

Статистический анализ сетевого трафика

Петров В.В.

МЭИ, ИРЭ, Москва, Красноказарменная 13,

v_petrov@skc.ru

Декабрь 2003

ОГЛАВЛЕНИЕ

| | |
|---|-----------|
| Используемые сокращения | 2 |
| Введение..... | 2 |
| 1. Реализации сетевого трафика..... | 4 |
| 1.1. BC-Oct89Ext.TL | 4 |
| 1.2. LBL-ПКТ-5.TCP | 6 |
| 1.3. LBL-TCP-3 | 8 |
| 2. Формирование временных рядов. | 10 |
| 1.1. Процедура агрегирования..... | 10 |
| 2.4. Тестовые реализации | 17 |
| 2.5. Логарифмированные реализации | 18 |
| 3. Классический анализ..... | 20 |
| 3.1. Плотности распределения | 21 |
| 3.2. Автокорреляционные функции..... | 23 |
| 3.3. Энергетические спектры..... | 27 |
| 4. Исследование показателя Хэрста реализаций..... | 29 |
| 5. Исследование реализаций методами нелинейной динамики | 31 |
| 5.1. Концепция суррогатных данных | 32 |
| 5.2. Идея реконструкции аттрактора | 34 |
| 5.3. Ложные ближайшие соседи (FNN)..... | 34 |
| 5.4. Вычисление корреляционного интеграла | 36 |
| 5.5. Проверка гипотезы о статистической независимости (BDS-тест)..... | 40 |
| 6. Исследование перепадов в реализациях сетевого трафика | 41 |
| 7. Анализ полученных результатов и основные выводы..... | 42 |
| Литература | 46 |

Используемые сокращения

| | |
|------|--------------------------------------|
| БУЗ | Быстро убывающая зависимость |
| МУЗ | Медленно убывающая зависимость |
| АКФ | Автокорреляционная функция |
| ФГШ | Фрактальный гауссовский шум |
| ССШС | Строгое самоподобие в широком смысле |
| СУС | Самоподобие в узком смысле |
| СКО | Среднеквадратическое отклонение |
| БД | Броуновское движение |
| ФБД | Фрактальное броуновское движение |

Введение

Многочисленные исследования трафика компьютерной сети свидетельствуют о том, что он обладает свойством самоподобия [1]. При этом методы расчета компьютерной сети (пропускной способности каналов, емкости буферов и пр.) основанные на марковских моделях и формулах Эрланга, которые с успехом используются при проектировании телефонных сетей, дают неоправданно оптимистические решения и приводят к недооценке нагрузки. Кроме того, самоподобный трафик имеет особую структуру, сохраняющуюся на многих масштабах - в реализации всегда присутствует некоторое количество очень больших выбросов при относительно небольшом среднем уровне трафика. Данное явление значительно ухудшает характеристики (увеличивает потери, задержки, джиттер пакетов) при прохождении самоподобного трафика через узлы сети. В этом смысле силы разработчиков, комитетов по стандартизации в области телекоммуникаций должны быть направлены на изучение особенностей трафика и формирование новых алгоритмов для его наиболее оптимальной обработки.

Целью настоящей работы является исследование доступными методами структуры реального сетевого трафика, направленное на выявление его характерных особенностей. По мнению автора, данный анализ должен позволить подойти вплотную к проблеме увеличения эффективности обработки самоподобного трафика.

Работа построена следующим образом:

- в первой главе рассматриваются три различные реализации реального сетевого трафика, представленные в [1]. Реализации выбирались с тем расчетом, чтобы по возможности исключить влияния человеческого фактора, особенностей анализируемой сети, протоколов и пр. Две реализации (*LBL-PKT-5.TCP* и *LBL-TCP-3*) были получены при анализе трафика одного протокола (TCP) на одной и той же сети (Lawrence Berkeley Laboratory) в разное время (28 января и 20 января 1994 года

соответственно) одним и тем же человеком (*Vern Paxson*). Третья реализация представляет собой трафик протокола другого уровня (протокол Ethernet) и получена другими авторами (*Will Leland* и *Dan Wilson*) на другой сети (*Bellcore Morristown Research and Engineering facility*). Показывается, тем не менее, что данные реализации обладают рядом схожих особенностей. В главе изучаются экспериментальные плотности распределения (гистограммы) длительностей единиц передачи информации (кадров и дэйтаграмм);

- во второй главе подробно описывается процедура агрегирования и анализируются возникающие при этом ошибки. Одновременно вводятся некоторые хорошо изученные тестовые реализации (хаотическая и белый шум), которые далее анализируются наравне с реализациями сетевого трафика для упрощения идентификации результатов анализа последнего;
- в главе 3 производится классический анализ: оцениваются дисперсии, средние, АКФ, энергетические спектры, плотности распределения изучаемых временных рядов;
- в главе 4 оценивается коэффициент Херста 7-ю доступными методами;
- в пятой главе реализации изучаются с помощью алгоритмов нелинейной динамики. В частности изучаются суррогатные данные, ложные ближайшие соседи, вычисляется корреляционный интеграл, а также BDS-статистика;
- выводы и обсуждение результатов исследования делаются в заключительной части.

Ценность настоящей работы состоит в том, что, пожалуй, впервые производится столь обширное исследование сетевого трафика с позиций различных методов и алгоритмов анализа временных рядов. Особый интерес представляет исследование сетевого трафика с точки зрения теории нелинейной динамики.

В исследовании использовались свободно распространяемые версии программных реализаций популярных алгоритмов [6], [25], [26], [10]. Некоторые используемые процедуры (агрегирования реализаций, регрессионного анализа АКФ, исследования перепадов в сетевом трафике) создавались автором самостоятельно.

1. Реализации сетевого трафика

Для дальнейшего анализа выберем известные реализации сетевого трафика, которые представлены в Internet traffic archive [1].

1.1. BC-Oct89Ext.TL

Первая из изучаемых нами реализаций была получена *Will Leland* и *Dan Wilson*. Она названа авторами *BC-Oct89Ext.TL* и представляет собой зависимость от времени размеров сетевых кадров протокола второго уровня модели OSI (Ethernet-кадров в данном случае), проходящих через вычислительную сеть *Bellcore Morristown Research and Engineering facility* в глобальную сеть Internet. Реализация исследуется в работах [2] и [3].

Данные представлены двумя колонками чисел в ASCII формате. Первый столбец представляет собой время в секундах. Второй – длину поля данных Ethernet-кадра в байтах (не включая преамбулу, заголовка и контрольной суммы CRC). Заметим, что существует как минимум четыре основных разновидности кадров Ethernet:

- Ethernet Type II
- Ethernet 802.3
- Ethernet 802.2
- Ethernet SNA

Так как, авторы не уточняют спецификацию кадра, который использовался при записи трафика, напомним общий формат кадров Ethernet (рис. 1.1.)

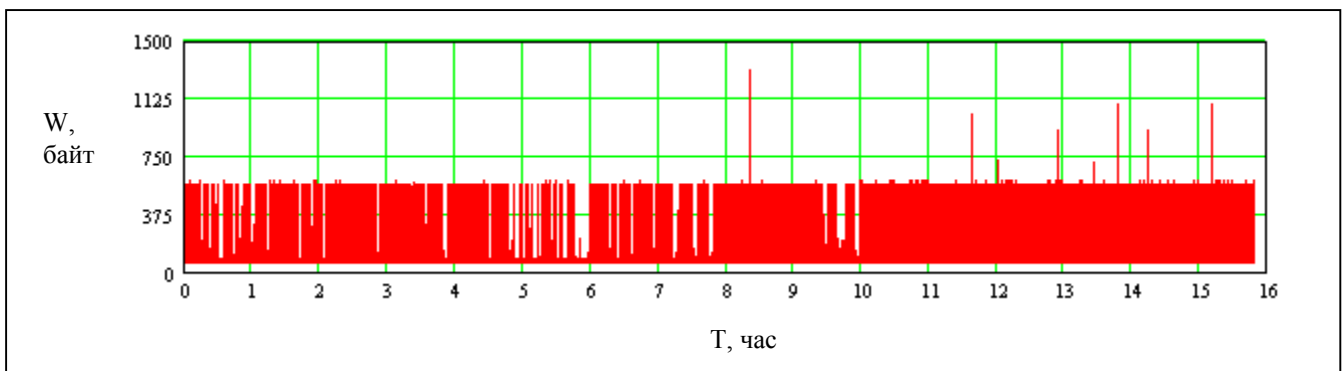


Рис. 1.1. Общий формат кадра Ethernet.

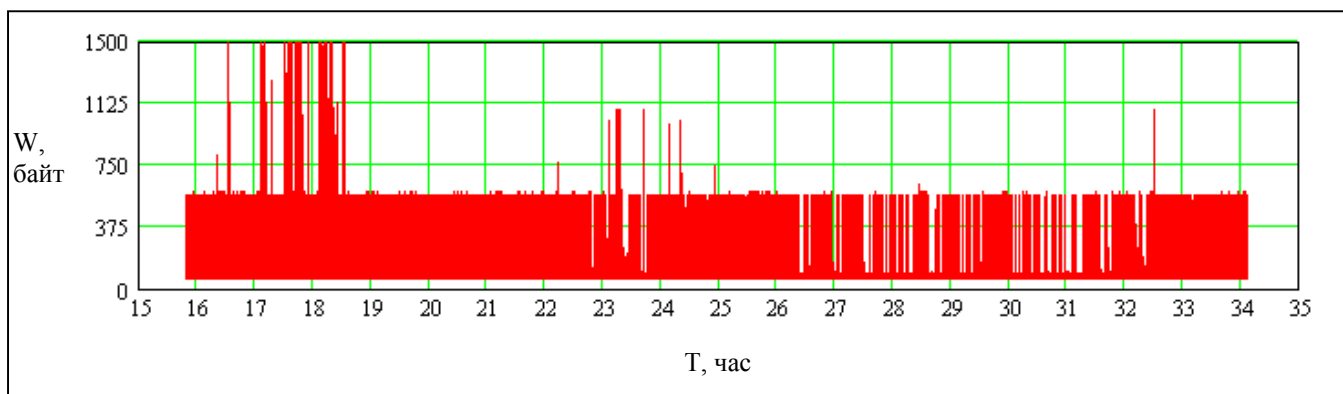
Реализация *BC-Oct89Ext.TL* начинается 3 октября 1989 года (день недели – вторник) в 23 час. 44 мин. по восточноевропейскому времени, заканчивается спустя 122797.83 секунд

(34,11 часов) и содержит 1 миллион записей об Ethernet-кадрах, проходящих в это время между сетью корпорации Bellcore и сетью Internet.

Исследователи утверждают, что аппаратура измерения не отбрасывала каких-либо Ethernet-кадров, проходящих через нее. Тем не менее, испорченные пакеты (возникающие, например, при коллизиях) в реализацию не включались. Кроме того, 99.5% инкапсулированных в Ethernet-кадры пакетов относятся к протоколу IP. По данным первоисточника точность временных отсчетов составляет порядка 10 микросекунд. Заметим, что отсчеты не эквидистантны, а соответствуют моментам прихода кадров. То есть для дальнейшего анализа реализацию необходимо будет привести к эквидистантной шкале по оси времени (это делается в главе 2). График реализации *BC-Oct89Ext.TL* представлен на рис 1.2: здесь по оси абсцисс отложены временные отсчеты “Т”, а по оси ординат – длительности “W” Ethernet-кадров.



а



б

Рис. 1.2. Зависимость длин кадров реализации *BC-Oct89Ext.TL* от времени. а) первые 16 часов, б) оставшиеся 18 часов.

Минимальная и максимальная длительность кадра в данной реализации составляет 64 байта и 1 515 байт соответственно. Для наглядности построим распределение длительностей кадров – гистограмму (см. рис. 1.3).

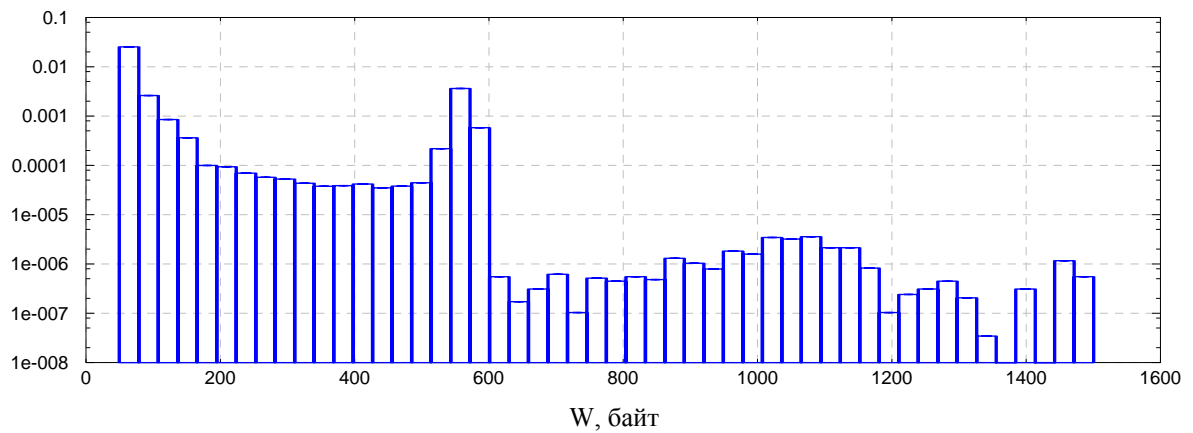


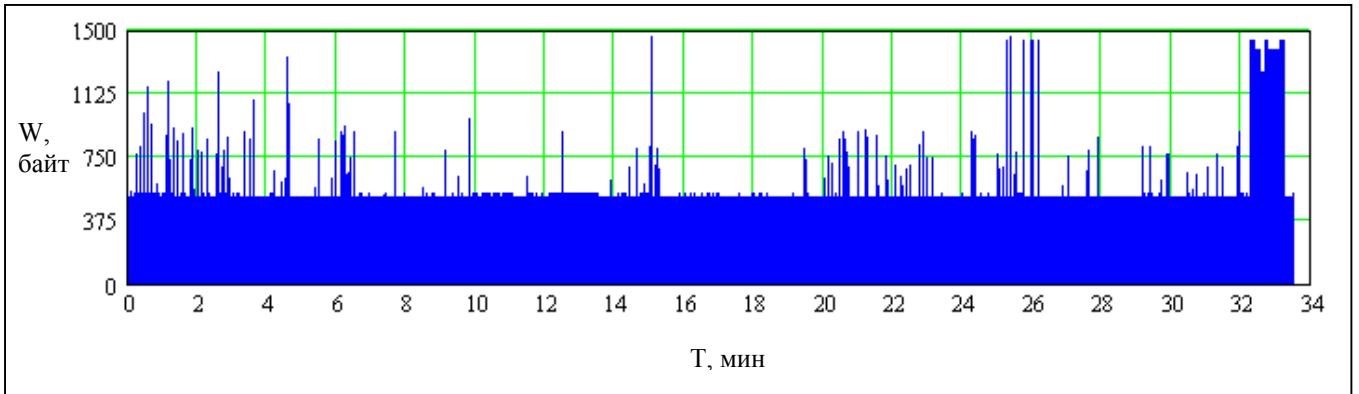
Рис. 1.3. Распределение длительностей Ethernet-кадров реализации *BC-Oct89Ext.TL*

Визуальный анализ гистограммы показывает, что большинство кадров имеют длительности порядка 64 и 570 байт. Может оказаться интересным, что частота появления кадров, длительность которых $W > 600$ байт существенно ниже (на два порядка), чем частота появления кадров, обладающих $W < 600$ байт.

