

АНАЛИЗ СТАТИСТИЧЕСКИХ ЗАКОНОМЕРНОСТЕЙ ИНФОРМАЦИОННЫХ ПОТОКОВ В IP СЕТЯХ

ANALYSIS OF STATISTICAL PATTERNS OF INFORMATION FLOWS IN IP NETWORKS

**A. Pugach
D. Stepanova
K. Gaipov**

Summary. This article studies the loss process in Internet Protocol (IP) networks. Obtaining information about network loss patterns is important for measuring network performance during multimedia traffic. To analyze the packet loss process, specific information was collected about when a loss occurs in the network, which packets are lost, whether the loss clusters into packets (episodes), the duration of the loss episodes, and what was the impact of network protocols on the loss. The main result of the work is a series of measurements in order to find out the mutual influence of information flows between interfaces on the value of the loss probability of a single stream, as well as to find out whether there will be any dependence on packet losses. This article includes 16 pages, 6 figures and 9 references.

Keywords: quality of service, traffic generator, IP packet.

Пугач Алексей Витальевич

Сибирский государственный университет науки
и технологий имени академика М.Ф. Решетнева,
Красноярск
dark.penek@mail.ru

Степанова Дарья Семёновна

Сибирский государственный университет науки
и технологий имени академика М.Ф. Решетнева,
Красноярск
052299@bk.ru

Гаипов Константин Эдуардович

К.т.н., доцент, Сибирский государственный
университет науки и технологий имени академика
М.Ф. Решетнева, Красноярск
gaipovke@yandex.ru

Аннотация. В данной статье производится исследование процесса потери в сетях Интернет-протокола (IP). Получение информации о закономерностях потерь в сети важно для измерения производительности сети во время передачи мультимедийного трафика. Для анализа процесса потери пакетов была собрана конкретная информация о том, когда потеря появляется в сети, какие пакеты теряются, кластеризируются ли потери в пакеты (эпизоды), продолжительность эпизодов потерь, и каково было влияние сетевых протоколов на потери. Основным результатом работы является ряд измерений с целью выяснения взаимного влияния потоков информации между интерфейсами на значение вероятности потерь отдельно взятого потока, а также выяснение того, будет ли проявляться какая-либо зависимость у потерь пакетов. Данная статья включает в себя 16 страниц, 6 рисунков и 9 источников списка литературы.

Ключевые слова: качество обслуживания, генератор трафика, IP пакет.

Введение

В настоящее время качество обслуживания сетей с большим количеством пользователей является довольно серьезным объектом для анализа, так как исходя из индивидуальных запросов каждого в сети, необходимо продумывать то, как же предоставить необходимые услуги без каких-либо потерь, задержек и с оптимальной пропускной способностью. Чтобы предоставить пользователю все вышеперечисленное, в данной работе использован имитационный подход, суть которого заключена в том, чтобы собрать данные потоков о потерях в сегментах сети, через которые проходит трафик пользователей. После их сбора, будут проанализированы потери пакетов с каждого потока, так и их совокупность. В качестве основных пакетов для передачи

использовались пакеты с протоколом UDP, так как они участвуют в процессе приема или передачи каких-либо данных они являются наиболее распространенными.

Для того, чтобы проводить сбор данных, необходимо определиться с характеристиками, которые будут описывать поток трафика в сети и инструментами для анализа статистических данных.

Среди таких характеристик, можно выделить следующие

1. Время задержки
2. Вероятность потерь
3. Коэффициент самоподобия
4. Длина очереди
5. Коэффициент пачечности

Низкая задержка — важный фактор, обеспечивающий надежную работу и высокую производительность сетей. Приложения для общения в реальном времени, стриминга и проведения транзакций сильно зависят от времени ожидания. Увеличение задержки всего на несколько миллисекунд может привести к искажению изображения и голосов, «зависанию» приложений и финансовым потерям[1].

Провайдеры стараются следить за пропускной способностью сети и колебанием задержек, но увеличение «ширины» канала часто не влияет на задержку в работе сети.

