

АНАЛИЗ ОСОБЕННОСТЕЙ СЕТЕВОГО ТРАФИКА В МАГИСТРАЛЬНОМ КАНАЛЕ: «НУЛЕВОЙ КАНАЛ»

*Вал. В. Иванов^а, Викт. В. Иванов^{а,б,1}, А. В. Крянев^{а,б},
А. С. Приказчикова^б, И. И. Татаринев^{а,в}*

^а Объединенный институт ядерных исследований, Дубна

^б Национальный исследовательский ядерный университет «МИФИ», Москва

^в АО «Лаборатория Касперского», Москва

Ранее нами было показано, что при агрегации измерений сетевого трафика, начиная с некоторого окна агрегации, формируется стабильное распределение, которое с высокой точностью аппроксимируется логнормальным законом. При этом, несмотря на то, что в целом измерения сетевого трафика следуют данному закону, в области малых величин интенсивности трафика и скорости передачи сетевых пакетов наблюдается заметное расхождение между указанным законом и наблюдаемыми измерениями. В настоящей работе обсуждается проблема формирования наблюдаемого пика, возникающего в области малых интенсивностей и скоростей при агрегировании измерений сетевого трафика. Принимая во внимание местоположение рассматриваемого пика, мы назвали его «нулевой канал».

Earlier, we showed that when aggregating network traffic measurements, starting from a certain aggregation window, a stable distribution is formed, which is approximated with a high accuracy by a lognormal law. Wherein, despite the fact that, in general, measurements of network traffic follow this law, in the region of small values of traffic intensity and transmission rate of network packets, there is a noticeable discrepancy between this law and the observed measurements. In this paper, we discuss the problem of the formation of the observed peak arising in the low-intensity region during aggregation of network traffic measurements. Taking into account the location of the peak in question, we called it the “zero channel”.

PACS: 2.50.-r; 05.45.Tp; 07.05.Bx; 89.70.Eg

ВВЕДЕНИЕ

Передача данных в компьютерных сетях, хотя, на первый взгляд, и носит случайный характер, подчиняется определенным законам [1]. Изучение этих законов, понимание причин их возникновения, разработка методов для их применения открывает новые возможности по оптимизации работы компьютерных систем, обнаружению ошибок при передаче данных, защите от вредоносных воздействий на компьютерные сети и т. д.

¹E-mail: ivanov@jinr.ru

Как показал анализ сетевого трафика в магистральном канале [2], проведенный нами в работах [3] и [4], плотность вероятности распределения (ПВР) интенсивности сетевого трафика (рис. 1) и ПВР скорости передачи сетевых пакетов (рис. 2) с высокой точностью описываются логнормальным законом [5]:

$$f(x) = \frac{a}{\sqrt{2\pi}\sigma} \frac{1}{x} \exp \left[-\frac{1}{2\sigma^2} (\ln x - \mu)^2 \right], \quad (1)$$

где x — переменная; σ и μ — параметры распределения; a — нормировочный множитель.

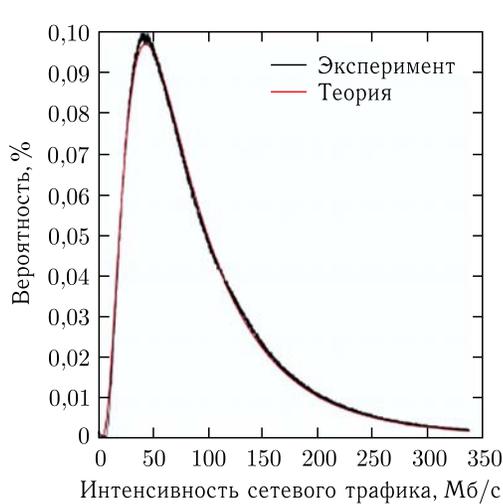


Рис. 1. Аппроксимация ПВР интенсивности сетевого трафика логнормальным законом (размер окна агрегации 10 мс)

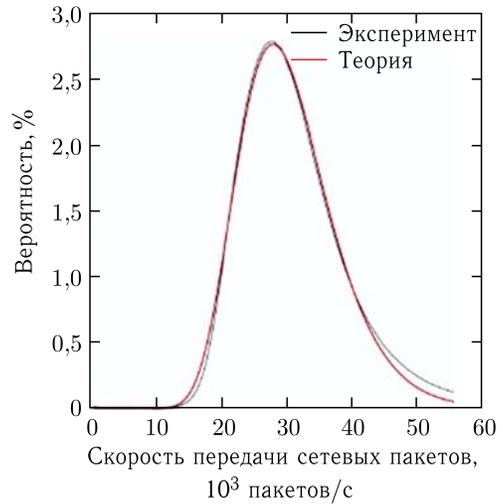


Рис. 2. Аппроксимация ПВР скорости передачи сетевых пакетов логнормальным законом (размер окна агрегации 10 мс)