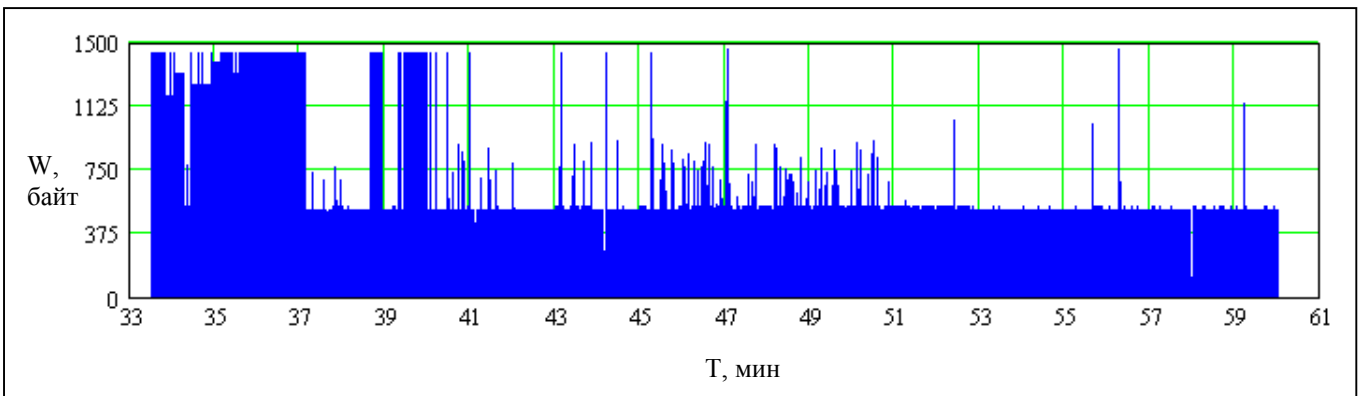
1.2. LBL-PKT-5.TCP

Следующей реализацией, которую мы будем исследовать, будет реализация *LBL-PKT-5.TCP*, полученную *Vern Paxson* и изучаемую, например в [4]. Также как предыдущая, она представляет собой зависимость от времени размеров дэйтаграмм протокола четвертого уровня модели OSI (TCP в данном случае), проходящих между сетью *Lawrence Berkeley Laboratory*, расположенной в *Berkeley, California*, и глобальной сетью *Internet*.

Данные представлены шестью колонками чисел в ASCII формате, из которых нас будут интересовать только первый (время в секундах) и шестой столбцы (длительность дэйтаграмм в байтах). Реализация записывалась 28 января 1994 года (пятница) с 14 час. 00 мин. до 15. час. 00 мин (по Тихоокеанскому поясному времени) и содержит 677 846 отсчетов, сделанных через неодинаковые интервалы времени, точность которых – 1 микросекунда. График реализации *BC-Oct89Ext.TL* представлен на рис 1.4. Минимальная и максимальная длительность дэйтаграмм в данной реализации составляет 0 байт (пакеты "ping-ack") и 1 460 байт соответственно. Распределение длительностей дэйтаграмм приводится на рис. 1.5.



а



б

Рис. 1.4. Зависимость длин кадров реализации *LBL-PKT-5.TCP* от времени. а) первые 33 минуты, б) оставшиеся 27 минут.

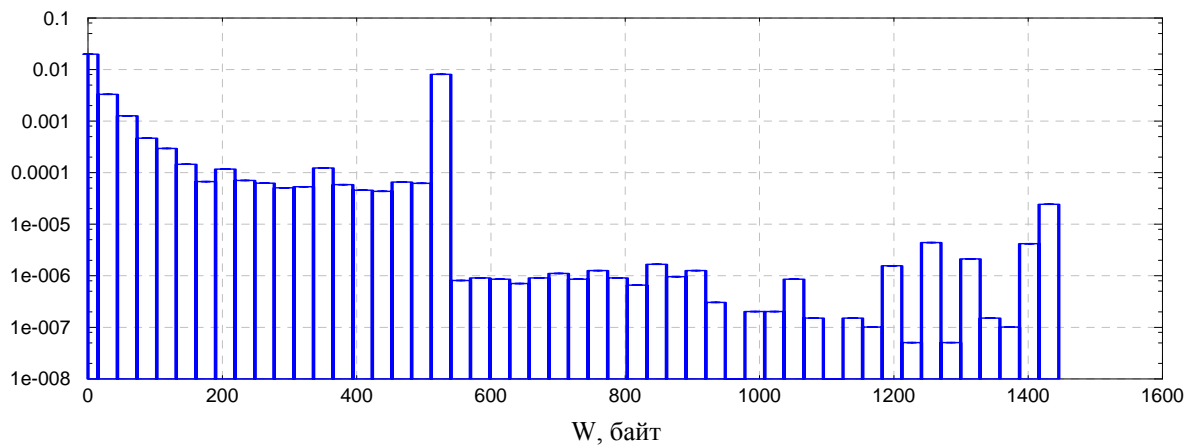
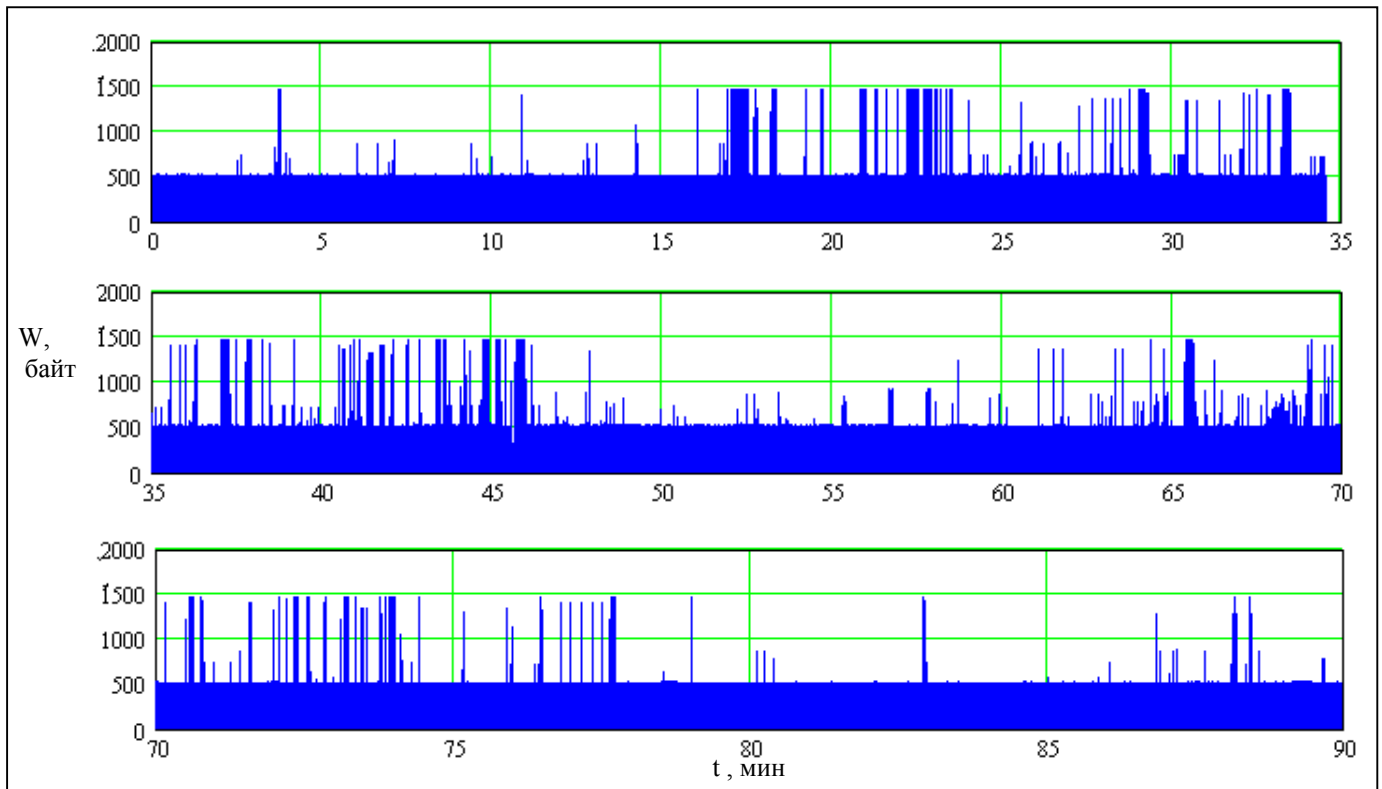


Рис. 1.5. Распределение длительностей дэйтаграмм реализации *LBL-PKT-5.TCP*

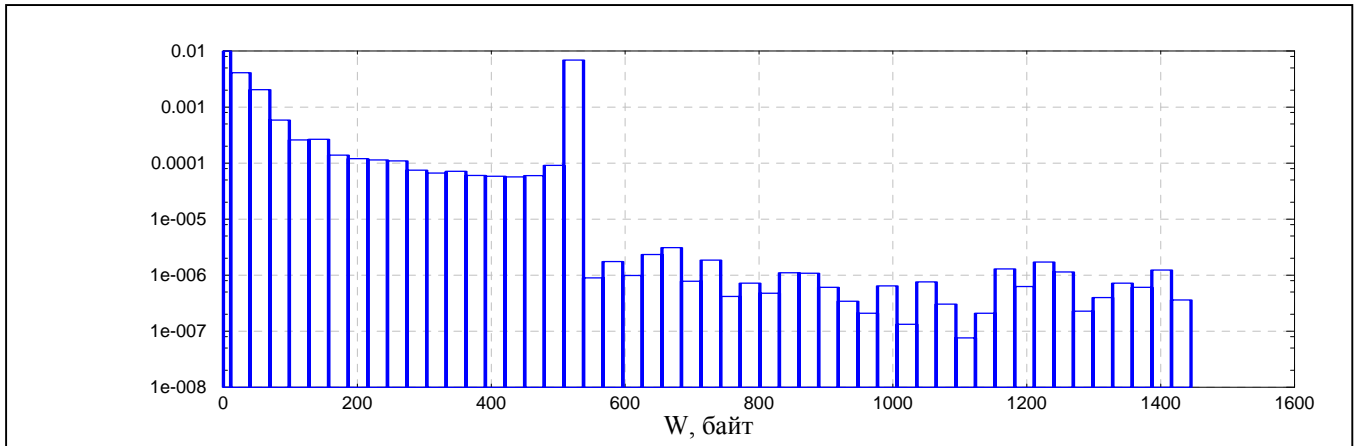
Изучая гистограмму реализации *LBL-PKT-5.TCP* можно отметить, большинство дэйтаграмм имеют длительности 0 байт и 512 байт. Кроме того, наблюдается характерный спад частот появления дэйтаграмм, длительность которых превышает значение 512 байт.

1.3. LBL-TCP-3

Возьмем еще одну реализацию *LBL-TCP-3*. Эта реализация аналогична *LBL-PKT-5.TCP* и представляет собой запись TCP-трафика между сетью Internet и локальной вычислительной сетью Lawrence Berkeley Laboratory. Трафик наблюдался Vern Paxson 20 января 1994 (четверг) с 14 час. 10 мин. до 16 час. 10 мин. Соответствующий временной ряд содержит информацию об 1.3 миллионе TCP-дэйтаграмм и представлен, также как и предыдущий, шестью столбцами чисел: временные отсчеты, хост-источник, хост-получатель, TCP-порт источника, TCP-порт получателя и количество байт в дэйтаграмме (ноль для "rueg-ask" – пакетов). По информации первоисточника [1] временные отсчеты имеют точность – 1 микросекунда. Данная реализация изучается в [4]. График реализации и соответствующее распределение длительностей TCP-дэйтаграмм представлены на рис. 1.6.



а



а

Рис. 1.6 а) Зависимость количества байт от времени в реализации *LBL-TCP-3*
 б) Распределение длительностей дэйтаграмм реализации *LBL-TCP-3*.

Сравнивая гистограмму реализации *LBL-TCP-3* (рис. 1.6) с гистограммами рассмотренных выше реализаций (рис. 1.5 и рис. 1.3) можно заметить, что все они обладают схожими структурными особенностями. А именно, наблюдается плавное уменьшение частот появления дэйтаграмм, длительности которых находится в интервале 0 ...512 байт, далее следует характерный всплеск в районе 512 байт практически на 2 порядка, после чего - резкое уменьшение (на 4 порядка) частот появления дэйтаграмм, длительность которых >512 байт (> 600 для Ethernet-трафика).

Так как дэйтаграммы протокола TCP инкапсулируются в Ethernet-кадры, можно сделать вывод, что распределение длин Ethernet-кадров определяются распределением TCP-дэйтаграмм, а не наоборот. Интересно заметить, что распределение Ethernet-кадров наследует особенности распределения TCP-дэйтаграмм, несмотря на то, что они относятся к разным реализациям.

2. Формирование временных рядов.

2.1. Процедура агрегирования

Как уже упоминалось ранее, процедура агрегирования необходима для приведения исходных реализации к эквидистантной (с постоянным шагом ΔT) шкале по оси времени. При этом величину ΔT будем называть уровнем агрегации. Для этого ось времени разбивается на интервалы ΔT , каждому из которых ставится в соответствие отношение количества прошедшей за данный интервал времени информации (в байтах) к длительности интервала (в секундах).

Допустим $\xi = \{\xi(t_1), \xi(t_2), \dots, \xi(t_n)\}$ - исходная (не эквидистантная) реализация графика, длительностью n отсчетов. Причем $|t_{i+1} - t_i| \neq const$, где $i \in 1, n$. На рис. 2.1 схематично показана процедура агрегирования.

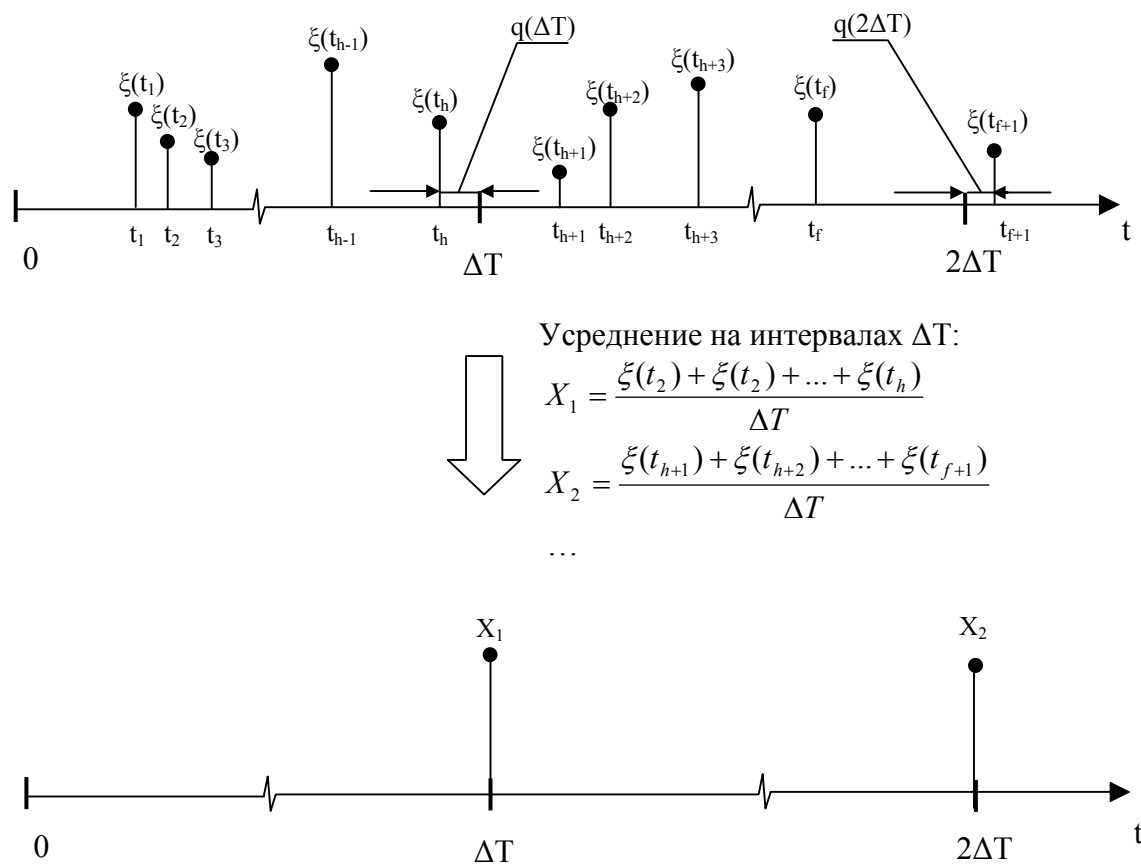
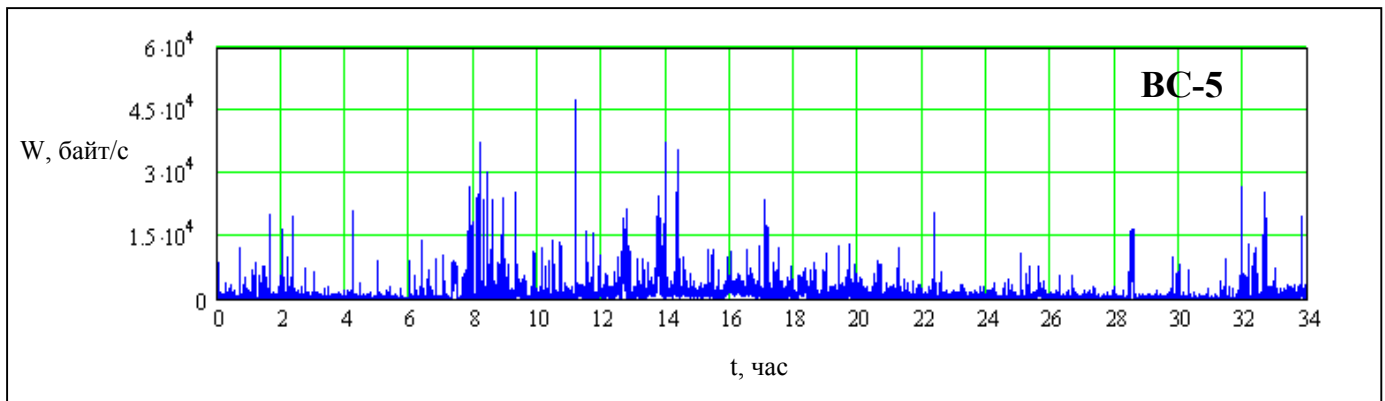


Рис. 2.1 Агрегирование реализации.