В сетях, основанных на пакетном обмене, связь между задержкой и пропускной способностью неоднозначна и сложна в определении. При этом время ожидания складывается из следующих компонентов:

1. задержка сериализации — время, необходимое порту для передачи пакета;
2. задержка распространения — время, требуемое биту информации для достижения приёмника (обуславливается законами физики);
3. задержка при перегрузке — время, которое кадр проводит в выходной очереди элемента сети;
4. задержка передачи — время, которое сетевой элемент тратит на анализ, обработку и передачу пакета.

Кроме этого, потеря пакетов с данными также влияют на приложения и их работу, которые зависят от передачи данных в режиме реального времени, например, программы для потоковой передачи аудио или видео, а также других различных цифровых системных платформ.

Обмен пакетами в интернете производится с помощью стека протоколов TCP/IP[2] (TCP — протокол обмена передачей / IP — межсетевой протокол). Это несколько протоколов, которые обеспечивают online передачу данных, определяют их форму, способ отправки и так далее.

Так же, под IP понимают адрес какого-либо ресурса непосредственно в интернете. При попытке обращения к этому ресурсу по адресу (IP) высылаются пакеты данных. Чтобы эти данные не терялись и не дублировались, их контролирует протокол обмена (TCP). Если данные были доставлены, адресат в свою очередь отправляет ответные пакеты. На любом этапе может произойти ошибка, что будет являться потерей пакета. Одной из возможных причин потерь являются переполнение сети, когда достигается максимальная пропускная способность и возрастающий трафик не может быть передан в виду сильной загруженности, что мо-

жет привести к сбою в соединении и потери пакетов. Другой же является проблемы сетевого оборудования и программного обеспечения, когда для запуска системы используются устаревшие аппаратные устройства, пакеты могут быть потеряны из-за медленной передачи данных. Компаниям и частным лицам рекомендуется постоянно обновлять или модернизировать свое аппаратное обеспечение для оптимизации производительности сетевых процессов. Это необходимо, чтобы избежать задержек в работе сети, потери пакетов или даже полной потери соединения с системой.

Под самоподобием подразумевается повторяемость распределения нагрузки во времени при различных масштабах. Если набор значений самоподобной функции (т.е. проявляющей признаки самокорреляции) разделить на равные группы, а затем просуммировать значения внутри групп, то набор сумм будет подчиняться той же самой корреляционной функции, что и исходные данные.

В отличие от пуассоновских самоподобные процессы характеризуются наличием последствия: вероятность наступления очередного события зависит не только от времени, но и от предыдущих событий (предыстории). Это означает, что число текущих событий может зависеть от числа предыдущих событий в отдаленные промежутки времени. Поэтому одним из основных свойств самоподобного процесса является медленно убывающая зависимость между объемами трафика в разные моменты времени[3].

Самоподобный процесс часто носит взрывной (burst) характер, что проявляется в наличии выбросов при относительно низкой скорости потока событий. Эти локализованные во времени скопления (congestions) вызывают значительные потери пакетов, даже когда суммарная потребность в обслуживании всех потоков далека от максимально допустимых значений, поскольку расчеты требуемых характеристик современных сетей используют лишь усредненные свойства трафика. В частности, для самоподобного потока пакетов при увеличении размера буфера на входе сетевого процессора вероятность потерь падает значительно медленнее, чем для экспоненциального закона, используемого в классических моделях телетрафика. Алгоритмы обработки трафика сетевым процессором, предназначенные для работы с простейшим потоком, неэффективны для самоподобных потоков.

Возможные причины самоподобия сетевого трафика — в особенностях распределения файлов по серверам, их размерах, а также в типичном поведении пользователей. Изначально не проявляющие свойств самоподобия потоки данных, пройдя обработку на уз-

ловых серверах и активных сетевых элементах, начинают проявлять ярко выраженные признаки самокорреляции.

Очередь — это механизм, позволяющий управлять приемом и передачей пакетов. Очередь является физическим объектом в памяти и содержит пакеты, поступившие на сетевой интерфейс.

