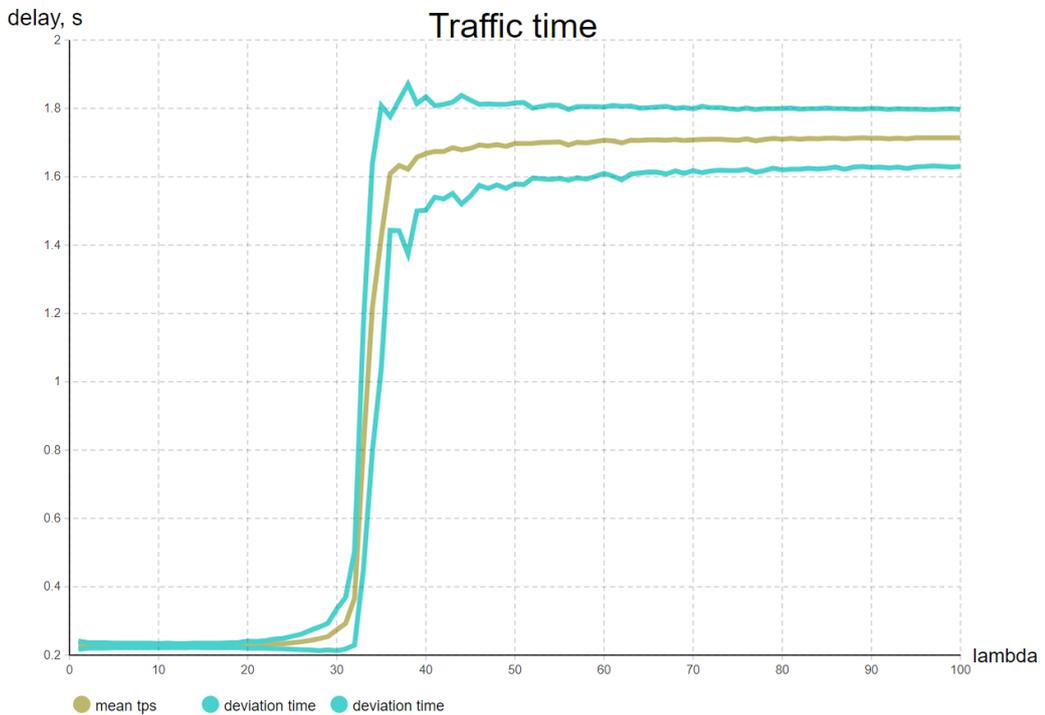


Рисунок 15. Время доставки сообщения

Рисунок 16 позволяет детальнее момент до роста нагрузки в 100 MPS на ММЕ – видно увеличение СКО времени обработки сообщения. До 93 MPS время обработки увеличилось не более, чем в 2 раза, а при 99 MPS среднее время увеличилось примерно в 3 раза.

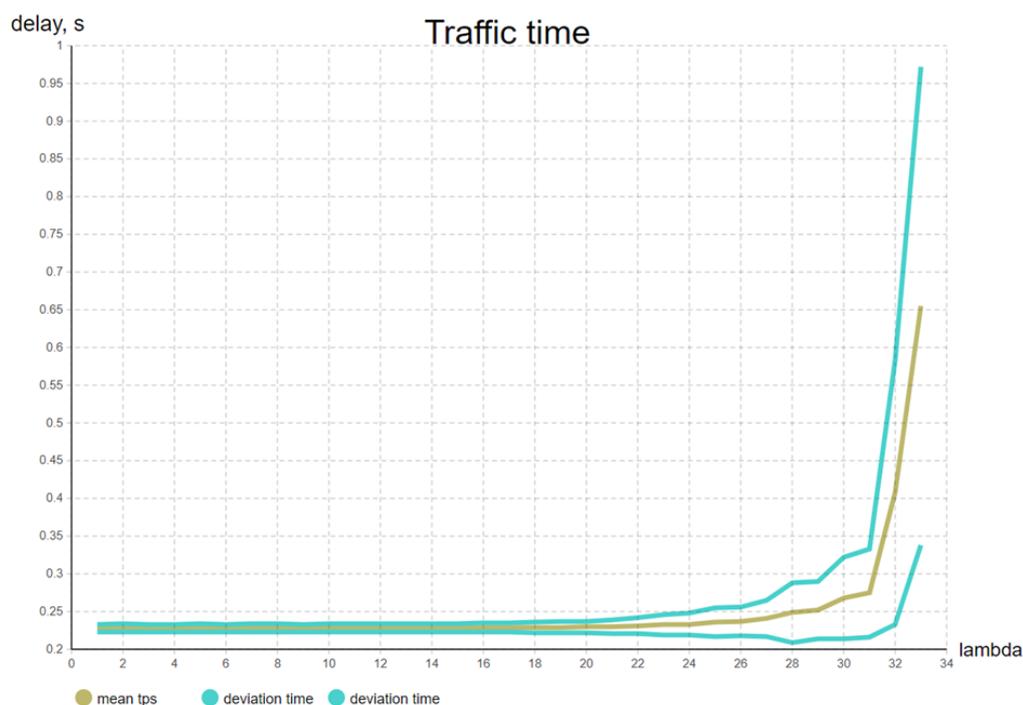


Рисунок 16. Время доставки сообщения (в масштабе)

Таким образом, были проанализированные полученные данные и выявлен характер роста нагрузки на элементы EPC (evolved packet core) при росте числа устройств в сети. Выявлено, что при приближении средней нагрузки к значению производительности системы, происходит небольшой рост задержки времени установления соединения/передачи данных. А при превышении производительности – резкий скачок.

References

1. В. А. Кокотушкин, Одно обобщение теоремы Пальма–Хинчина, Теория вероятн. и ее примен., 1974, том 19, выпуск 3.
2. А. Я. Хинчин, Математические методы теории массового обслуживания, Тр. МИАН СССР, 1955, том 49.
3. AnyLogic – имитационное моделирование. URL: <https://www.anylogic.ru/>.