В работе [3] нами анализировался сетевой трафик, переданный по магистральному каналу в течение 15 мин, в следующей работе [4] — за 48 ч, включая как дневные, так и ночные интервалы суток. В обоих случаях ПВР интенсивности сетевого трафика и ПВР скорости передачи сетевых пакетов с высокой точностью описываются логнормальным законом как для разных режимов передачи сетевых пакетов (включающих симплексный режим, при котором сетевые пакеты передаются только от одного МАК-адреса ко второму, или от второго к первому, и дуплексный режим, при котором сетевые пакеты передаются как одновременно и от одного МАК-адреса ко второму и от второго к первому), так и для разных типов сетевых пакетов (включая служебные пакеты и ТСП-пакеты) и различных размеров окна агрегации (от 1 мс до 60 с).

Несмотря на то, что в целом ПВР интенсивности сетевого трафика и ПВР скорости передачи сетевых пакетов с высокой точностью описываются логнормальным законом, в области малых величин интенсивности сетевого трафика и скорости передачи сетевых пакетов наблюдается заметное расхождение между логнормальным законом и наблюдаемыми данными. Эта особенность отмечалась нами в работе [6] и обсуждалась позднее в работе [3].

1. «НУЛЕВОЙ КАНАЛ»

С учетом местоположения и группировки отсчетов в области малых величин интенсивности сетевого трафика и скорости передачи сетевых пакетов указанная особенность названа нами «нулевым каналом».

Под «нулевым каналом» (НК) понимается самый левый канал ПВР, содержащий события (случай регистрации сетевых пакетов) в диапазоне $[0 \dots w_x]$ (где w_x — ширина канала). Для ПВР интенсивности сетевого трафика нулевой канал — это канал, содержащий события в диапазоне от 0 до w_x мбит/с (например, от 0 до 1 мбит/с, где ширина канала равна 1 мбит/с).

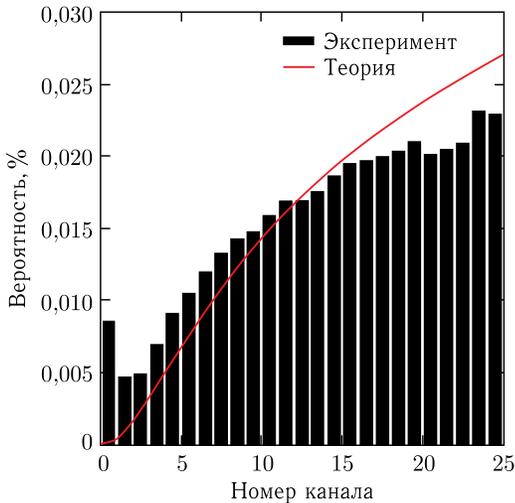


Рис. 3. Аппроксимация ПВР интенсивности сетевого трафика логнормальным законом в области малых интенсивностей (размер окна агрегации 10 мс, ширина канала 1 Мбит/с)

Наибольшее расхождение между теоретическими (логнормальный закон) и реальными (регистрируемые ПВР) данными наблюдается в самом первом канале в диапазоне $[0 \dots w_x]$ (рис. 3).

В зависимости от того, на основании каких данных строится ПВР интенсивности сетевого трафика или ПВР скорости передачи сетевых пакетов, учитывающих режим сбора сетевых пакетов, тип сетевых пакетов, размер окна агрегации и т. д., вклад НК в общее распределение может колебаться от пренебрежимо малых величин до достаточно заметных. При этом НК при аппроксимации экспериментальных данных может оказывать заметное влияние на параметры логнормального распределения.

В этой связи необходимо разобраться с тем, как указанные данные влияют на изучаемый нами процесс и как правильно обрабатывать данные, сгруппированные в области НК.

Понимание причин, приводящих к формированию НК в ПВР интенсивности сетевого трафика и ПВР скорости передачи сетевых пакетов, позволит понять, как учитывать вклад НК при анализе регистрируемого трафика, а также как аппроксимировать указанные ПВР теоретическим законом.

Кроме того, при анализе вклада НК можно также понять, какому закону подчиняются ПВР интенсивности сетевого трафика и ПВР скорости передачи сетевых пакетов. С этой целью ниже будут рассмотрены несколько наиболее вероятных распределений, которые могут отвечать за формирование НК.