Очереди способны контролировать скорость передачи пакетов, ограничивая нежелательный сетевой трафик по скорости (позволяет избежать выход из строя сервера или отдельных демонов в результате DoS- и даже DDoS- атак). Позволяет осуществлять распределение нагрузки между несколькими сетевыми интерфейсами. С помощью очередей также можно добиться существенного увеличения производительности сети в целом при помощи разделения различных видов трафика (например, интерактивные данные должны обрабатываться быстрее) на основе поля ToS (type of service — тип услуг)[4].

Механизмы очередей используются в любом сетевом устройстве, где применяется коммутация пакетов — маршрутизаторе, коммутаторе локальной или глобальной сети, конечном узле.

Необходимость в очереди возникает в периоды временных перегрузок, когда сетевое устройство не успевает передавать поступающие пакеты на выходной интерфейс. Если причиной перегрузки является процессорный блок сетевого устройства, то необработанные пакеты временно помещаются во входную очередь, т.е. в очередь на входном интерфейсе. В случае, когда причина перегрузки заключается в ограниченной скорости выходного интерфейса (а она не может превышать скорость поддерживаемого протокола), то пакеты временно хранятся в выходной очереди[5].

Оценка возможной длины очередей в сетевых устройствах позволила бы определить параметры качества обслуживания при известных характеристиках трафика.

Коэффициент «пачечности» (концентрации ошибок) — величина, которая показывает отношение количества ошибок, полученных в данном интервале времени по отношению к ожидаемому среднему по всем интервалам.

Коэффициент пачечности трафика определяется как отношение между максимальным и средним трафиком соответствующего сервиса. Коэффициент пачечности вычисляется по формуле:

$$K = \frac{\bar{\psi}}{\psi}$$

Средняя длительность пика $\bar{T}^{(P)}$. Средняя длительность интервала времени, в течение которого, соответствующий сервис генерирует пиковый трафик, вычисляется по формуле:

$$\bar{T}^{(P)} = \frac{1}{N^{(P)}} \sum_{i=t}^{N^{(P)}} T_i^{(P)}$$

где $N^{(P)}$ — число пиков в течение сеанса связи; $T_i^{(P)}$ — длительность i -пика процесса $B(t)$, $i = 1, N^{(P)}$, а длительность i -пика определяется выражением:

$$T_i^{(P)} = t_i^{(B)} - t_i^{(S)}$$

где $t_i^{(B)}$, $t_i^{(S)}$ — моменты начала и окончания i -пика, которые определяются следующими выражениями:

$$t_i^{(B)} = \min t, \text{ где } t_0^{(S)} \text{ и } t_0^{(B)} = 0$$

Перечисленные выше параметры используются для описания трафика соответствующего сервиса в течение одного сеанса связи с абонентом сервиса.

Интенсивность запросов λ на получение обслуживания абонентами сети у соответствующего сервиса определяется как среднее число поступивших запросов на обслуживание в единицу времени.

Средняя длительность сеанса связи \bar{T}^S — средняя продолжительность интервала времени, в течение которого соответствующий сервис обслуживает поступивший запрос.

Максимальный размер пакета \hat{S} — максимальный размер элемента трафика в битах (элемент трафика передается адресату как единое целое).

Среди инструментов для анализа статистических данных были выбраны приложения под названием DITG и MathCad

DITG — приложение, которое позволяет создавать симуляцию интернет трафика для двух версий протокола IP (IPv4, IPv6), также позволяет на основе полученных данных из данного приложения можно строить стохастические модели (модели, в которых рассматриваемые величины являются случайными. В данном случае этими величинами являются размер пакета (packet size) интервалы между вызовами пакета (inter departure time)). Данные модели позволяют рассматривать поведение протоколов уровня приложений. Помимо этого, в данном генераторе есть большое разнообразие распределений как для интервалов между вызовами, так и длины пакета, что выявляет около 45 комбинаций для