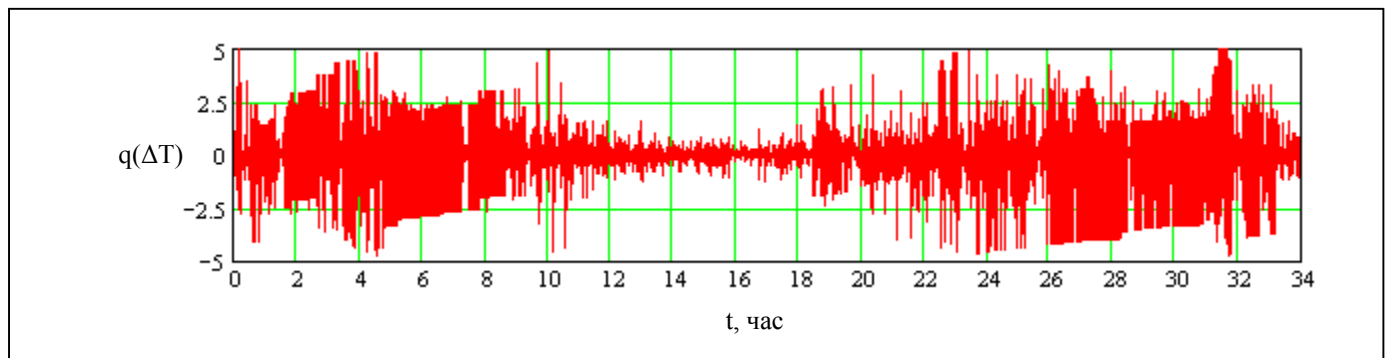
Так как отсчеты $\Delta T, 2\Delta T, \dots, N\Delta T$ чаще всего не совпадают с отсчетами t_i исходной реализации, примем за правило (при усреднении) в качестве последнего элемента, принадлежащего некоторому интервалу ΔT брать ближайший отсчет ξ . То есть к интервалу ΔT - $2\Delta T$ отнесем отсчеты $\xi(t_{h+1}), \xi(t_{h+2}), \dots, \xi(t_{f+1})$. Очевидно, чем меньше значения ошибок

$|q(\Delta T)|, |q(2\Delta T)|, |q(3\Delta T)|, \dots, |q(N\Delta T)|$ тем меньше искажений вносится на данном этапе. Таким образом, получается агрегированная эквидистантная реализация трафика $X = \{X(\Delta T), X(2\Delta T), \dots, X(N \cdot \Delta T)\}$, содержащая N элементов. Физический смысл каждого ее элемента – средняя скорость (байт/сек) на соответствующем интервале ΔT .

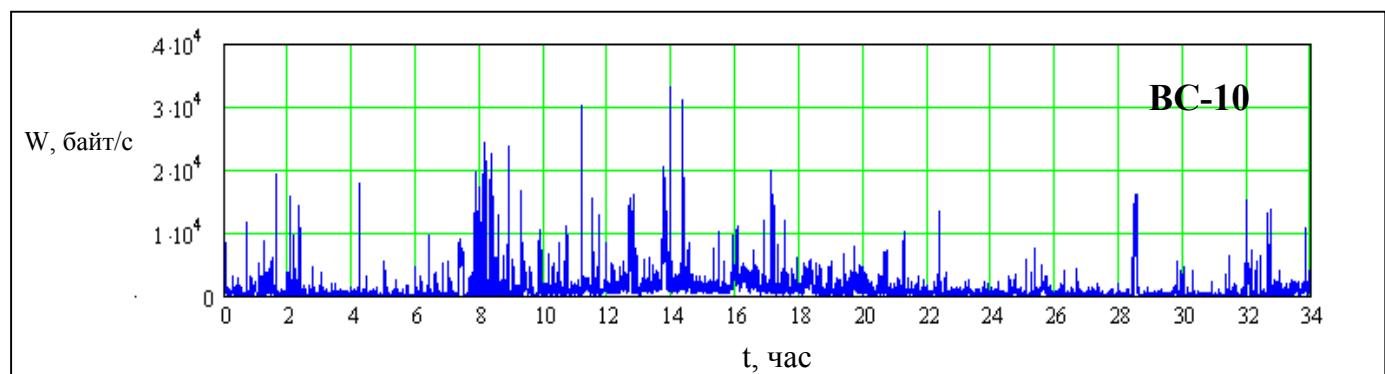
Проведем агрегирование реализации *BC-Oct89Ext.TL* с уровнями агрегации $\Delta T=5с$, $\Delta T=10с$ и $\Delta T=100с$. В дальнейшем изложении будем обозначать получившиеся временные ряды BC-5, BC-10 и BC-100 соответственно. На рис. 2.2 показан внешний вид рядов BC-5, BC-10 и BC-100, а также значения ошибок $q(\Delta T), q(2\Delta T), q(3\Delta T), \dots, q(N\Delta T)$.



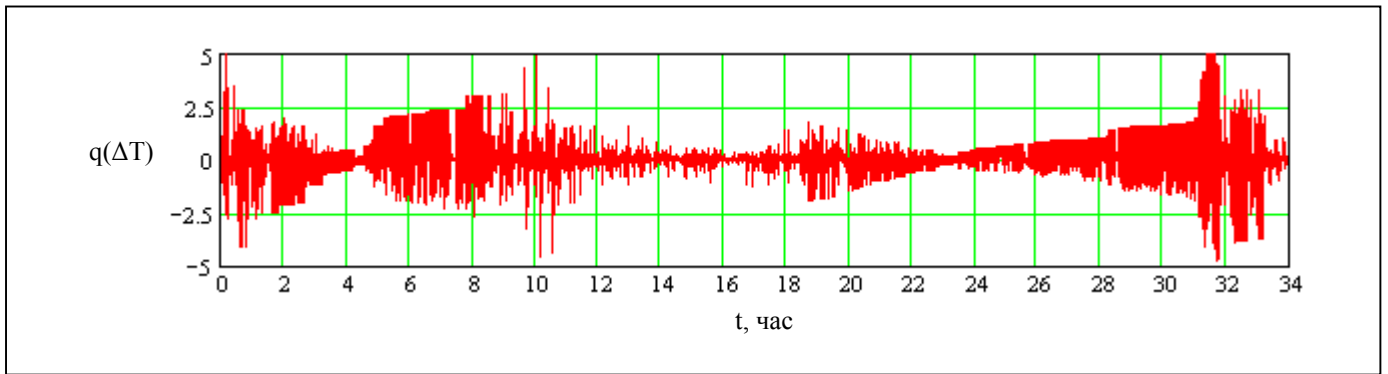
а



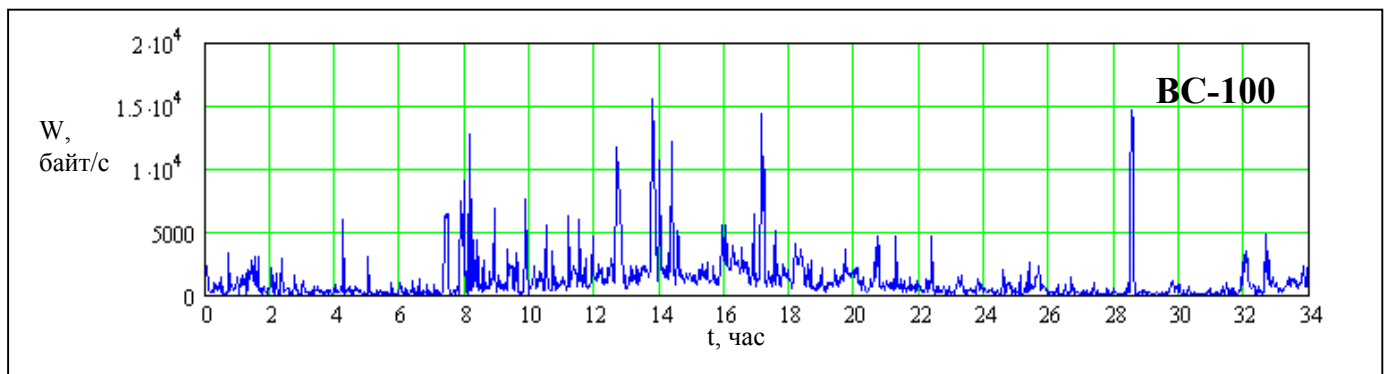
б



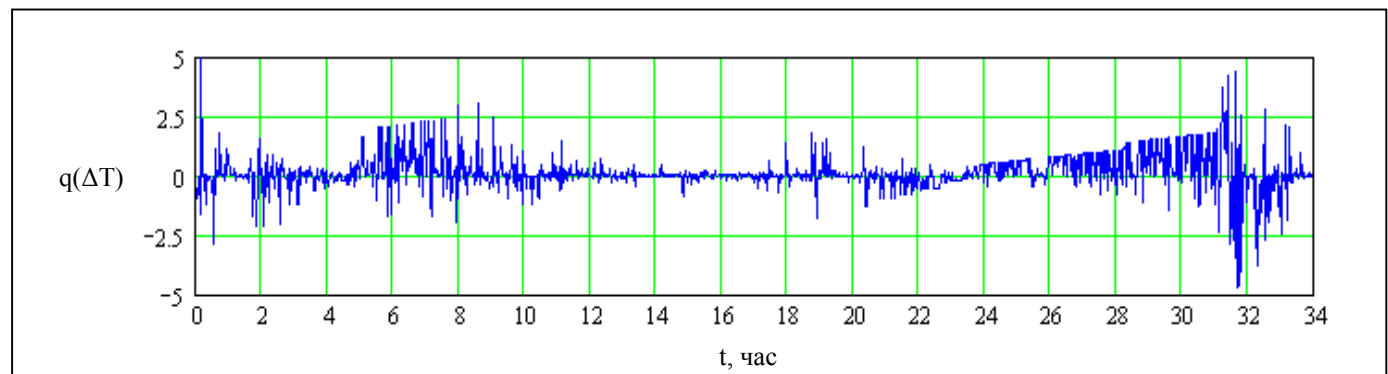
в



г



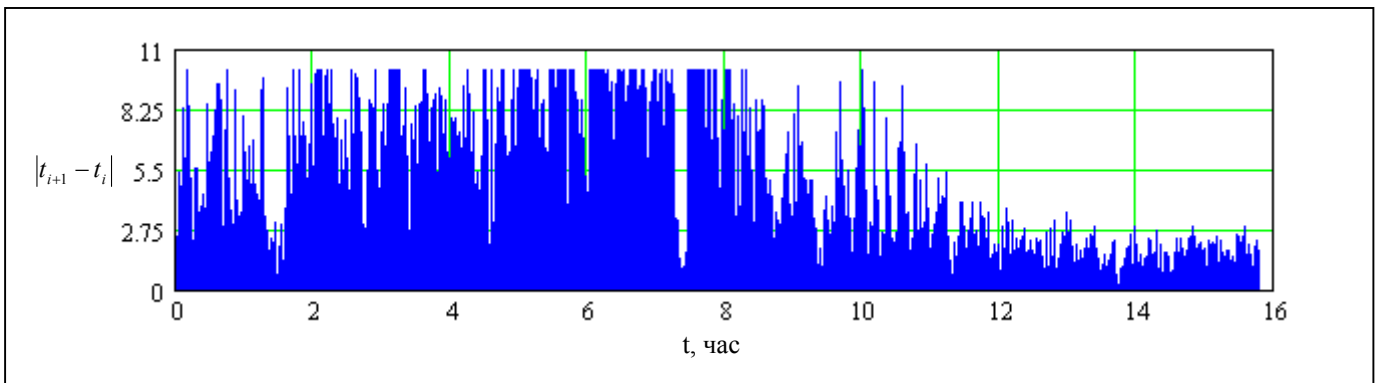
д



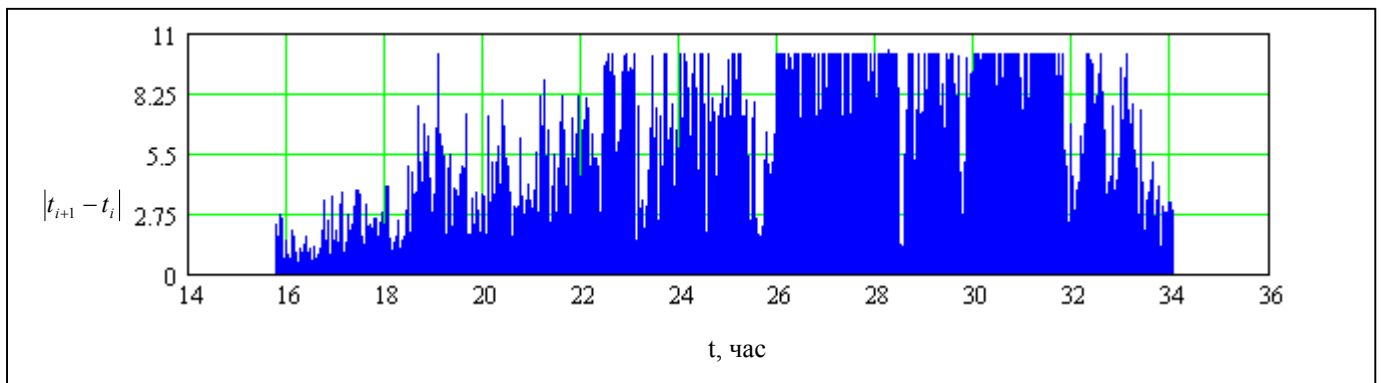
е

Рис. 2.2. а) Временной ряд BC-5;
 б) Ошибки $q(\Delta T)$, $q(2\Delta T)$, $q(3\Delta T)$, ..., $q(N\Delta T)$ при получении ряда BC-5;
 в) Временной ряд BC-10;
 г) Ошибки $q(\Delta T)$, $q(2\Delta T)$, $q(3\Delta T)$, ..., $q(N\Delta T)$ при получении ряда BC-10;
 д) Временной ряд BC-100;
 е) Ошибки $q(\Delta T)$, $q(2\Delta T)$, $q(3\Delta T)$, ..., $q(N\Delta T)$ при получении ряда BC-100;

Еще одной характеристикой, интересной с точки зрения изучения реализации трафика может служить зависимость $|t_{i+1} - t_i|$ от времени, где $i \in 1, n$, (см. рис. 2.1) которая служит наглядным отражением степени неэквидистантности исходной реализации (см. рис. 2.3).