Наблюдение выделяющегося НК, т. е. ситуации, когда количество событий в диапазоне $[0 \dots w_x]$ (где w_x — ширина канала) значительно превышает количество событий в соседних диапазонах $[w_x \dots n \times w_x]$ (где n — количество каналов, расположенных в зоне малой интенсивности), может быть связано с тем, что в сетевом трафике:

1) наблюдается заметное количество последовательных сетевых пакетов, временные интервалы между которыми больше окна агрегации ПВР и значительно превосходят величину среднего интервала между последовательными сетевыми пакетами:

$$N(\Delta t > w_a) \gg 1;$$

2) присутствует последовательность сетевых пакетов (даже при относительно малом их количестве), временной интервал между которыми значительно превосходит размер окна агрегации:

$$\Delta t \gg w_a,$$

где w_a — размер окна агрегации.

В упомянутых нами работах [3,4] наблюдаются описанные случаи как для объема переданных данных (рис. 4), так и для количества сетевых пакетов (рис. 5).

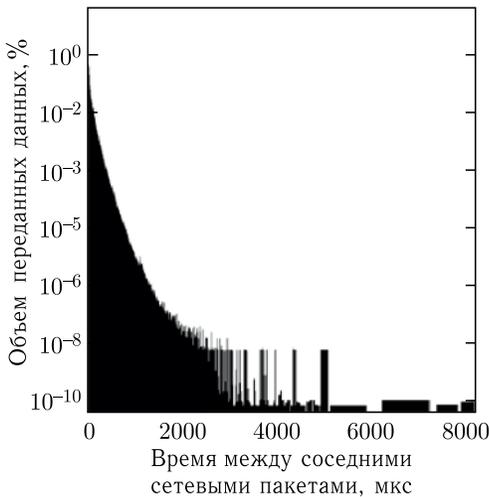


Рис. 4. ПВР объема переданных данных в зависимости от величины временного интервала между последовательными сетевыми пакетами

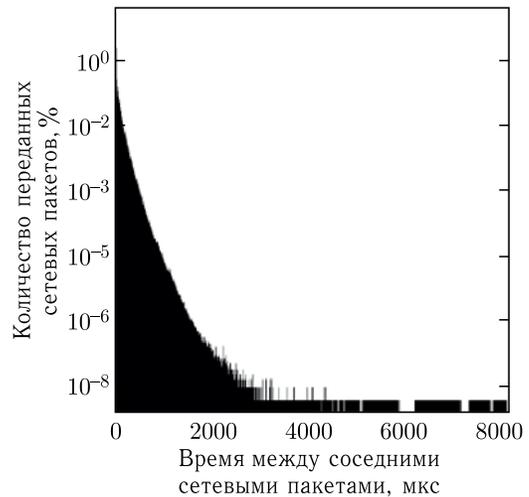


Рис. 5. ПВР количества сетевых пакетов в зависимости от величины временного интервала между последовательными сетевыми пакетами

В проанализированном нами сетевом трафике были зарегистрированы:

- последовательные сетевые пакеты, между которыми зафиксированы большие временные интервалы (более 8000 мкс, при этом среднее значение временного интервала $\overline{\Delta t}$ между пакетами составило 83 мкс);
- последовательные сетевые пакеты, временной интервал между которыми превышает размер окна агрегации (около 0,001 %).

2. АНАЛИЗ РАСПРЕДЕЛЕНИЯ ПАКЕТОВ СЕТЕВОГО ТРАФИКА

Рассмотрим, как ПВР временных интервалов между последовательными сетевыми пакетами влияет на ПВР интенсивности сетевого трафика и на ПВР скорости передачи сетевых пакетов.

Пусть $f(t)$ — ПВР временных интервалов между последовательными сетевыми пакетами, а w_a — размер окна агрегации. Тогда ПВР сетевых пакетов $g(n)$ в указанном окне может быть вычислена с использованием следующего выражения:

$$g(n) = \begin{cases} \int_{w_a}^{\infty} f(t) dt, & n = 0, \\ \int_0^{w_a} \int_0^{w_a - t_1} \dots \int_0^{w_a - t_1 - \dots - t_{n-1}} f(t_1) \dots f(t_n) dt_1 dt_2 \dots dt_n, & n > 0. \end{cases} \quad (2)$$

ПВР интенсивности сетевого трафика и ПВР скорости передачи сетевых пакетов можно представить в следующем виде:

$$h(n) \sim \frac{g(n)}{w_a}.$$

При этом, на основании (2), вклад НК в общее распределение будет определяться из следующего соотношения:

$$k = \frac{g(0)}{1 - g(0)} = \frac{\int_{w_a}^{\infty} f(t) dt}{1 - \int_{w_a}^{\infty} f(t) dt}. \quad (3)$$

Из выражения (3) можно сделать вывод о том, что распределение временных интервалов с «тяжелыми» хвостами может приводить к тому, что в ПВР интенсивности сетевого трафика или в ПВР скорости передачи сетевых пакетов вклад в НК будет заметным по сравнению с вкладами в другие каналы.