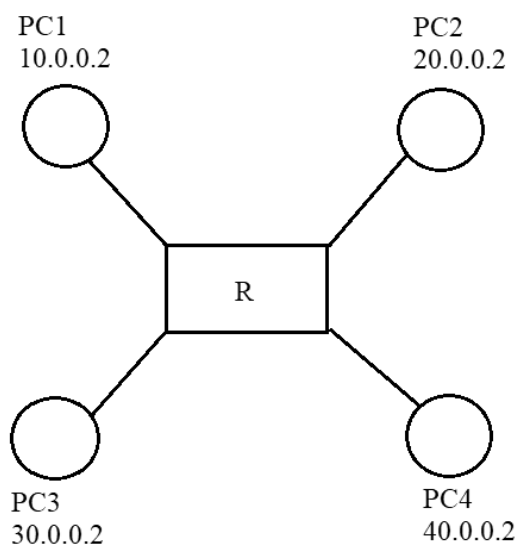


Рис. 1. Топология сети

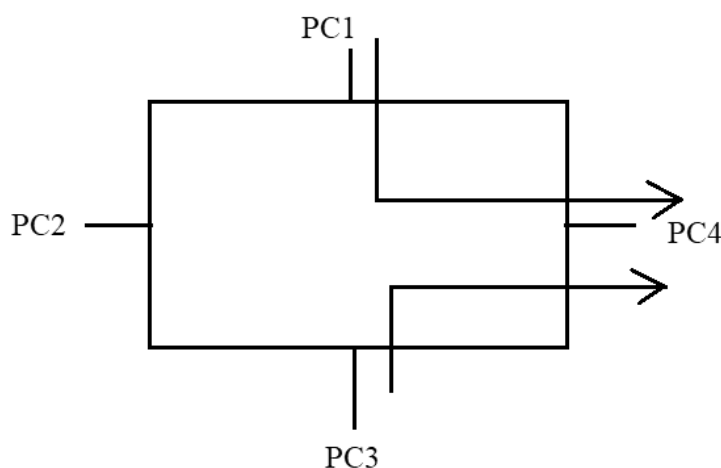


Рис. 2. Топология сети с индивидуальными потоками из PC1 и PC3 на PC4

наблюдения и изучения, выявляя тем самым вариации для ситуации с наименьшими потерями данных и наивысшей пропускной способностью[6].

Также данный генератор способен измерять наиболее распространенные для сети показатели производительности в виде задержки, пропускной способности, джиттера (колебание задержки при передаче пакета), потери пакетов.

Интерполяция включает использование существующих точек данных для прогнозирования значений между этими точками данных. Mathcad позволяет либо соединять точки данных прямыми линиями (линейная интерполяция), либо соединять их с участка-

ми кубического полинома (кубическая сплайн-интерполяция).

В отличие от функций регрессии, эти функции интерполяции возвращают кривую, которая должна проходить через указанные вами точки. Из-за этого результирующая функция очень чувствительна к ложным точкам данных. Если ваши данные зашумлены, вам следует подумать об использовании вместо них функций регрессии[7].

Линейное прогнозирование предполагает использование существующих значений данных для прогнозирования значений, выходящих за рамки существующих. Mathcad предоставляет функцию, которая позволяет

Таблица 1. Два источника с одинаковой скоростью

№	Заданная скорость (пакеты в секунду)	Реальная скорость (пакеты в секунду)	Потери пакетов (%)	Суммарная скорость (пакеты в секунду)
1	1120	1045	0	2090
2	1130	1055	0,08	2110
3	1140	1064	0,29	2128
4	1150	1073	0,59	2146
5	1160	1083	1,3	2166
6	1170	1093	2,41	2186
7	1180	1102	3,73	2204
8	1200	1121	5,52	2242
9	1300	1208	14,52	2416