а

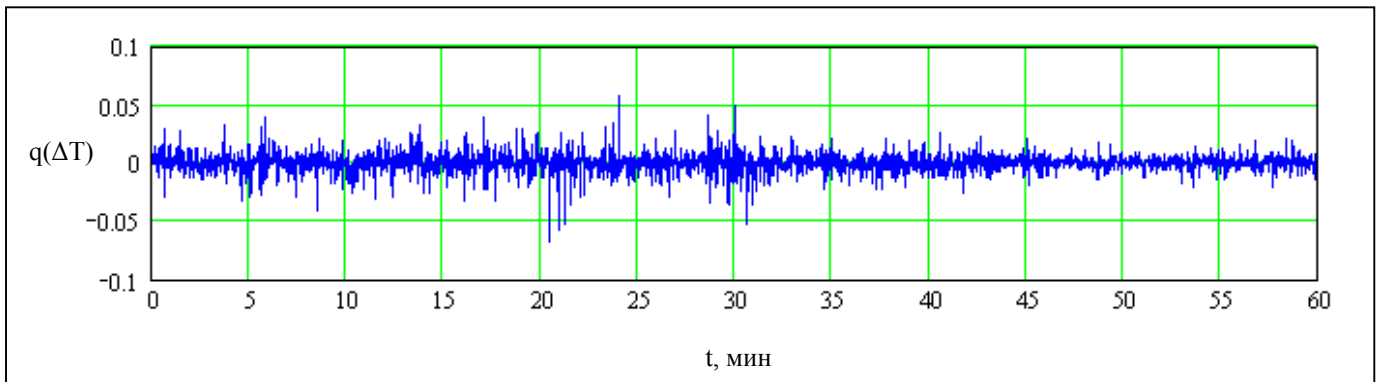
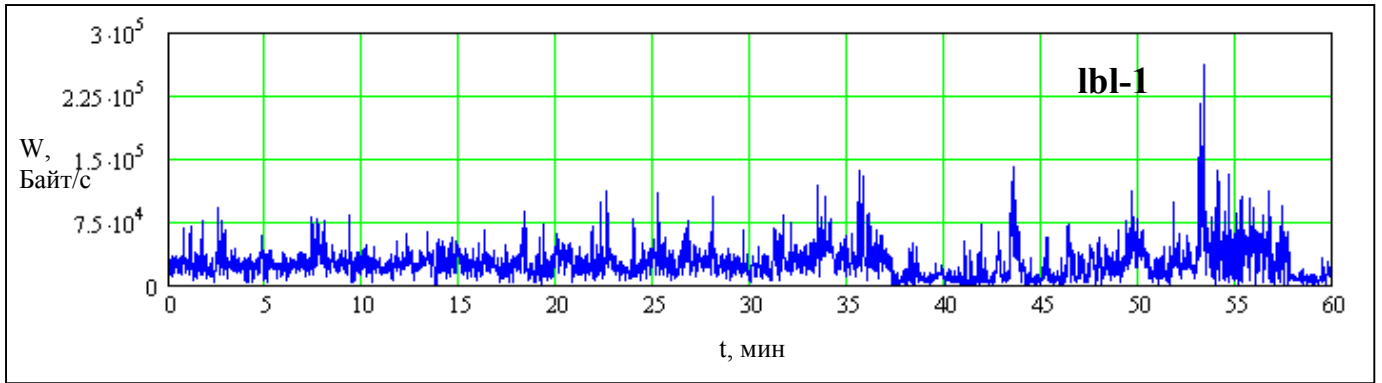
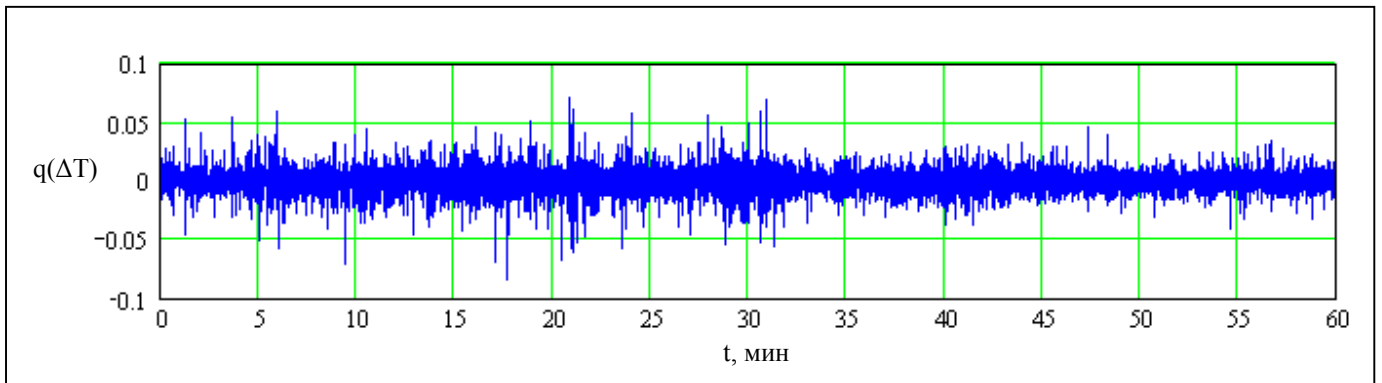
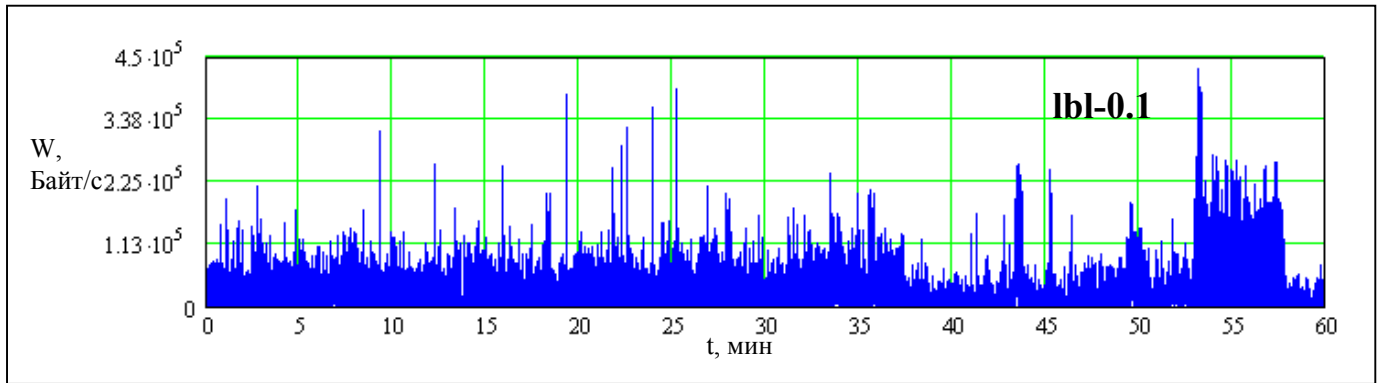


б

Рис. 2.3. Первые разности временных отсчетов исходной реализации *BC-Oct89Ext.TL*

Интересно, что в промежуток примерно с 11:00 до 18:00 наблюдается минимальные значения $|t_{i+1} - t_i|$, которые Ethernet-кадры на данном интервале времени следуют друг за другом максимально плотно. Этот факт может объясняться высокой активностью использования сетевых ресурсов в указанной время суток, что в свою очередь, скорее всего, связано с продолжительностью рабочего дня. Аналогично, на можно отметить, что на данном интервале времени наблюдаются минимальные ошибки (см. рис. 2.2) при агрегировании. Очевидно, последнее является следствием высокой плотности следования кадров.

Произведем агрегирование реализации *LBL-PKT-5.TCP* с уровнями $\Delta T=0.1$ с и $\Delta T=1$ с и $\Delta T=2$ с. Получившиеся ряды назовем *lbl-0.1*, *lbl-1*, и *lbl-2* соответственно (см. рис. 2.4).



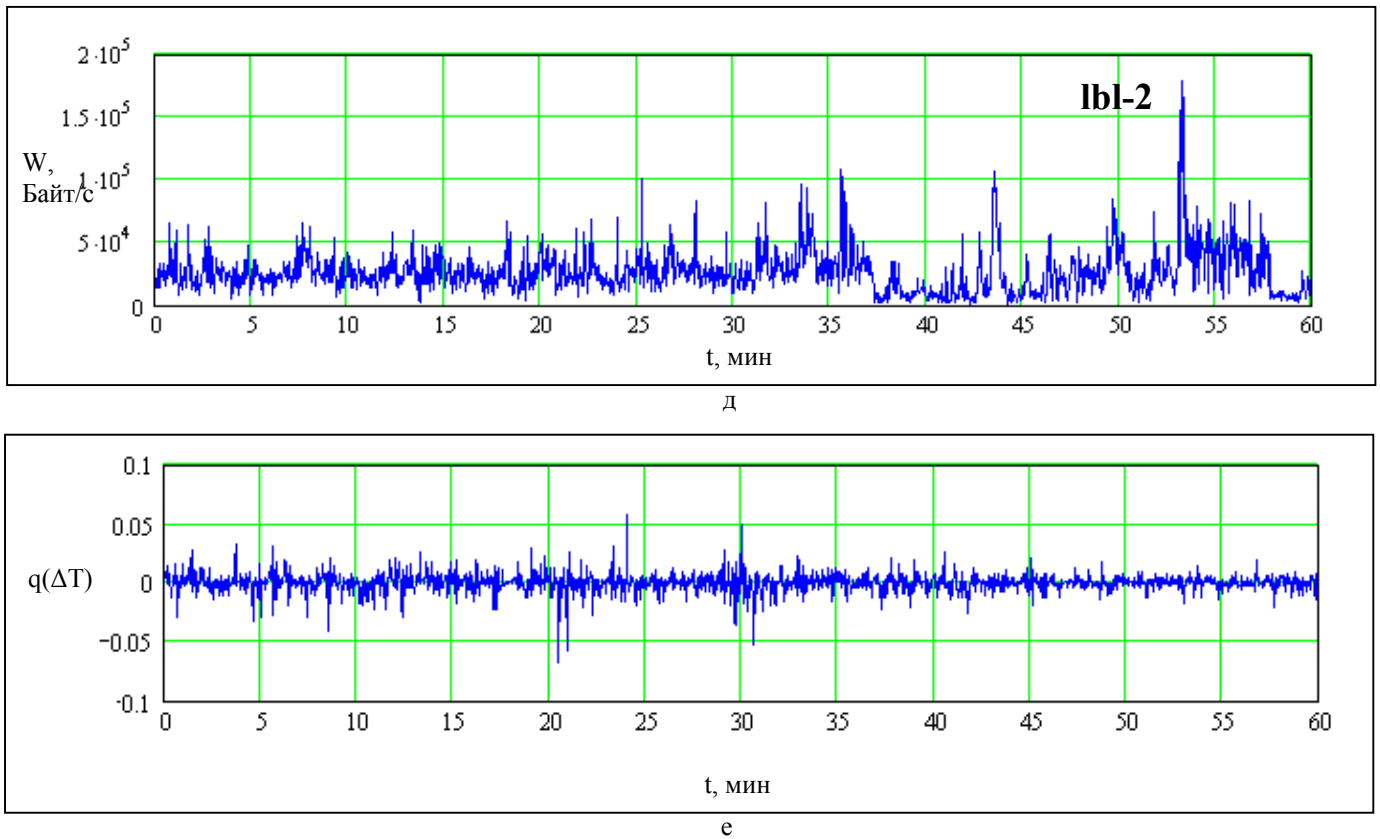
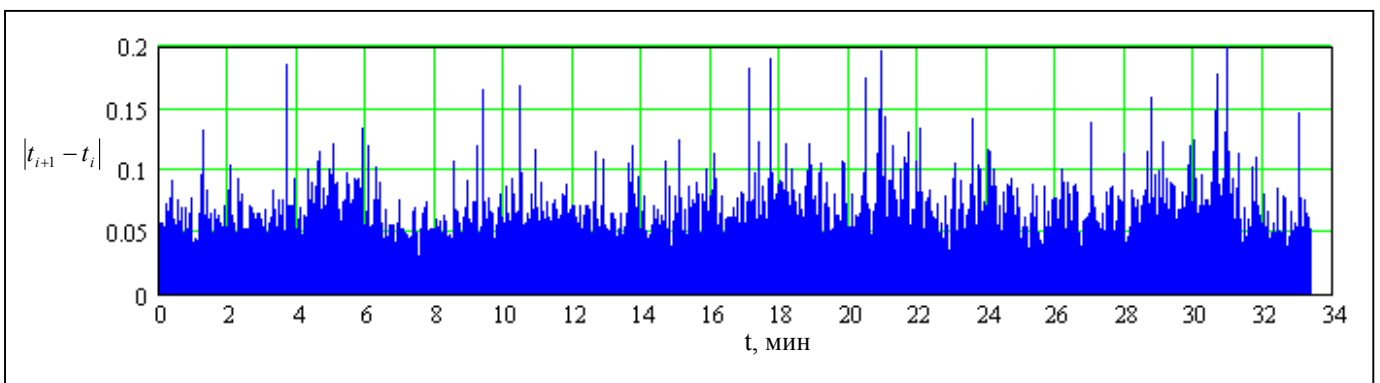


Рис. 2.4. а) Временной ряд lbl-0.1;
 б) Ошибки $q(\Delta T)$, $q(2\Delta T)$, $q(3\Delta T)$, ..., $q(N\Delta T)$ при получении ряда lbl-0.1;
 в) Временной ряд lbl-1;
 г) Ошибки $q(\Delta T)$, $q(2\Delta T)$, $q(3\Delta T)$, ..., $q(N\Delta T)$ при получении ряда lbl-1;
 д) Временной ряд lbl-2;
 е) Ошибки $q(\Delta T)$, $q(2\Delta T)$, $q(3\Delta T)$, ..., $q(N\Delta T)$ при получении ряда lbl-2;

Первые разности временных отсчетов исходной реализации *LBL-PKT-5.TCP* приведены на рис. 2.5. В отличие от реализации *BC-Oct89Ext.TL* трафик *LBL-PKT-5.TCP* (возможно благодаря более короткой длительности последнего) характеризуется более равномерным поступлением дэйтаграмм.