В качестве примера рассмотрим некоторые статистические распределения, которые могут быть ответственными за формирование НК:

- нормальное распределение — распределение, описывающее многие случайные процессы [7];
- логнормальное распределение — распределение [8], описывающее ПВР интенсивности сетевого трафика [3];
- распределение Парето — распределение, описывающее, в частности, размеры файлов в интернет-трафике по протоколу TCP [9].

2.1. Нормальное распределение:

$$f(t) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(t - \mu)^2}{2\sigma^2}\right), \quad (4)$$

$$g(0) = \operatorname{erf}\left(\frac{t - \mu}{\sqrt{2}\sigma}\right)\Big|_{w_a}^{\infty} = 1 - \left(\frac{w_a - \mu}{\sqrt{2}\sigma}\right), \quad k = \frac{\sqrt{2}\sigma}{w_a - \mu} - 1,$$

здесь $\operatorname{erf}(\cdot)$ — функция ошибок [10].

Из выражений (4) видно, что для того чтобы при нормальном распределении временных интервалов между сетевыми пакетами пик НК был ярко выражен, требуется, чтобы $w_a \approx \mu$.

Однако для сетевого трафика в магистральном канале для $w_a = 10$ мс имеет место соотношение $w_a \gg \mu = 83$ мкс [2]. Это, в свою очередь, приводит к тому, что величина вклада k слабо зависит от μ . Поэтому наблюдение пика в области НК практически невозможно.

2.2. Логнормальное распределение:

$$f(t) = \frac{1}{\sqrt{2\pi}\sigma t} \exp - \frac{(\ln t - \mu)^2}{2\sigma^2},$$

$$g(0) = \operatorname{erf} \left(\frac{\ln t - \mu}{\sqrt{2}\sigma} \right) \Big|_{w_a}^{\infty} = 1 - \left(\frac{\ln w_a - \mu}{\sqrt{2}\sigma} \right), \quad k = \frac{\sqrt{2}\sigma}{\ln w_a - \mu} - 1. \quad (5)$$

Из выражений (5) видно, что для того чтобы при логнормальном распределении временных интервалов между сетевыми пакетами пик НК был ярко выражен, требуется выполнение более слабого, чем в случае нормального распределения, условия, а именно: $\ln w_a \approx \mu$. Данное условие заметно слабее (по сравнению с условием $w_a \gg \mu$) ограничивает вклад k в область НК. Указанное обстоятельство делает возможным наблюдение пика в НК.

2.3. Распределение Парето:

$$f(t) = \frac{1}{t^a}, \quad a > 1,$$

$$g(0) = \frac{1}{(1-a)t^{a-1}} \Big|_{w_a}^{\infty} = \frac{1}{(a-1)w_a^{a-1}}, \quad k = \frac{1}{1 - (a-1)w_a^{a-1}}. \quad (6)$$

Из выражений (6) видно, что для того чтобы при распределении по закону Парето временных интервалов между сетевыми пакетами проявился пик в области НК, не требуется никаких дополнительных условий, кроме задания окна агрегации. Поэтому даже при небольших значениях параметров a и w_a в области НК должен наблюдаться искомый пик (см. разд. 3).

3. ЧИСЛЕННОЕ МОДЕЛИРОВАНИЕ СЕТЕВОГО ТРАФИКА

Для тестирования приведенных в разд. 2 распределений, на предмет формирования нулевого канала, использовалось численное моделирование суточного сетевого трафика. Моделировалась передача данных в симплексном режиме ТСП сетевых пакетов с фиксированным размером 1600 байт (равным среднему размеру сетевых пакетов в сетевом трафике магистрального канала [2]) и с временными интервалами между последовательными сетевыми пакетами, удовлетворяющими заданным распределениям.

Поскольку размер сетевых пакетов в моделируемом сетевом трафике был фиксированным, то, в отличие от исходного сетевого трафика, он (размер) не оказывал никакого влияния на итоговую ПВР интенсивности моделируемого трафика. Поэтому

при анализе вклада распределения в НК рассматривалась только ПВР скорости передачи сетевых пакетов. При этом ПВР интенсивности сетевого трафика имела ровно такой же вид.