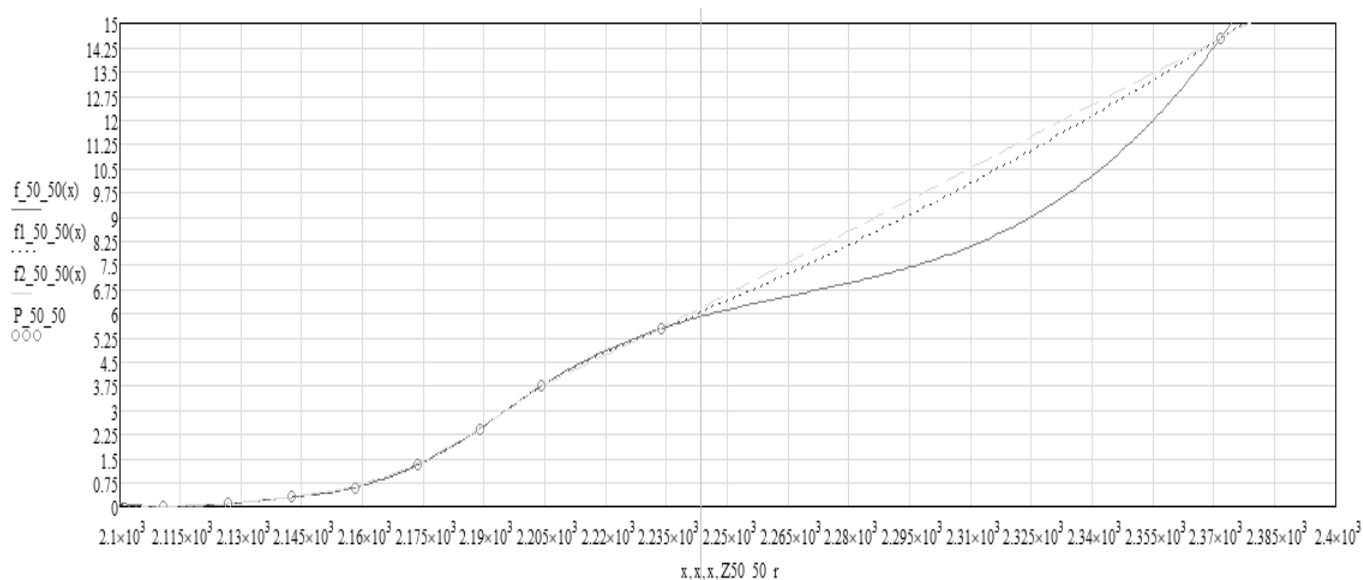


Рис. 3. Зависимость одинаковых потоков данных от потерь

прогнозировать будущие точки данных на основе прошлых точек данных[8].

Разобравшись с параметрами и инструментами, необходимыми для сбора данных, изобразим топологию, состоящая из 4 компьютеров и маршрутизатора.

Исходя из данной топологии, с помощью генератора DITG будут создаваться потоки, которые будут одновременно запущены со определенных компьютеров и будут проходить через маршрутизатор в разные пункты назначения внутри топологии. После того, как данные будут отосланы, занесем результаты в таблицу и построим соответствующие графики.

Запустим индивидуальные потоки с компьютеров PC1 и PC3 на PC4

В ходе этих измерений была составлена таблица 1, с помощью Mathcad был построен график с использованием функции interp:

Коэффициент корреляции между заданной скоростью пакетов и потерей пакетов равна 0,98, что говорит о присутствии сильной прямой взаимосвязи между этими множествами данных.

Следующие измерения были с один фиксированным потоком в 200, 400, 800 пакетов с продолжительностью

Таблица 2. Фиксированный поток в 200 пакетов

№	Заданная скорость (пакеты в секунду)	Реальная скорость (пакеты в секунду)	Потери пакетов (%)	Суммарная скорость (пакеты в секунду)
1	2120	1889	0	2082
2	2140	1910	0,05	2103
3	2160	1926	0,16	2122
4	2180	1935	0,44	2136
5	2200	1941	0,94	2152
6	2220	1951	1,14	2166
7	2240	1937	2,2	2172
8	2260	1933	3,78	2199
9	2280	1926	4,49	2205
10	2300	1923	5,55	2222