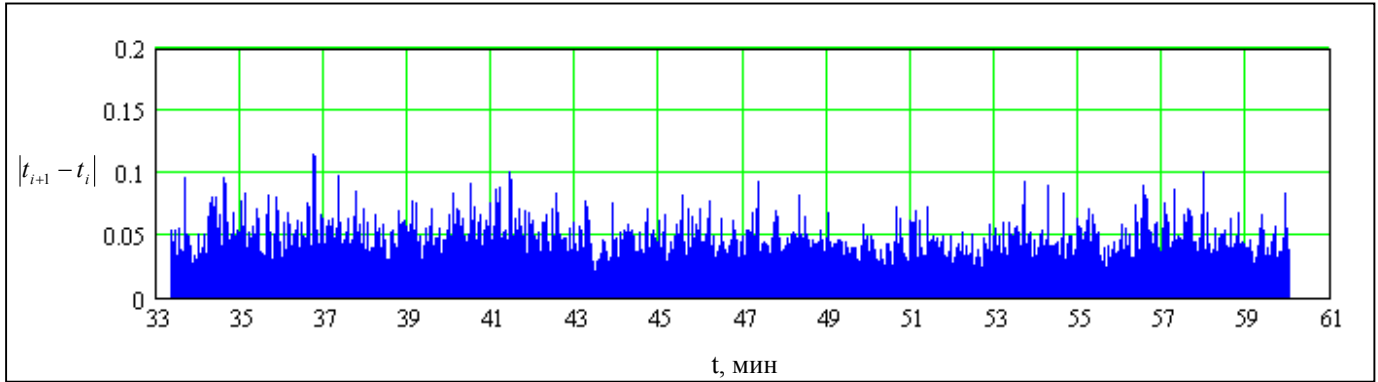
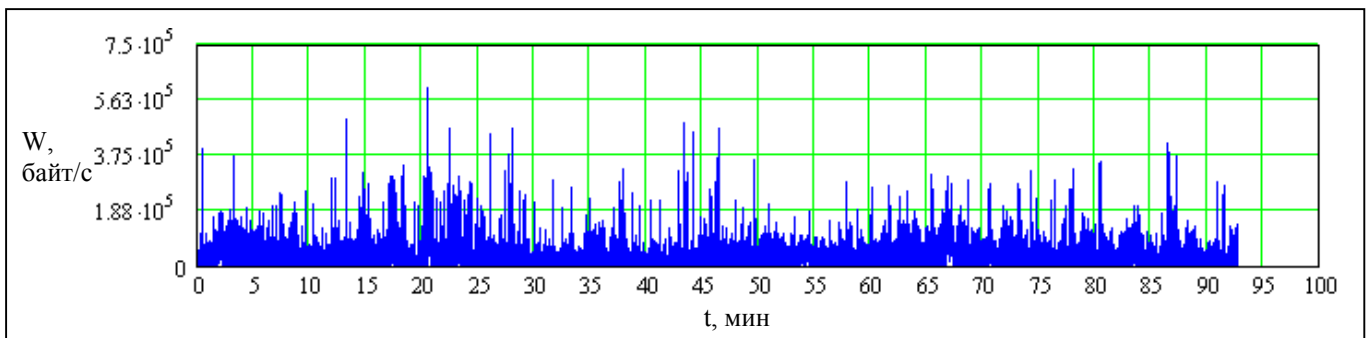
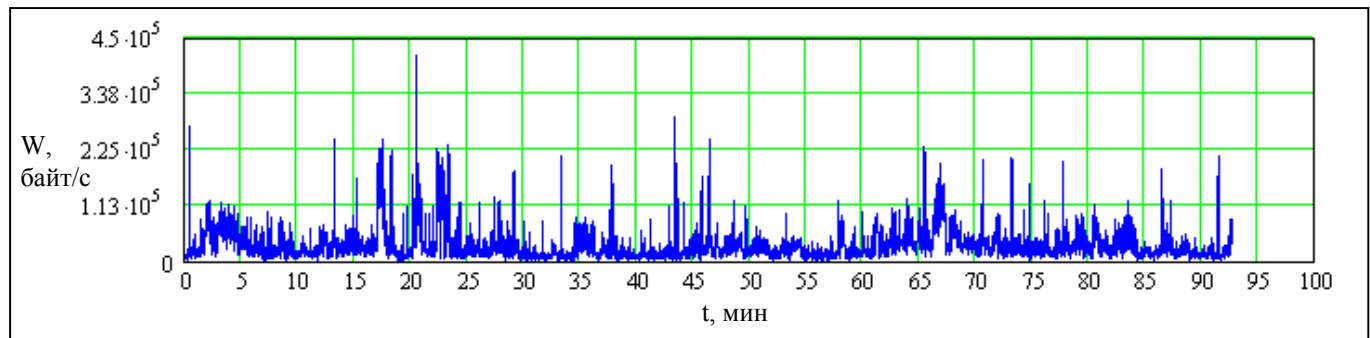


Рис. 2.5. Зависимость $|t_{i+1} - t_i|$ от времени, где $i \in 1, n$ для реализации *LBL-ПКТ-5.TCP*.

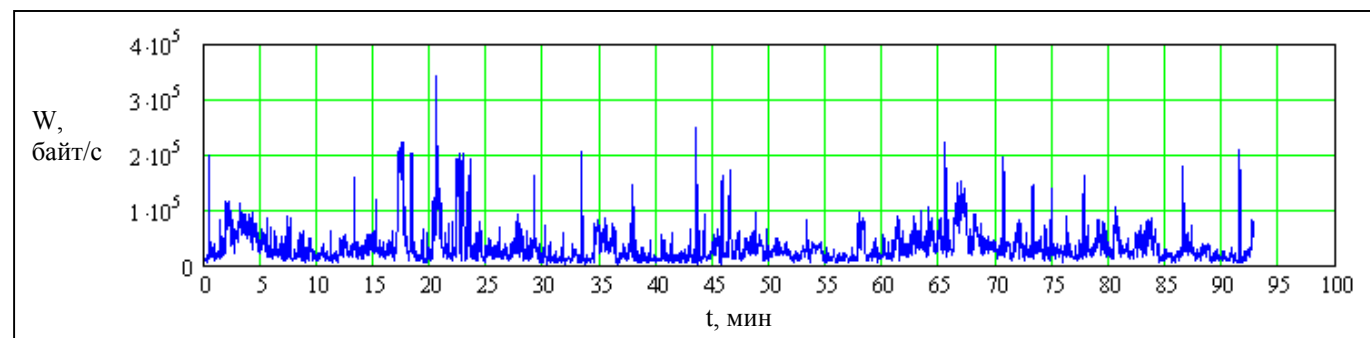
Агрегированная по уровням $\Delta T=0.1$ с и $\Delta T=1$ с и $\Delta T=2$ с реализация *LBL-TCP-3.TCP* представлена на рис. 2.6.



а

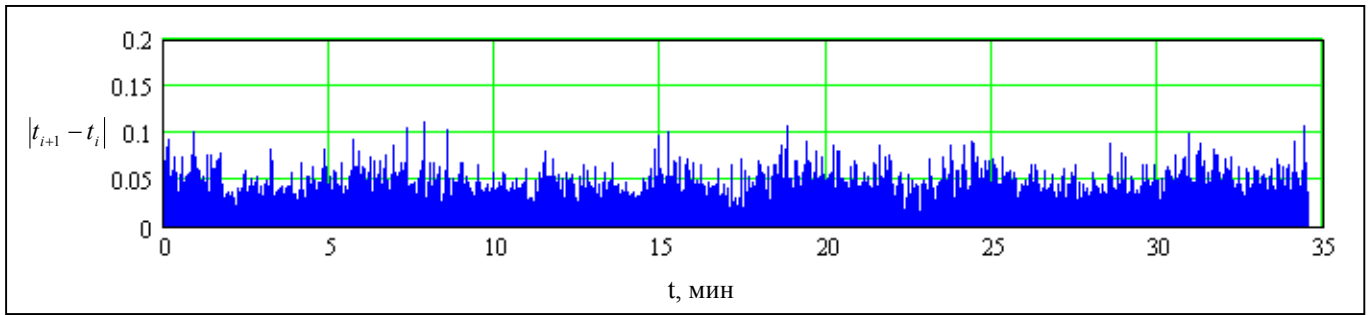


б

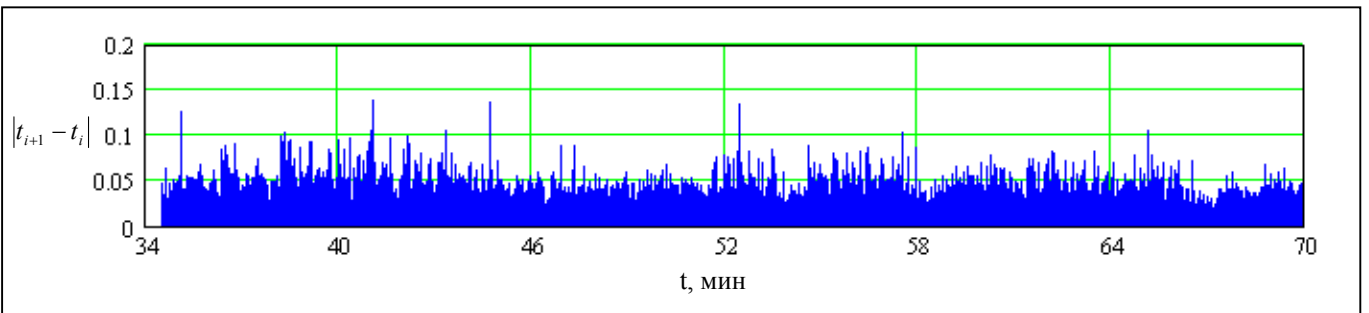


в

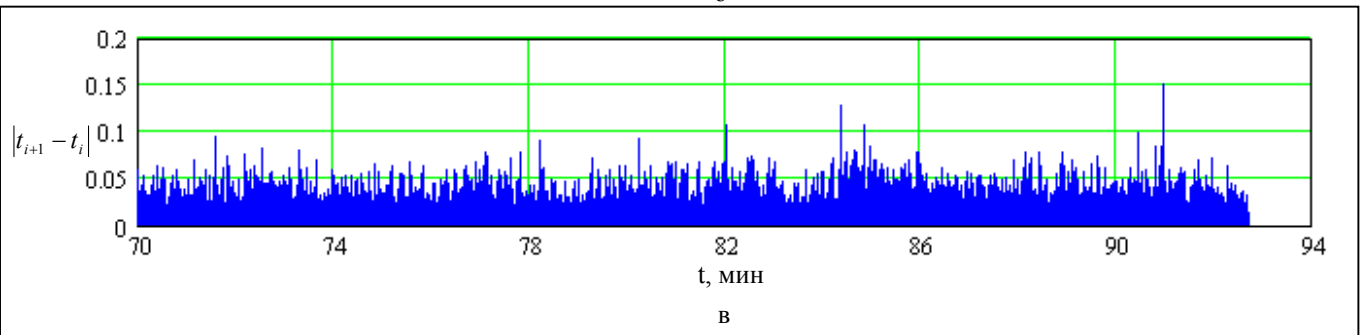
Рис. 2.6. Реализация *LBL-TCP-3.TCP* при различных уровнях усреднения: а) $\Delta T=0.1$ с; б) $\Delta T=1$ с; в) $\Delta T=2$ с.



а



б



в

Рис. 2.7. Первые разности временных отсчетов реализации LBL-TCP-3.TCP

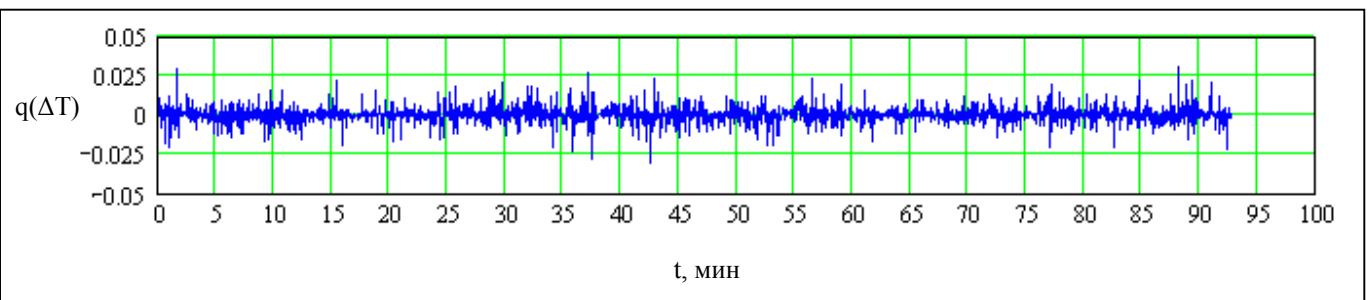


Рис. 2.8. Ошибки $q(\Delta T)$, $q(2\Delta T)$, $q(3\Delta T)$, ..., $q(N\Delta T)$ при получении ряда TCP-2

2.4. Тестовые реализации

Для дальнейшего анализа в качестве тестовых рядов нам также понадобятся реализации чисто детерминированного и чисто случайного процессов. В качестве первой

выберем часто используемую в литературе зависимость (ряд L далее) одной из переменных хаотической системы Лоренца [9] (см. рис. 2.9):

$$\begin{aligned}\frac{dx}{dt} &= \sigma(y - x) \\ \frac{dy}{dt} &= rx - y - xz \\ \frac{dz}{dt} &= -bz + xy\end{aligned}$$

при “стандартных” параметрах $\sigma = 10$, $r = 28$, $b = 8/3$.

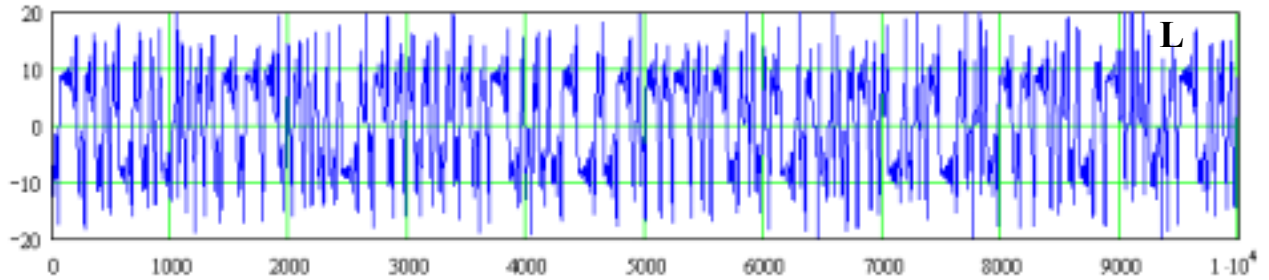


Рис. 2.9. Зависимость от времени одной из переменных хаотической системы Лоренца. Получена при помощи программы Fractan 4.4 [10].

В качестве чисто случайного процесса выберем реализацию (ряд N далее) классического белого гауссовского шума с нулевым средним и стандартным отклонением σ .

Пример такого процесса при $\sigma = 1000$ представлен на рис.

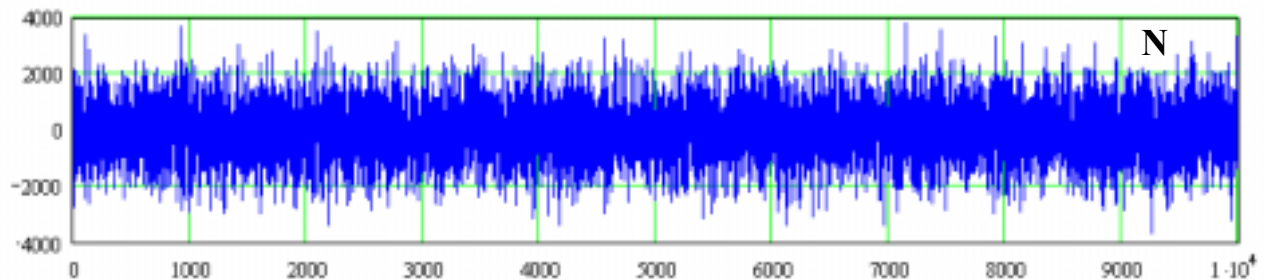


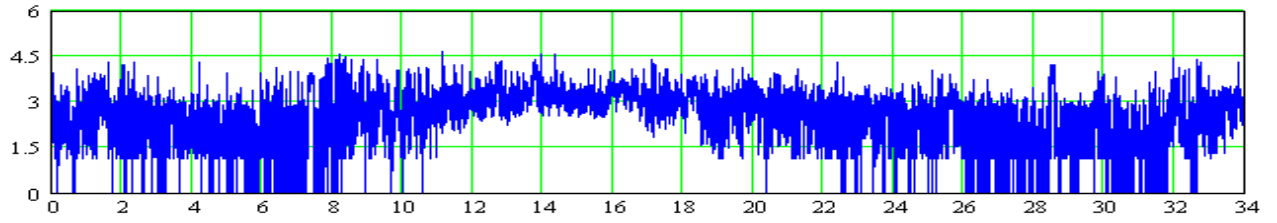
Рис. 2.10. Пример реализации белого гауссовского шума. Ряд получен при помощи программы Fractan 4.4 [10].