3.1. Нормальное распределение. Для нормального распределения (4) временных интервалов между соседними сетевыми пакетами коэффициенты σ , μ выбирались таким образом, чтобы максимум распределения соответствовал средней скорости сетевых пакетов в исходном трафике.

Как прогнозировалось выше в п. 2.1, НК не должен был наблюдаться ни в окрестности соседних каналов (рис. 6), ни в области всего распределения ПВР (рис. 7). Из приведенных рисунков видно, что при нормальном распределении временных интервалов между пакетами пик в области НК отсутствует.

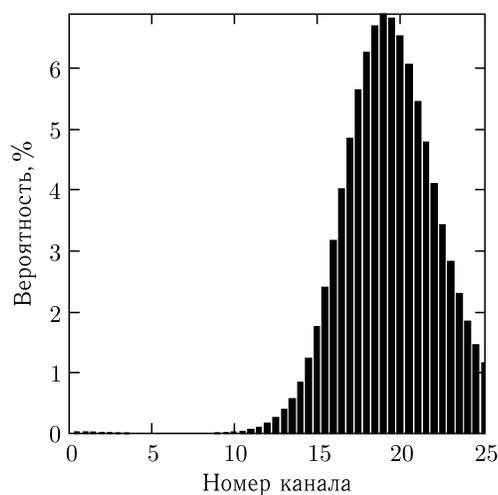


Рис. 6. Моделирование ПВР скорости передачи сетевых пакетов в области малых скоростей (размер окна агрегации 10 мс, ширина канала 1000 пакетов/с) для нормального распределения временных интервалов между соседними пакетами

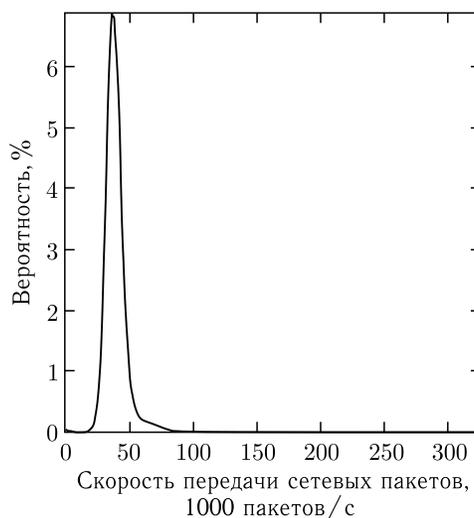


Рис. 7. Моделирование ПВР скорости передачи сетевых пакетов в широком диапазоне скоростей (размер окна агрегации 10 мс, ширина канала 1000 пакетов/с) для нормального распределения временных интервалов между соседними пакетами

Поскольку НК не наблюдается, а сам вид используемого при моделировании распределения временных интервалов между соседними пакетами (нормальное распределение, симметричное относительно средней скорости передачи сетевых пакетов) отличается от ПВР скорости передачи сетевых пакетов, полученной из анализа экспериментальных данных [3], можно сделать вывод о том, что реальное распределение временных интервалов между соседними пакетами не следует нормальному закону.

3.2. Логнормальное распределение. Для логнормального распределения (5) временных интервалов между соседними сетевыми пакетами коэффициенты σ и μ выбирались таким образом, чтобы максимум распределения соответствовал средней скорости сетевых пакетов в исходном трафике.

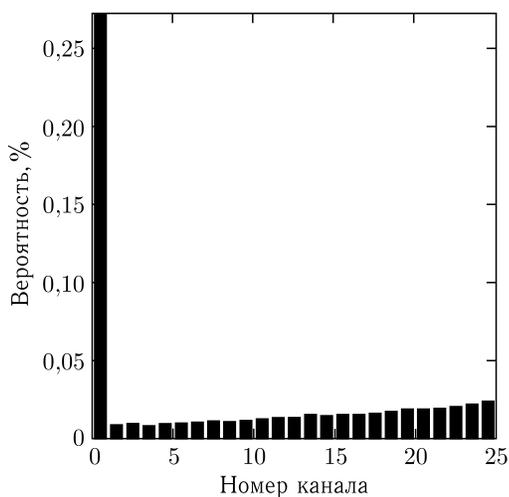


Рис. 8. Моделирование ПВР скорости передачи сетевых пакетов в области малых скоростей (размер окна агрегации 10 мс, ширина канала 1000 пакетов/с) для логнормального распределения временных интервалов между соседними пакетами