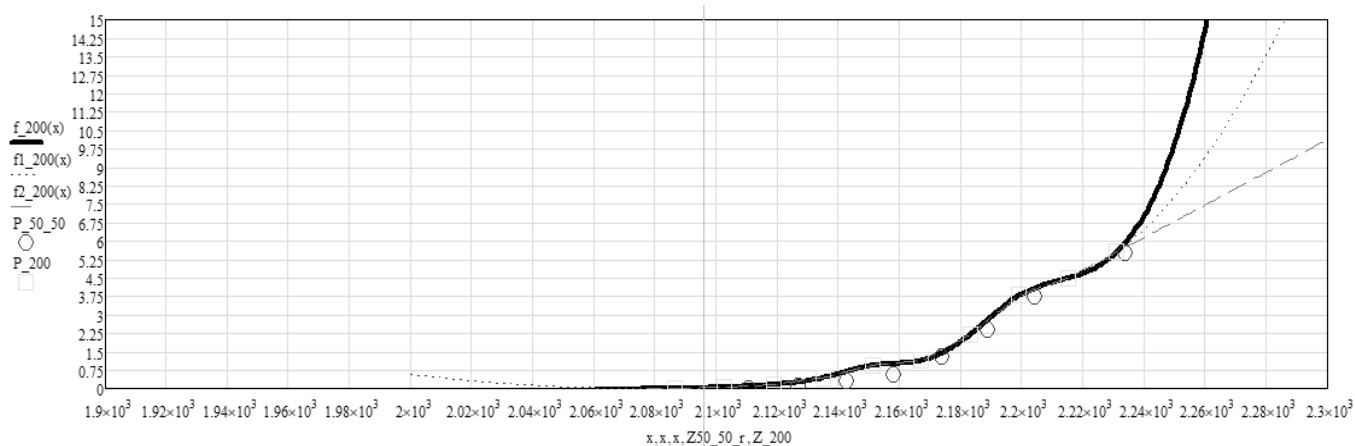


Рис. 4. Зависимость одного фиксированного потока данных в 200 пакетов от потерь

в 30 минут с PC3 на PC4, а с PC1 на PC4 поток продолжительностью 2 минуты

Коэффициент корреляции между заданной скоростью пакетов и потерей пакетов равна 0,95, что говорит о присутствии прямой зависимости между этими множествами данных.

Коэффициент корреляции между заданной скоростью пакетов и потерей пакетов равен 0,96, что говорит о присутствии сильной прямой взаимосвязи между этими множествами данных[9].

Вывод: по результатам измерений видно, что два потока идущие на один порт не влияют друг на друга и вероятность потерь зависит только от суммарного потока, при отсутствии других потоков между другими портами.

Так же было проверено, что если подавать от трех источников, то вероятность потерь зависит только от суммарной интенсивности.

Коэффициент корреляции между заданной скоростью пакетов и потерей пакетов равен 0,99, что говорит о присутствии сильной прямой взаимосвязи между этими множествами данных.

Вывод: Вероятность потерь от трех потоков идущих на один интерфейс зависит только от суммарного потока на выходном интерфейсе. На графиках видно, что вероятность для двух и трех потоков идущих на один и тот же интерфейс совпадает.

В ходе данной выпускной квалификационной работы на основе маршрутизатора Cisco 2811 были проведены ряд измерений с целью выяснения взаимного

Таблица 3. Фиксированный поток в 400 пакетов

№	Заданная скорость (пакеты в секунду)	Реальная скорость (пакеты в секунду)	Потери пакетов (%)	Суммарная скорость (пакеты в секунду)
1	1840	1661	0	2045
2	1860	1686	0	2070
3	1880	1697	0	2081
4	1900	1712	0	2096
5	1920	1714	0,05	2098
6	1940	1725	0,14	2111
7	1960	1742	0,38	2132
8	1980	1750	0,5	2142
9	2000	1757	2,17	2179
10	2020	1752	3,15	2191
11	2040	1757	3,88	2209
12	2060	1745	5,35	2222
13	2080	1743	6,2	2235
14	2100	1735	7,67	2252
15	2120	1742	7,93	2264
16	2140	1723	9,47	2270
17	2160	1716	10,99	2288