2.5. Логарифмированные реализации

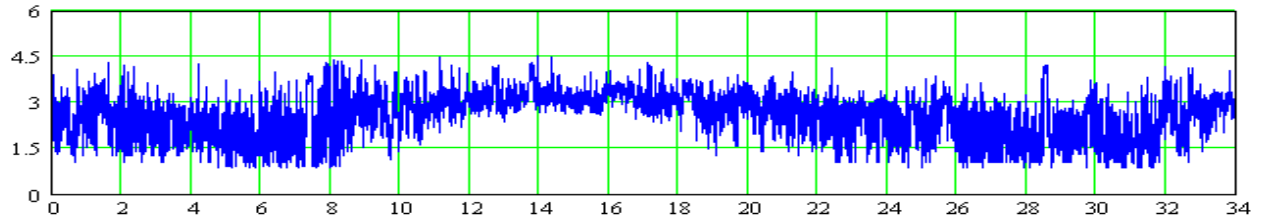
Построим также логарифмы эквидистантных рядов BC-5, BC-10, BC-100, LBL-0.1, LBL-1, LBL-2, TCP-0.1, TCP-1, TCP-2 в соответствии с (2.1)

$$XL_i = \log(X_i + 1), \quad i \in 1, N \quad (2.1)$$

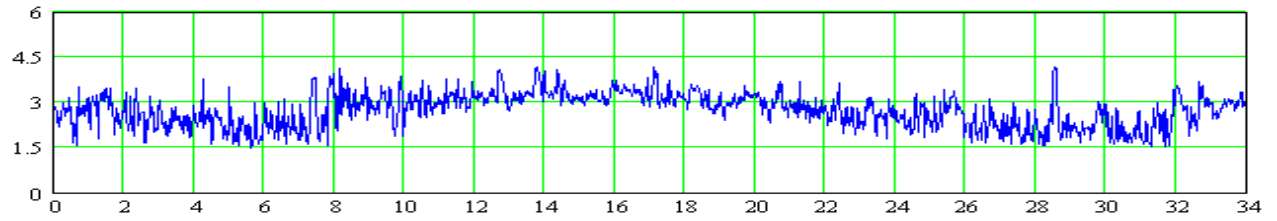
Полученные ряды назовем BC-5-LOG, BC-10-LOG, BC-100-LOG, LBL-0.1-LOG, LBL-1-LOG, LBL-2-LOG, TCP-0.1-LOG, TCP-1-LOG и TCP-2-LOG.



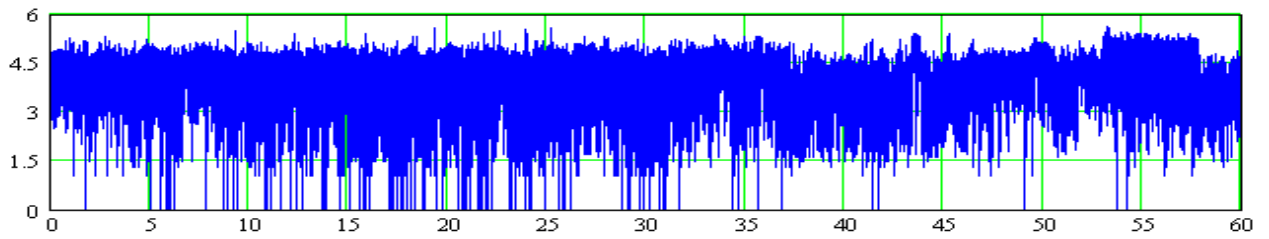
а



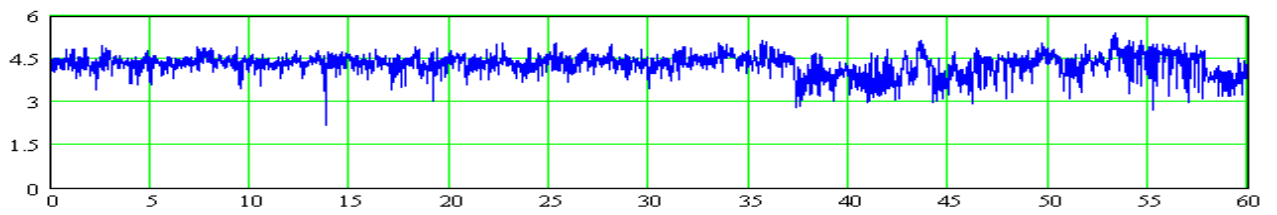
б



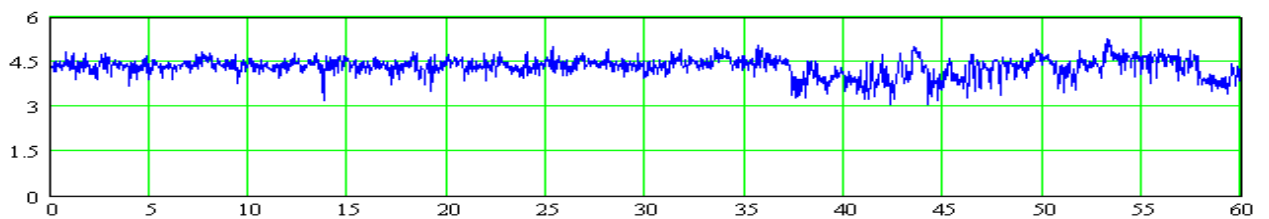
в



г



д



е

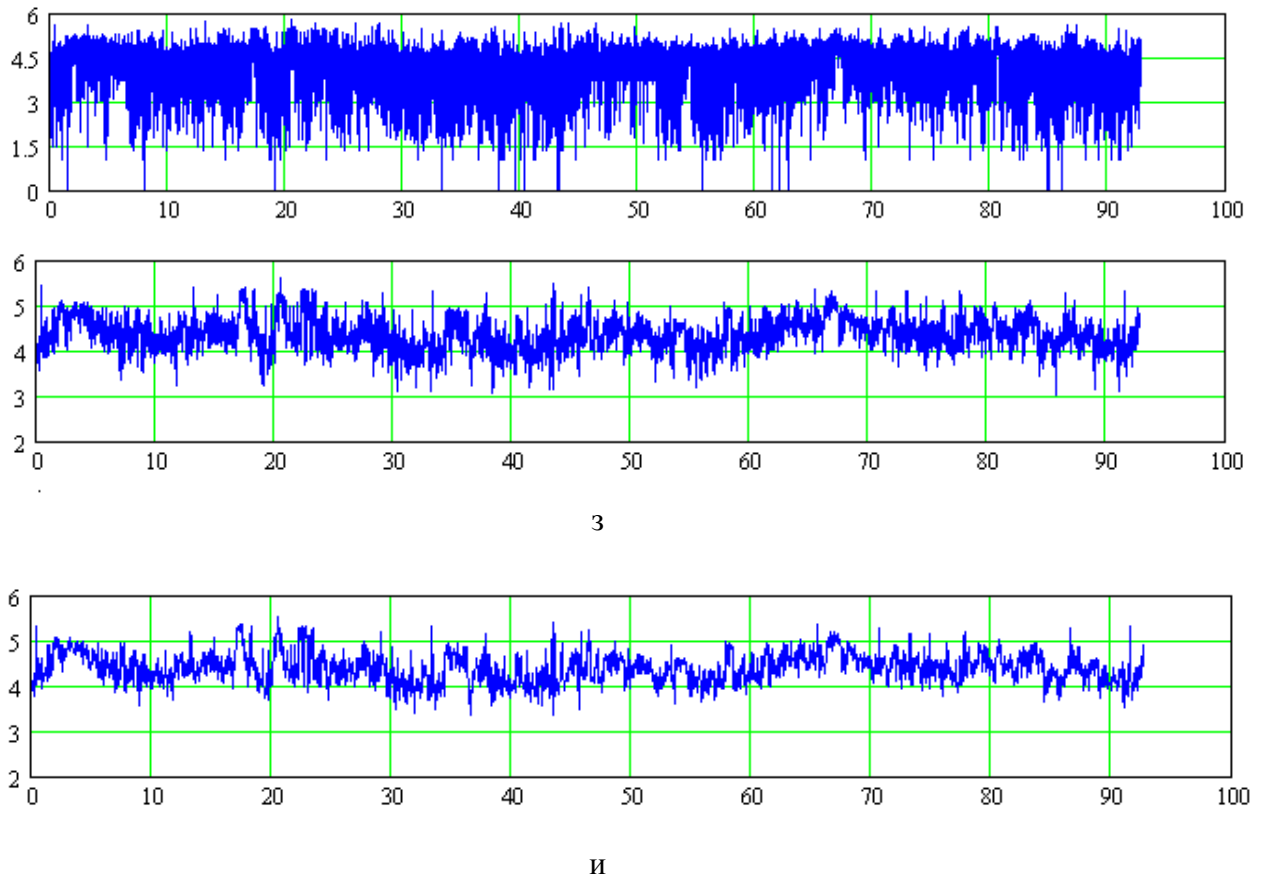


Рис. 2.11. Временные ряды

- а) BC-5-LOG; б) BC-10-LOG; в) BC-100-LOG; г) LBL-0.1-LOG; д) LBL-1-LOG;
 е) LBL-2-LOG, ж) TCP-0.1-LOG, з) TCP-1-LOG, и) TCP-2-LOG.

3. Классический анализ

В таблице 3.1 приведены основные характеристики исследуемых реализаций: среднее квадратическое отклонение σ , среднее m , количество N отсчетов (членов) ряда, а также максимально возможная корреляционная размерность $D_{max} = 2 \lg(N)$, которую можно оценить при помощи корреляционного интеграла по ряду данной длины (см. ниже).

| Реализация/ Характеристика | σ | m | N | D_{max} |
|----------------------------|----------|--------|--------|-----------|
| BC-5 | 2405 | 1 142 | 24 560 | 8.782 |
| BC-10 | 2243 | 1 142 | 12 279 | 8.178 |
| BC-100 | 1779 | 1 142 | 1 227 | 6.178 |
| BC-5-LOG | 0.775 | 2.508 | 24 560 | 8.782 |
| BC-10-LOG | 0.702 | 2.572 | 12 279 | 8.178 |
| BC-100-LOG | 0.565 | 2.721 | 1 227 | 6.178 |
| BC-5-surr-001 | 2409 | 1 139 | 24 300 | 8.771 |
| BC-10-LOG-surr-001 | 0.704 | 2.567 | 12 150 | 8.169 |
| LBL-0.1 | 32 250 | 26 146 | 36 000 | 9.113 |
| LBL-1 | 19 331 | 26 151 | 3 600 | 7.113 |

| | | | | |
|----------------------|--------|--------|--------|-------|
| LBL-2 | 17 174 | 26 157 | 1 800 | 6.511 |
| LBL-0.1-LOG | 0.821 | 3.977 | 36 000 | 9.113 |
| LBL-1-LOG | 0.365 | 4.297 | 3 600 | 7.113 |
| LBL-2-LOG | 0.328 | 4.320 | 1 800 | 6.511 |
| LBL-0.1-surr-001 | 32 520 | 26 639 | 34 992 | 9.088 |
| LBL-0.1-LOG-surr-001 | 0.824 | 3.989 | 34 992 | 9.088 |
| TCP-0.1 | 43 420 | 34 107 | 55 634 | 9.491 |
| TCP-1 | 34 773 | 34 109 | 5 563 | 7.491 |
| TCP-2 | 32 528 | 34 100 | 2 781 | 6.888 |
| TCP-0.1-LOG | 0.674 | 4.175 | 55 634 | 9.491 |
| TCP-1-LOG | 0.373 | 4.374 | 5 563 | 7.491 |
| TCP-2-LOG | 0.344 | 4.396 | 2 781 | 6.888 |
| TCP-0.1-surr-001 | 43 474 | 34 092 | 55 296 | 9.485 |
| TCP-0.1-LOG-surr-001 | 0.674 | 4.175 | 55 296 | 9.485 |
| N | 1 010 | -7.892 | 10 000 | 8.000 |
| L | 8.450 | 0.528 | 10 000 | 8.000 |

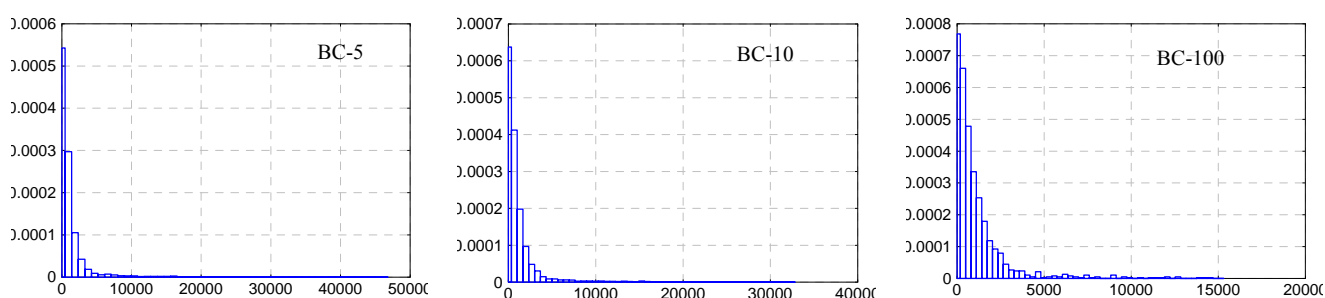
Табл. 3.1. Основные характеристики исследуемых реализаций

Некоторые из перечисленных в табл. 1 реализаций (с суффиксами “-surr-”) также будут нами рассматриваться ниже.

Изучая результаты измерений можно заметить, что для не логарифмированных реализаций (BC, LBL, TCP) при увеличении уровня агрегирования наблюдается уменьшение среднеквадратического отклонения на фоне практически неизменного среднего. Данный факт говорит о “сглаживании” реализации при увеличении уровня агрегирования. Аналогичные наблюдения можно сделать для логарифмированных реализаций, для которых, тем не менее, наблюдается некоторое увеличение среднего при увеличении уровня агрегирования, однако общая тенденция к сглаживанию сохраняется и в этом случае.

3.1. Плотности распределения

В настоящем разделе производится анализ плотностей распределения агрегированных реализаций BC, LBL, TCP а также тестовых реализаций N и L. Оценка плотностей распределения производится на основании геометрического изображения (гистограмм) относительных частот появления предварительно заданных участков данных (см. рис. 3.1).