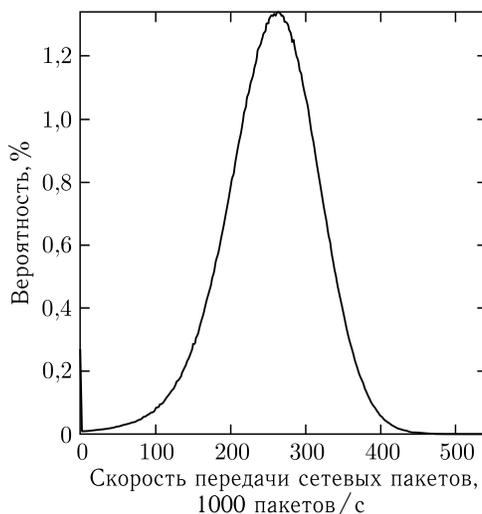


Рис. 9. Моделирование ПВР скорости передачи сетевых пакетов в широком диапазоне скоростей (размер окна агрегации 10 мс, ширина канала 1000 пакетов/с) для логнормального распределения временных интервалов между соседними пакетами

Как прогнозировалось ранее в п. 2.2, НК отчетливо наблюдается в окрестности соседних каналов (рис. 8), а также на фоне всего распределения ПВР (рис. 9).

Таким образом, мы видим, что при логнормальном распределении временных интервалов между сетевыми пакетами пик в области НК четко выделяется.

В этом случае при аппроксимации экспериментальных данных логнормальным распределением (с использованием, в частности, метода наименьших квадратов [emulation]) отчетливо выраженный пик НК может оказывать заметное влияние на основные параметры ($\sigma_{\text{аппрок}}$, $\mu_{\text{аппрок}}$) исследуемого распределения. Это, в свою очередь, отражается на средних характеристиках изучаемой ПВР таких, как интенсивность сетевого трафика, скорость передачи сетевых пакетов.

3.3. Распределение Парето. Для распределения временных интервалов между соседними сетевыми пакетами согласно закону Парето (6) выбиралось такое значение коэффициента a , при котором пик распределения соответствовал средней скорости передачи сетевых пакетов в исходном трафике.

Как прогнозировалось ранее в п. 2.3, в этом случае также наблюдается пик НК (хотя не так отчетливо, как это имело место для логнормального распределения) как в области соседних каналов (рис. 10), так и на фоне всего ПВР распределения (рис. 11).

При этом следует заметить, что, несмотря на то, что с помощью распределения Парето удастся воспроизвести пик в области НК, у данного распределения, в отличие от ПВР скорости передачи сетевых пакетов для реальных данных и логнормального распределения, полностью отсутствуют отсчеты в области малых скоростей.

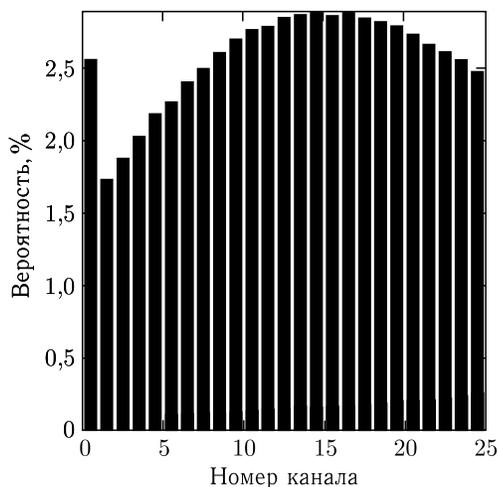


Рис. 10. Моделирование ПВР скорости передачи сетевых пакетов в области малых скоростей (размер окна агрегации 10 мс, ширина канала 1000 пакетов/с) для распределения временных интервалов между последовательными сетевыми пакетами по закону Парето



Рис. 11. Моделирование ПВР скорости передачи сетевых пакетов в широком диапазоне скоростей (размер окна агрегации 10 мс, ширина канала 1000 пакетов/с) для распределения временных интервалов между последовательными сетевыми пакетами по закону Парето

Отмеченная особенность позволяет нам сделать вывод о том, что реальное распределение временных интервалов между соседними пакетами, вероятнее всего, не следует распределению Парето.

Таким образом, можно сделать вывод о том, что из трех рассмотренных нами статистических распределений для описания распределения временных интервалов между соседними пакетами наилучшим образом подходит логнормальное распределение, для которого прежде чем аппроксимировать экспериментальные данные, нужно исключить вклад, связанный с НК.