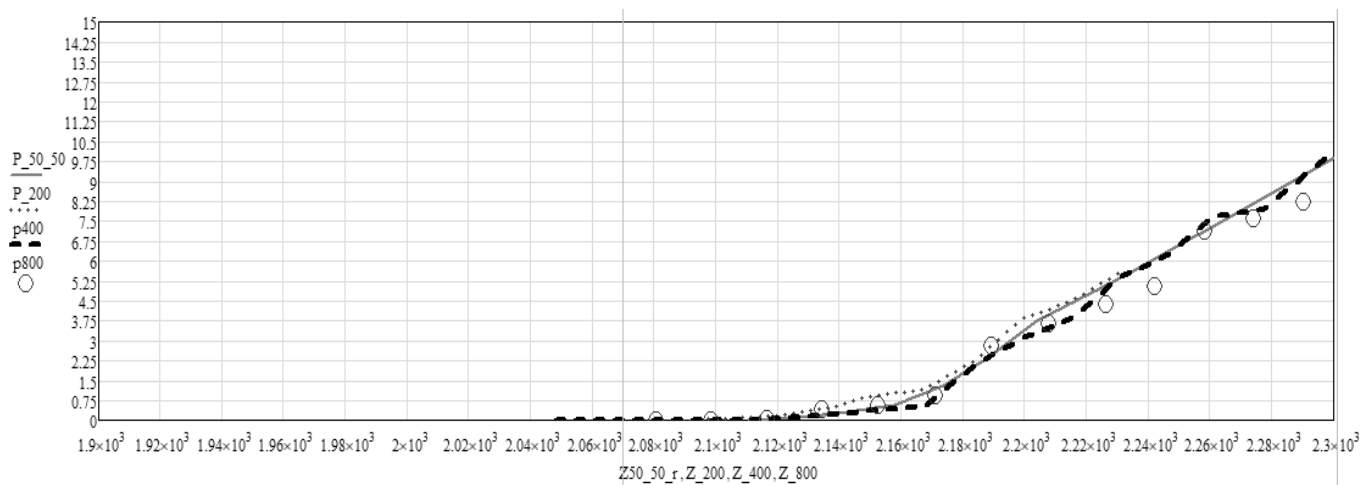


Рис. 5. Зависимость одного фиксированного потока данных в 400 пакетов от потерь

влияния потоков информации между интерфейсами на значение вероятности потерь отдельно взятого потока. Исследуемый маршрутизатор был укомплектован четырьмя портами каждый из которых был переведен в режим передачи данных на скорости 10Мбит/с с полным дуплексом, настройки интерфейса находились в заводских настройках, то есть ни каких специфических дисциплин обслуживания на интерфейсе не было сконфигурировано.

В результате проведенных измерений можно сделать вывод о том, что потоки в рамках маршрутизатора не оказали влияния друг на друга. Минимальная скорость генерации пакетов 163, максимальная 2200 пакетов в секунду, следовательно, за две минуты было сгенерировано достаточно большое количество данных на основании, которых можно говорить об отсутствии взаимного влияния потоков, на скорости моделирования до 10 Мбит/с.

Таблица 4. Три источника с одинаковой скоростью

№	Заданная скорость (пакеты в секунду)	Реальная скорость (пакеты в секунду)	Потери пакетов (%)	Суммарная скорость (пакеты в секунду)
1	746	707	0,53	2132
2	753	712	0,62	2149
3	760	712	1,22	2162
4	767	715	2,3	2194
5	773	713	2,78	2198
6	780	710	3,99	2214
7	787	710	4,64	2228
8	793	704	6,34	2245
9	800	704	7,5	2270
10	807	705	7,9	2282
11	813	706	8,41	2296
12	820	705	9,33	2312
13	827	702	10,09	2318
14	833	697	11,41	2329
15	840	704	11,43	2353
16	847	695	12,93	2352
17	853	697	13,44	2372