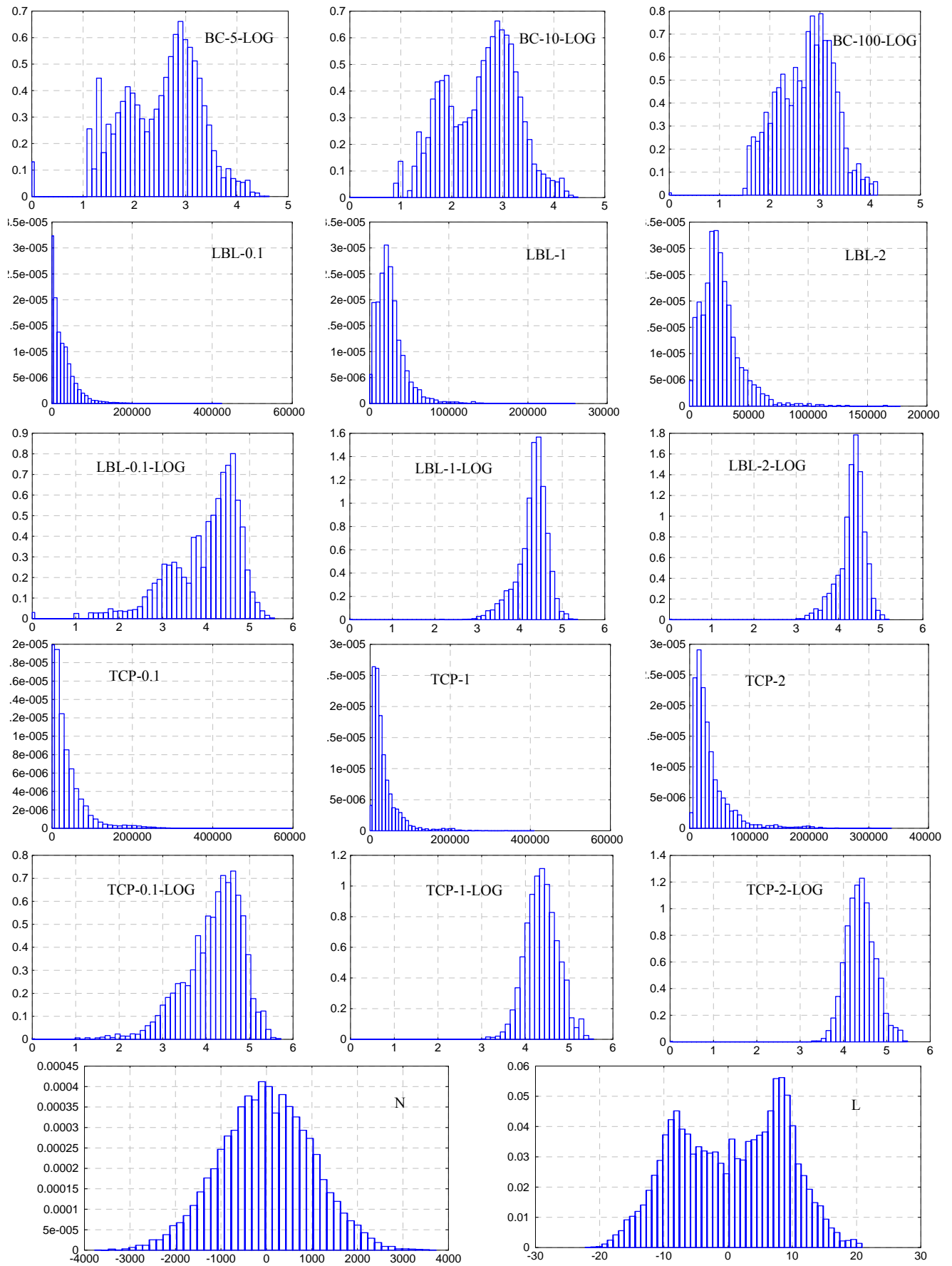


Рис. 3.1. Гистограммы изучаемых реализаций.

Визуальный анализ гистограмм изучаемых временных рядов позволяет сделать некоторые выводы:

- ряды ВС-х, LBL-х а также TCP-х скорее всего подчиняются некоторому распределению с так называемым “тяжелым хвостом” (например Парето);
- увеличение уровня агрегации приводит к изменению параметров “тяжелохвостого” распределения. Более точный ответ может дать исследование распределений методом максимального правдоподобия.
- логарифмирование рядов приводит к нормализации их распределений;
- гистограмма хаотической реализации L имеет два характерных максимума (при значениях ~ -9 и ~ 9). Аналогичные максимумы наблюдаются в реализации ВС-10-LOG и (менее выраженные) в LBL-0.1-LOG.

3.2. Автокорреляционные функции

На рис. 3.2 приведены автокорреляционных функций исследуемых временных рядов. Вычисления производились по классической формуле (3.1)

$$r(k) = \frac{1}{N - \tau} \cdot \frac{\sum_i (X_i - \bar{X})(X_{i+k} - \bar{X})}{\sigma^2(X)} \quad (3.1)$$

где \bar{X} - выборочное среднее ряда X , $\sigma^2(X)$ - выборочная дисперсия ряда X , $k \in Z_+ = \{0, 1, 2, \dots\}$.

Заметим, что АКФ рядов LBL-0.1 и LBL-0.1-LOG имеют отчетливую периодическую структуру (с частотой ~ 5 Гц). Менее выраженные периодичности (пульсации) АКФ имеются в рядах ВС-5-LOG и ВС-10-LOG, а также в рядах TCP-0.1 и TCP-0.1-LOG. Характерно, что с увеличением уровня агрегации периодичности исчезают. Можно предположить наличие соответствующих гармоник в спектрах данных рядов, в частности для рядов LBL-0.1, LBL-0.1-LOG, TCP-0.1 и TCP-0.1-LOG – выделяющейся гармонике с частотой 5 Гц.

Выясним, какие из исследуемых временных рядов обладают медленно убывающей зависимостью (МУЗ), а какие быстро убывающей зависимостью. Для начала напомним определения.

Определение [3]. Говорят, что процесс X обладает *медленно убывающей зависимостью* (МУЗ) [long-range dependence], если выполняется (3.2). То есть, процессы с МУЗ характеризуются автокорреляционной функцией, которая убывает гиперболически (по степенному закону) при увеличении временной задержки (лага).

$$r(k) \sim k^{-\beta} \cdot L_1(k), \quad k \rightarrow \infty \quad (3.2)$$

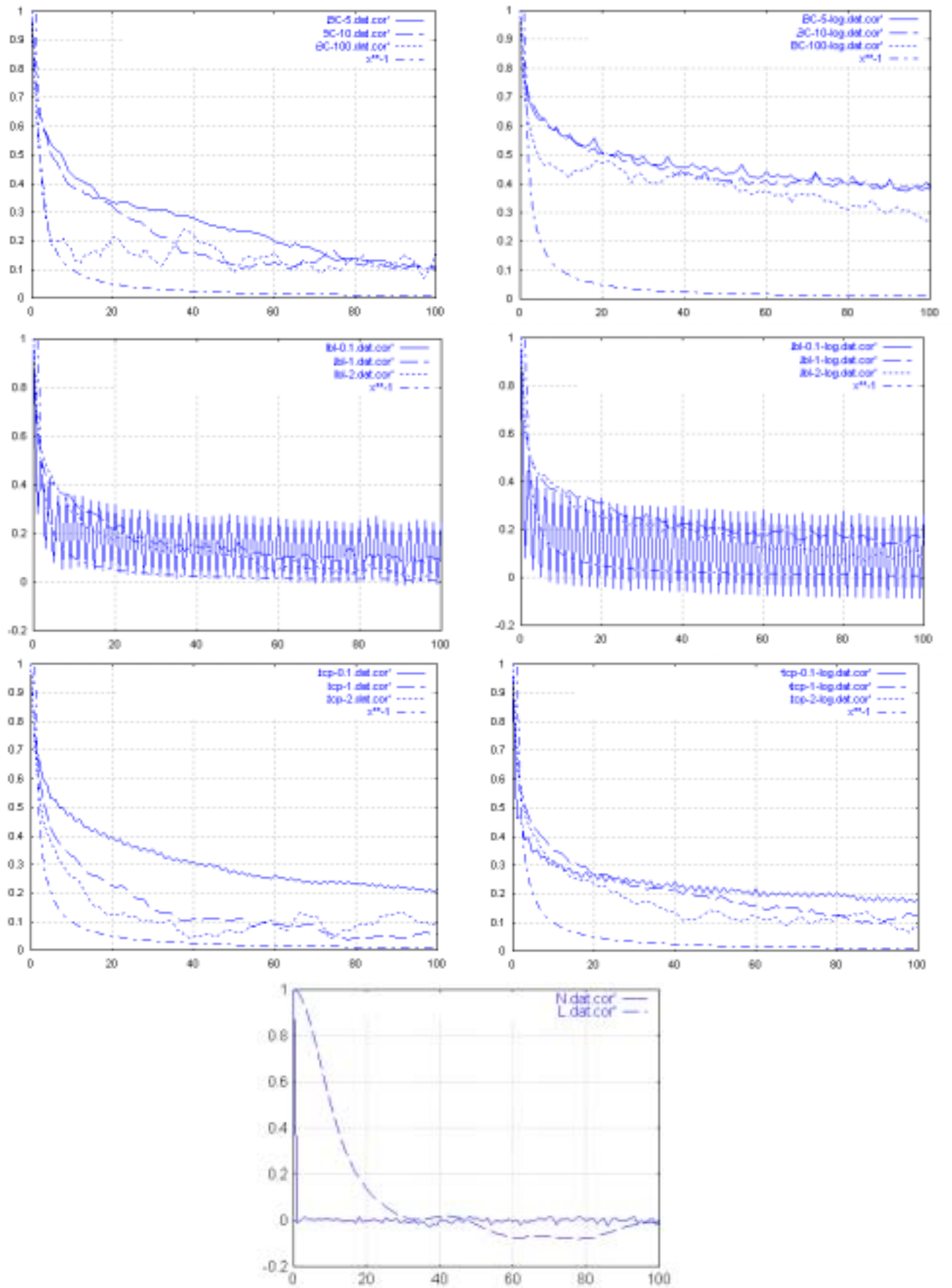


Рис. 3.2. Графики автокорреляционных функций $g(k)$

где $0 < \beta < 1$ и L_1 – медленно меняющаяся на бесконечности функция, то есть $\lim_{t \rightarrow \infty} \frac{L_1(tx)}{L_1(t)} = 1$ для всех $x > 0$ (примерами медленно меняющейся функции могут служить

$$L_1(t) = \text{const}, L_1(t) = \log(t).$$

На каждом графике рис. 3.2 для сравнения также изображена зависимость $C(k) = k^{-\beta}$ при $\beta = 1$, соответствующая предельному случаю существования МУЗ.

В отличие от процессов с МУЗ, процессы с **быстро убывающей зависимостью** (БУЗ) [short-range dependence] обладают экспоненциально спадающей АКФ вида

$$r(k) \sim \rho^k, \quad k \rightarrow \infty, \quad 0 < \rho < 1 \quad (3.3)$$

Итак, для того чтобы проверить гипотезу о присутствии МУЗ в некотором временном ряде необходимо решить задачу регрессии, то есть вычислить по экспериментальной АКФ методом наименьших квадратов параметры A_0 и β модели (3.4):

$$r(k) = A_0 \cdot k^{-\beta}, \quad \text{при } k \rightarrow \infty \quad (3.4)$$

где $0 < \beta < 1$ и $A_0 = \text{const}$. Для оценки степени точности необходимо найти дисперсию $\sigma_{\text{МУЗ}}^2$ разности экспериментальной АКФ и полученной ее аппроксимации (3.4).

Одновременно проверяется гипотеза о присутствии БУЗ в этом же ряде. Для этого аналогичным образом вычисляются параметры B_0 и ρ модели (3.5):

$$r(k) = B_0 \cdot \rho^k, \quad \text{при } k \rightarrow \infty \quad (3.5)$$

где $0 < \rho < 1$ и $B_0 = \text{const}$. Для оценки степени точности необходимо найти дисперсию $\sigma_{\text{БУЗ}}^2$ разности экспериментальной АКФ и полученной ее аппроксимации (3.5).

Вывод о присутствии МУЗ в ряде логично сделать, если $\sigma_{\text{МУЗ}}^2$ – достаточно мало и значение коэффициента

$$LR = \frac{\sigma_{\text{АКФ}}^2}{\sigma_{\text{АКФ}}^2}$$

меньше 1. Напротив, будем говорить, что в исследуемом ряде присутствует БУЗ если $LR > 1$ и $\sigma_{\text{БУЗ}}^2$ – мало. Конечно, условие малости $\sigma_{\text{МУЗ}}^2$ или $\sigma_{\text{БУЗ}}^2$ достаточно условно и отражает лишь тот факт, что степень аппроксимации экспериментальной зависимости моделью (3.4) или (3.5) нас удовлетворяет в плане корректности выбора модели. Выполнимость данного условия проще всего проконтролировать по графикам, на которых изображены одновременно экспериментальная АКФ и модели. Поскольку (3.4) и (3.5) справедливы для $k \rightarrow \infty$, хотя экспериментальные данные (АКФ) ограничены количеством отсчетов M , условимся, что регрессию будем осуществлять для $k = \frac{M}{10}, \dots, M$. То есть будем рассматривать асимптотику. Заметим, что реализации LBL-0.1 и LBL-0.1-LOG не имеет

смысла аппроксимировать моделями (3.4) или (3.5) поскольку первые имеют сильно выраженную периодическую структуру. Результаты вычислений приведены в таблице 3.2.

Анализируя полученные результаты отметим, что коэффициент $LR < 1$ практически для всех реализаций, что подтверждает присутствие в них МУЗ. Исключения составляют ряды BC-100, LBL-1, LBL-2, LBL-1-LOG и LBL-2-LOG (выделенные участки таблицы 3.2). Однако, рассматривая более внимательно параметры модели БУЗ согласно (3.5) можно заметить, что абсолютно для всех случаев $\rho \sim 1$. Другими словами модель (3.5) вырождается в (3.6)

$$r(k) \sim const \tag{3.6}$$

что говорит об очень медленном убывании АКФ.

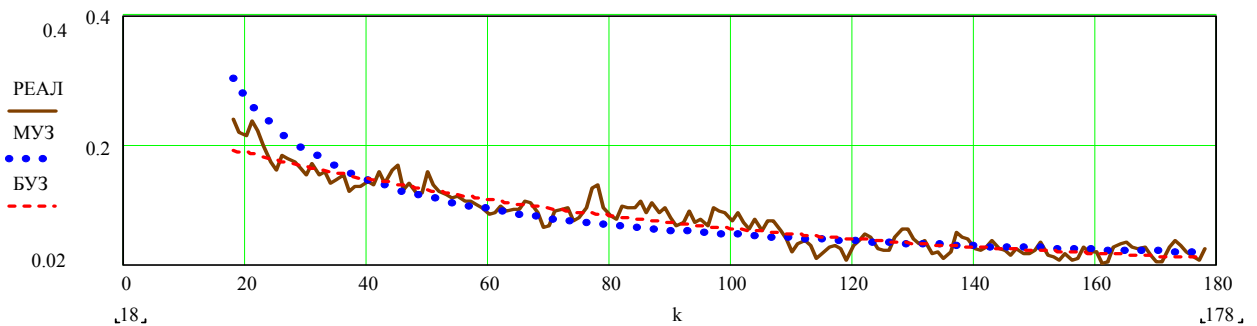


Рис. 3.3. Регрессионный анализ LBL-1

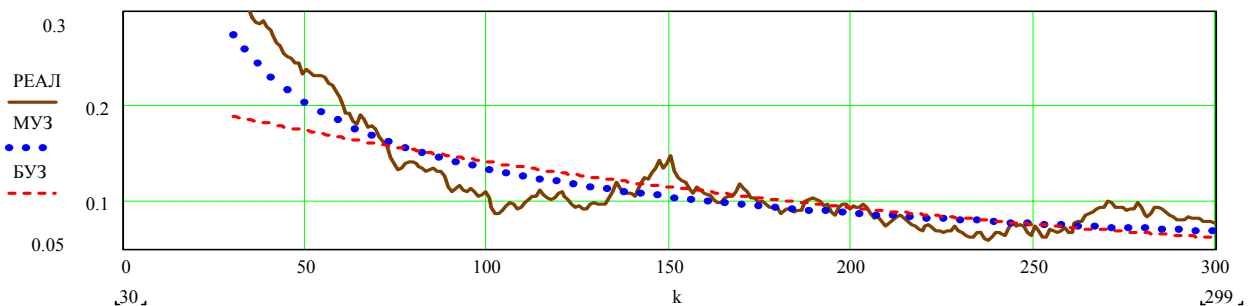


Рис. 3.4 . Регрессионный анализ BC-5

Для примера на рис. 3.3 и 3.4 даны результаты аппроксимации АКФ реализаций LBL-1 и BC-5.с помощью моделей (3.4) и (3.5).