4. ВЛИЯНИЕ РАЗМЕРА ОКНА АГРЕГАЦИИ НА «НУЛЕВОЙ КАНАЛ»

Как отмечалось в разд. 2, вклад НК в анализируемое статистическое распределение может зависеть от размера окна агрегации. В частности, в случае логнормального распределения временных интервалов между соседними сетевыми пакетами вклад НК обратно пропорционален логарифму размера окна агрегации: см. выражения в п. 2.2.

Для сравнения прогноза (5) с измерениями трафика в магистральном канале [3] вычислялись зависимости вкладов НК в ПВР интенсивности сетевого трафика и ПВР скорости передачи сетевых пакетов с использованием следующего выражения:

$$\eta = \frac{h(i)}{H}, \quad i = 0, 1, \dots, N,$$

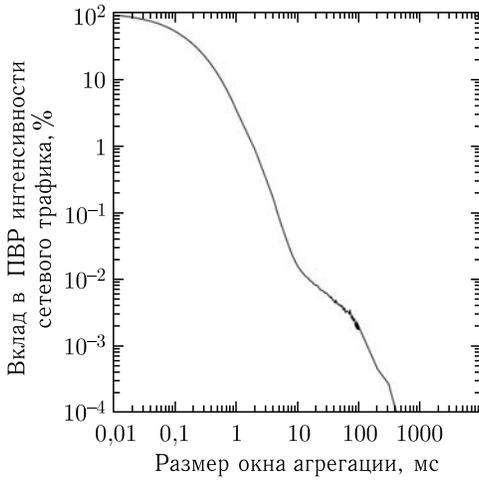


Рис. 12. Зависимость вклада НК в ПВР интенсивности сетевого трафика от размера окна агрегации для пакетов, переданных в симплексном режиме

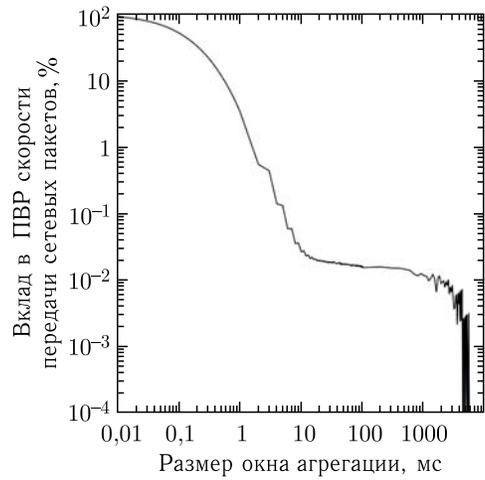


Рис. 13. Зависимость вклада НК в ПВР скорости передачи сетевых пакетов от размера окна агрегации для пакетов, переданных в симплексном режиме

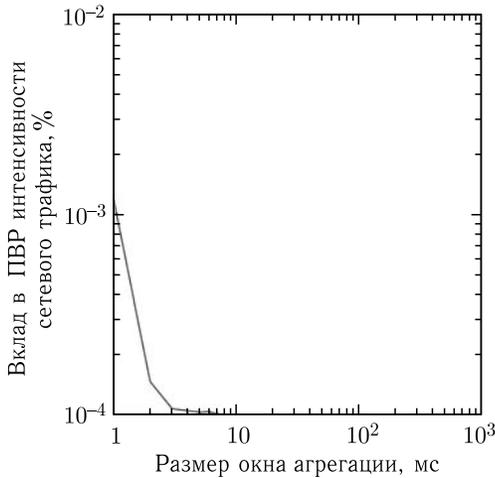


Рис. 14. Зависимость вклада НК в ПВР интенсивности сетевого трафика от размера окна агрегации для всех пакетов, переданных в симплексном режиме

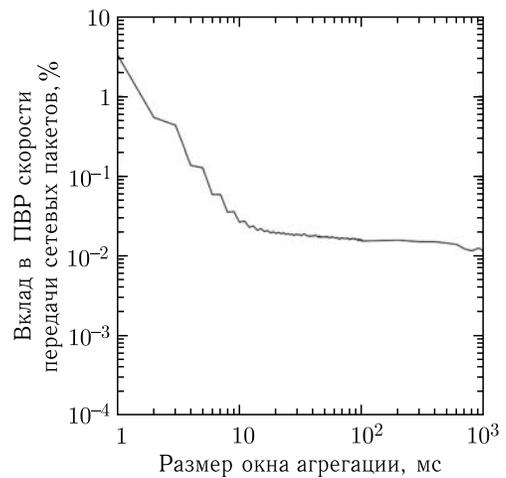


Рис. 15. Зависимость вклада НК в ПВР скорости передачи сетевых пакетов от размера окна агрегации для TCP-пакетов, переданных в симплексном режиме

где $h(i)$ — содержимое текущего окна агрегации; $H = \sum_{i=1}^N h_i$ — сумма по всем окнам агрегации, содержимое которых отличается от нуля.