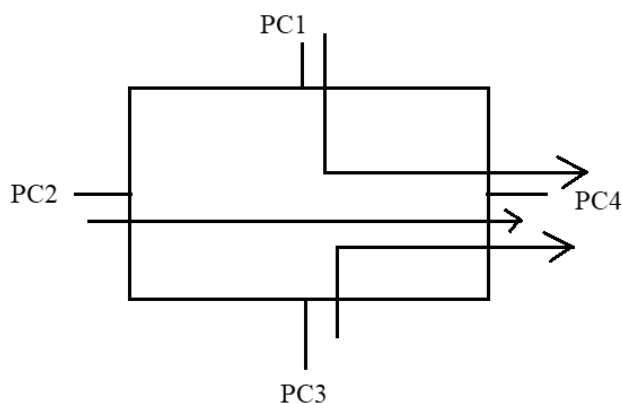


Рис. 6. Топология сети с тремя одинаковыми потоками из PC1, PC2, PC3 на PC4

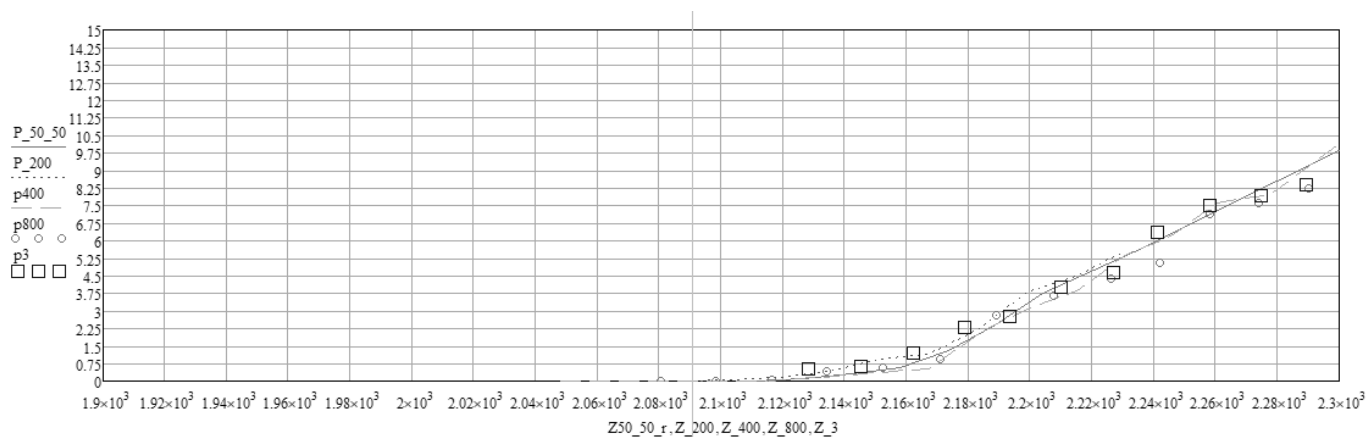


Рис. 7. Зависимость трёх одинаковых потоков данных

ЛИТЕРАТУРА

1. Гаврилов, М.В. Информатика и информационные технологии: учебник для прикладного бакалавриата / М.В. Гаврилов; под редакцией В.А. Климов. — Люберцы: Юрайт, 2016. — 383 с.
2. Гагарина, Л.Г. Информационные технологии: учебное пособие / Л.Г. Гагарина; под редакцией Я.О. Теплова, Е.Л. Румянцева. — Москва: Форум, 2018. — 144 с.
3. Гохберг, Г.С. Информационные технологии: учебное пособие / Г.С. Гохберг. — Москва: Академия, 2018. — 474 с.
4. Дарков, А.В. Информационные технологии: теоретические основы: учебное пособие / А.В. Дарков; под редакцией Н.Н. Шапошников. — Санкт-Петербург: Лань, 2016. — 448 с.
5. Емельянов, С.В. Информационные технологии и вычислительные системы / С.В. Емельянов. — Москва: Ленанд, 2015. — 96 с.
6. Жук, Ю.А. Информационные технологии: мультимедиа: учебное пособие / Ю.А. Жук. — Санкт-Петербург: Лань, 2018. — 210 с.
7. Затонский, А.В. Информационные технологии: разработка информационной модели: учебное пособие / А.В. Затонский. — Москва: Риор, 2017. — 384 с.
8. Коломейченко, А.С. Информационные технологии: учебное пособие / А.С. Коломейченко; под редакцией Н.В. Польшакова, О.В. Чеха. — Санкт-Петербург: Лань, 2018. — 228 с.
9. Коноплева, И.А. Информационные технологии. / И.А. Коноплева; под редакцией О.А. Хохлова, А.В. Денисов. — М.: Проспект, 2015. — 328 с.

© Пугач Алексей Витальевич (dark.penek@mail.ru),

Степанова Дарья Семёновна (052299@bk.ru), Гаипов Константин Эдуардович (gaipovke@yandex.ru).

Журнал «Современная наука: актуальные проблемы теории и практики»