На рис. 12 и 13 приведены зависимости вкладов НК в ПВР интенсивности сетевого трафика и ПВР скорости передачи сетевых пакетов от размера окна агрегации.

Из приведенных рисунков отчетливо видно, что основной вклад НК формируется в области малых значений окна агрегации (до 10 мс). При этом величина вклада быстро спадает (примерно обратно пропорционально размеру окна агрегации), достигая нулевых значений. Такое поведение указанной зависимости объясняется причинами, рассмотренными в разд. 2.

Детальный анализ измерений сетевого трафика [3] показывает, что вне зависимости от режима передачи данных и типа сетевых пакетов НК наблюдается во всех ПВР интенсивности сетевого трафика и ПВР скорости передачи сетевых пакетов. Однако, в зависимости от упомянутых выше параметров, вклад НК в ПВР может колебаться от тысячных долей процента (рис. 14) до целых процентов (рис. 15).

ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ И ВЫВОДЫ

В настоящей работе обсуждается проблема формирования наблюдаемого пика, возникающего в области малых интенсивностей при агрегировании измерений сетевого трафика.

Результаты проделанной работы можно резюмировать следующим образом:

- расхождения между теоретическими и экспериментальными значениями ПВР интенсивности сетевого трафика и ПВР скорости передачи сетевых пакетов в области НК обусловлены последовательными сетевыми пакетами, временной интервал между которыми значительно превышает его среднее значение;
- в результате проведенного численного моделирования показано, что ПВР интенсивности сетевого трафика и ПВР скорости передачи сетевых пакетов наилучшим образом согласуются с логнормальным законом;
- показано, что основной вклад НК формируется в области малых значений окна агрегации, при этом вклад НК быстро спадает, достигая нулевых значений;
- с учетом того, что вклад НК может оказывать серьезное влияние на результаты анализа измерений сетевого трафика, для корректной аппроксимации данных наблюдения логнормальным законом НК нужно исключать из набора экспериментальных данных.

СПИСОК ЛИТЕРАТУРЫ

1. Antoninou I., Ivanov V. V., Ivanov Valery V., Zrellov P. V. On the Log-Normal Distribution of Network Traffic // *Physica D*. 2002. V. 167. P. 72–85.
2. MAWI Working Group Traffic Archive. <http://mawi.wide.ad.jp/mawi/>.
3. Zrellov P. V., Ivanov Valery V., Ivanov Victor V., Kryukov Yu. A., Tatarinov I. I. Study of Internet-Traffic Features in the Trunk Channel // *Phys. Part. Nucl. Lett.* 2019. V. 16, No. 3. P. 289–299.
4. Иванов Валерий В., Иванов Виктор В., Крянев А. В., Татаринов И. И. Аппроксимация измерений интернет-трафика в магистральном канале суммой логнормальных распределений // *Вестн. МИФИ*. 2019. Т. 8, № 4. С. 380–394.
5. Eadie W. T., Dryard D., James F. E., Roos M., Sadoulet B. *Statistical Methods in Experimental Physics*. Amsterdam; London: North-Holland, 1971.

6. Крюков Ю. А., Чернягин Д. В. Исследование самоподобия трафика высокоскоростного канала передачи пакетных данных // III Междунар. науч. конф. «Современные проблемы информатизации в системах моделирования, программирования и телекоммуникациях». М., 2009; <http://econf.rae.ru/article/4819>.
7. Wasserman L. All of Statistics. New York, NY: Springer, 2004. P. 142.
8. Lognormal Distributions, Theory and Applications / Ed.: Crow E. L., Shimizu Kunio. 1988. V. 88. Statistics: Textbooks and Monographs. New York: Marcel Dekker, Inc., xvi+387p.
9. Guerriero V. Power Law Distribution: Method of Multi-Scale Inferential Statistics // J. Mod. Math. Frontier (JMMF). 2012. V. 1. P. 21–28.
10. Press W. H., Teukolsky S. A., Vetterling W. T., Flannery B. P. Sec. 6.2. Incomplete Gamma Function and Error Function // Numerical Recipes: The Art of Scientific Computing. 3rd ed. New York: Cambridge Univ. Press, 2007.

Получено 11 января 2021 г.