| | | |
|--|--|--|
| <p>BC-5, M=300, $A_0 = 2.25, \beta = 0.611, \sigma^2_{МУЗ} = 0.000411$ $B_0 = 0.214, \rho = 0.9959, \sigma^2_{БУЗ} = 0.00116$ LR= 0.354404</p> | <p>BC-10, M=300, $A_0 = 0.24, \beta = 0.162, \sigma^2_{МУЗ} = 0.000435$ $B_0 = 0.126, \rho = 0.9990, \sigma^2_{БУЗ} = 0.00051$ LR= 0.850971</p> | <p>BC-100, M=150, $A_0 = 0.422, \beta = 0.280, \sigma^2_{МУЗ} = 0.001214$ $B_0 = 0.180, \rho = 0.9958, \sigma^2_{БУЗ} = 0.00122$ LR= 0.992989</p> |
| <p>BC-5-LOG, M=300, $A_0 = 0.766, \beta = 0.145, \sigma^2_{МУЗ} = 0.000174$ $B_0 = 0.4398, \rho = 0.9990, \sigma^2_{БУЗ} = 0.00037$ LR= 0.473328</p> | <p>BC-10-LOG, M=300, $A_0 = 0.511, \beta = 0.054, \sigma^2_{МУЗ} = 0.000242$ $B_0 = 0.4131, \rho = 0.9997, \sigma^2_{БУЗ} = 0.000298$ LR= 0.811935</p> | <p>BC-100-LOG, M=150, $A_0 = 1.433, \beta = 0.3440, \sigma^2_{МУЗ} = 0.000681$ $B_0 = 0.5112, \rho = 0.9946, \sigma^2_{БУЗ} = 0.000223$ LR= 3.056881</p> |

| | | |
|---|--|--|
| | LBL-1, M=180, $A_0 = 4.4887, \beta = 0.916, \sigma^2_{МУЗ} = 0.00047$ $B_0 = 0.2405, \rho = 0.9883, \sigma^2_{БУЗ} = 0.000247$ LR= 1.906777 | LBL-2, M=91, $A_0 = 3.16996, \beta = 0.92473, \sigma^2_{МУЗ} = 0.00063$ $B_0 = 0.30474, \rho = 0.9769, \sigma^2_{БУЗ} = 0.000298$ LR= 2.117156 |
| | LBL-1-LOG, M=250, $A_0 = 4.3239, \beta = 0.7679, \sigma^2_{МУЗ} = 0.00056$ $B_0 = 0.9928, \rho = 0.9934, \sigma^2_{БУЗ} = 0.000224$ LR= 2.503366 | LBL-2-LOG, M=100, $A_0 = 2.24027, \beta = 0.69325, \sigma^2_{МУЗ} = 0.00058$ $B_0 = 0.36159, \rho = 0.98428, \sigma^2_{БУЗ} = 0.000264$ LR= 2.208149 |
| TCP-0.1, M=300, $A_0 = 2.259, \beta = 0.523, \sigma^2_{МУЗ} = 0.00012$ $B_0 = 0.3228, \rho = 0.9960, \sigma^2_{БУЗ} = 0.00015$ LR= 0.80758 | TCP-1, M=200, $A_0 = 0.2920, \beta = 0.2786, \sigma^2_{МУЗ} = 0.00093$ $B_0 = 0.09944, \rho = 0.9982, \sigma^2_{БУЗ} = 0.00120$ LR= 0.77632 | TCP-2, M=89, $A_0 = 0.55824, \beta = 0.47581, \sigma^2_{МУЗ} = 0.00112$ $B_0 = 0.14666, \rho = 0.9908, \sigma^2_{БУЗ} = 0.001906$ LR= 0.585866 |
| TCP-0.1-LOG, M=300, $A_0 = 0.7557, \beta = 0.314, \sigma^2_{МУЗ} = 0.00007$ $B_0 = 0.2342, \rho = 0.9976, \sigma^2_{БУЗ} = 0.00008$ LR= 0.89552 | TCP-1-LOG, M=200, $A_0 = 1.3356, \beta = 0.5201, \sigma^2_{МУЗ} = 0.00020$ $B_0 = 0.22649, \rho = 0.9945, \sigma^2_{БУЗ} = 0.00055$ LR= 0.36068 | TCP-2-LOG, M=100, $A_0 = 1.15918, \beta = 0.55224, \sigma^2_{МУЗ} = 0.00027$ $B_0 = 0.256227, \rho = 0.9885, \sigma^2_{БУЗ} = 0.000731$ LR= 0.36993 |

Табл. 3.2. Результаты регрессионного анализа.

Таким образом, в результате приведенного анализа логично сделать вывод о присутствии МУЗ во всех исследуемых реализациях трафика.

3.3. Энергетические спектры

В настоящем разделе производится расчет энергетических спектров исследуемых временных реализаций. Известно, что в частотной области МУЗ отражается на характерном степенном законе поведения спектральной плотности рассматриваемого процесса. Действительно, эквивалентно (3.2) можно констатировать, что процесс X обладает МУЗ, если

$$f(\lambda) \sim \lambda^{\beta-1} L_2(\lambda), \quad \lambda \rightarrow 0, \quad 0 < \beta < 1 \quad (3.4)$$

здесь L_2 – медленно изменяющаяся в нуле функция, $f(\lambda) = \sum_k r(k) e^{ik\lambda}$ - означает спектральную плотность. Таким образом, с точки зрения спектрального анализа процесс с МУЗ обладает спектральной плотностью с особенностью в нуле (т.е. спектральная плотность $f(\lambda)$ такого процесса стремится к бесконечности, по мере того как частота λ стремится к нулю). Такой процесс часто называют “1/f - шум” или “фликкер-шум”.

Рассчитанные энергетически спектры реализаций представлены на рис. 3.5.

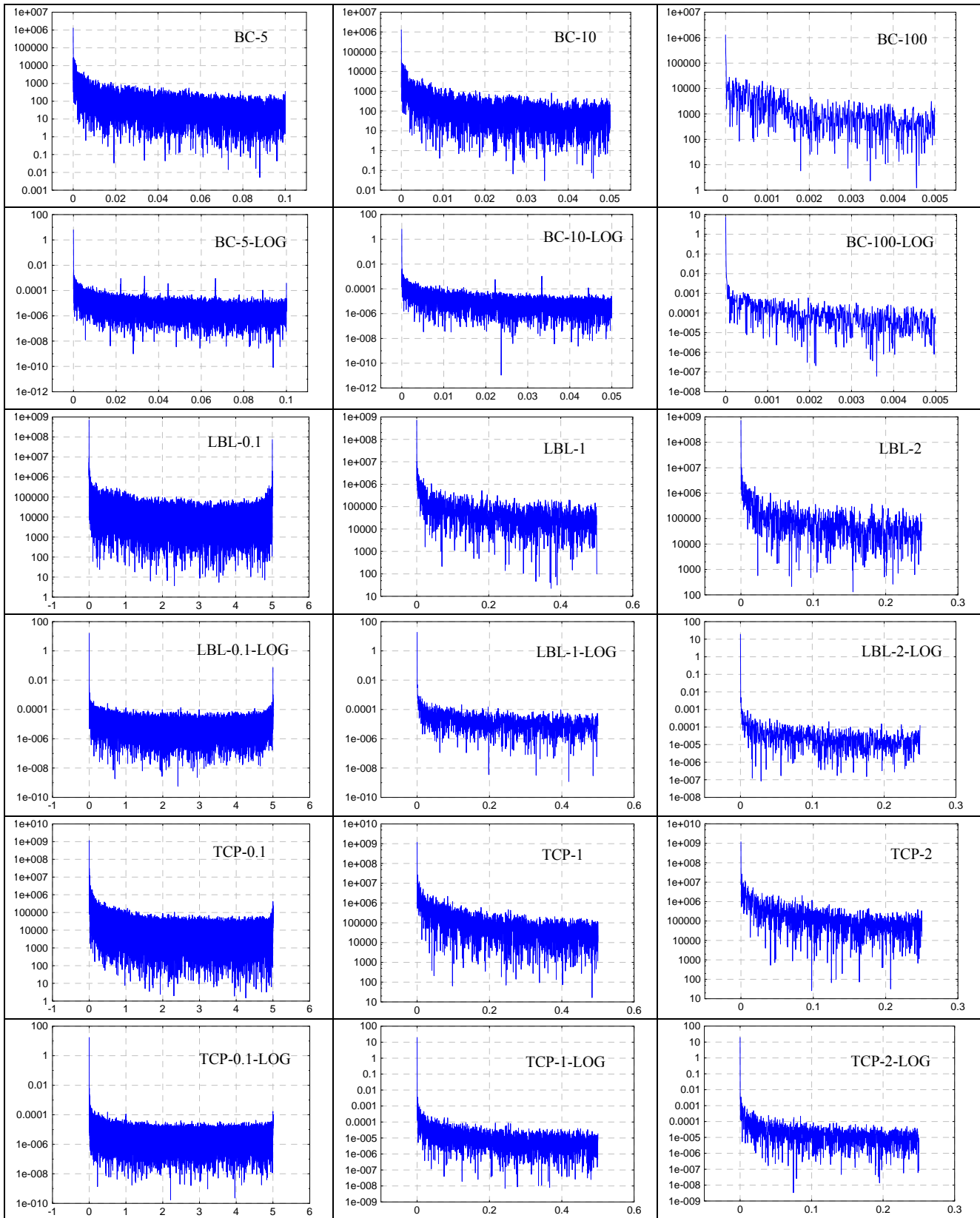


Рис. 3.5 . Энергетические спектры реализаций трафика. Здесь ось абсцисс отложена частота λ (Гц), по оси ординат – спектр $W(\lambda)$ в логарифмическом масштабе.

Анализируя полученные спектры можно заметить наличие сильной гармонической компоненты ~ 5 Гц в реализациях LBL-0.1, LBL-0.1-LOG. Данная гармоника также присутствует (хотя и менее существенна) в реализациях TCP-0.1 и TCP-0.1-LOG. При увеличении уровня агрегирования гармоника исчезает. Очевидно, периодичность АКФ реализаций LBL-0.1, LBL-0.1-LOG обусловлена именно присутствием в их спектрах гармоники 5 Гц. Также в реализациях BC-5-LOG и BC-10-LOG наблюдается несколько сильных гармонических компонент. Отметим, что в [13] и [14], при исследовании реализаций трафика получены схожие результаты и выявлены гармоники с частотами ~ 11.9 Гц и 1.25 Гц, присутствующие в wan-трафике. Тем не менее, пока непонятны причины появления подобных компонент в спектрах трафика.

Обнаруженное явление имеет принципиальное значение, поскольку выявляет присутствие регулярной детерминированной составляющей в агрегированном сетевом трафике, что может быть интересным при решении задач прогнозирования телетрафика.

Одновременно, можно отметить преобладание в полученных спектрах низкочастотных компонент, что согласуется с понятием фликкер-шума, о котором упоминалось ранее.

4. Исследование показателя Хэрста реализаций

Произведем оценку показателя Хэрста H реализаций. Напомним, что при $0.5 < H < 1$ говорят о *персистентном* (поддерживаемом) поведении процесса либо о том, что процесс обладает *длительной памятью*. Другими словами, если в течение некоторого времени в прошлом наблюдались положительные приращения процесса, то есть происходило увеличение, то и впредь в среднем будет происходить увеличение. Иначе говоря, вероятность того, что процесс на $i+1$ шаге отклоняется от среднего в том же направлении, что и на i шаге настолько велика, насколько параметр H близок к 1. Т.е. персистентные стохастические процессы обнаруживают четко выраженные тенденции изменения при относительно малом “шуме”.

В случае $0 < H < 0.5$ говорят о *антиперсистентности* процесса. Здесь высокие значения процесса следуют за низкими, и наоборот. Другими словами, вероятность того, что $i+1$ шаге процесс отклоняется от среднего в противоположном направлении (по отношению к отклонению на i шаге) настолько велика, насколько параметр H близок к 0.

При $H=0.5$ отклонения процесса от среднего являются действительно случайными и не зависят от предыдущих значений, что соответствует случаю БД.

Результаты измерений показателя Хэрста 7-ю методами приведены в табл. 4.1. Здесь через дробь приведены непосредственно сам показатель H и соответствующая ему точность измерения. Точность определяется качеством аппроксимации точек вычисляемых статистик отрезком прямой либо вероятностью нахождения параметра H в заданном доверительном

интервале (для методов Abru-Veitch Estimator и Whittle Estimator). Если точность невелика (заштрихованные области), то данную оценку H нельзя принимать во внимание.

| № | Реализация/ Метод | Коэффициент Хэрста | | | | | | |
|----|----------------------|-----------------------|---------------|------------------|---------------------|--------------------------|------------------------------|----------------------|
| | | Aggregate Variance | R/S | Periodo- gram | Absolute Moments | Variance of Residuals | Abru- Veitch Estimator | Whittle Estimator |
| 1 | BC-5 | 0.842 / 98.93 | 0.770 / 99.36 | 0.955 | 0.534 / 55.75 | 1.128 / 99.94 | 1.007 / 95 | 0.949 / 95 |
| 2 | BC-5-LOG | 0.945 / 95.86 | 0.639 / 94.54 | 0.945 | 0.943 / 98.98 | 1.056 / 99.47 | 0.959 / 95 | 0.889 / 95 |
| 3 | BC-10 | 0.837 / 98.95 | 0.745 / 99.08 | 0.920 | 0.479 / 56.61 | 1.111 / 99.58 | 1.037 / 95 | 0.951 / 95 |
| 4 | BC-10-LOG | 0.943 / 95.61 | 0.583 / 92.41 | 0.928 | 0.945 / 95.17 | 1.066 / 99.23 | 0.979 / 95 | 0.920 / 95 |
| 5 | BC-100 | 0.695 / 73.95 | 0.720 / 97.36 | 0.653 | 0.116 / 65.67 | 1.114 / 99.65 | 0.986 / 95 | 0.904 / 95 |
| 6 | BC-100-LOG | 0.771 / 63.99 | 0.436 / 80.52 | 0.759 | 0.793 / 54.78 | 1.059 / 99.53 | 0.900 / 95 | 0.873 / 95 |
| 7 | BC-5-surr-001 | 0.845 / 99.23 | 0.796 / 99.73 | 1.016 | 0.504 / 60.53 | 0.923 / 99.78 | 1.006 / 95 | 0.946 / 95 |
| 8 | BC-10-LOG-surr-001 | 0.961 / 83.43 | 0.667 / 96.21 | 0.914 | 0.963 / 96.40 | 1.062 / 99.00 | 0.972 / 95 | 0.916 / 95 |
| 9 | LBL-0.1 | 0.591 / 80.01 | 0.805 / 99.79 | 0.838 | 0.214 / 69.89 | 0.849 / 99.89 | 0.755 / 95 | 0.715 / 95 |
| 10 | LBL-0.1-LOG | 0.741 / 91.09 | 0.640 / 95.59 | 0.823 | 0.733 / 90.50 | 0.836 / 99.90 | 0.700 / 95 | 0.65 / 95 |
| 11 | LBL-1 | 0.677 / 98.79 | 0.598 / 98.31 | 0.762 | 0.018 / 66.87 | 0.703 / 99.46 | 0.881 / 95 | 0.823 / 95 |
| 12 | LBL-1-LOG | 0.637 / 98.64 | 0.153 / 70.05 | 0.699 | 0.709 / 87.77 | 0.698 / 99.40 | 0.921 / 95 | 0.785 / 95 |
| 13 | LBL-2 | 0.661 / 98.91 | 0.531 / 97.31 | 0.823 | 0.117 / 67.53 | 0.669 / 99.17 | 0.858 / 95 | 0.788 / 95 |
| 14 | LBL-2-LOG | 0.624 / 99.54 | 0.051 / 39.48 | 0.623 | 0.719 / 86.81 | 0.649 / 98.51 | 0.888 / 95 | 0.684 / 95 |
| 15 | LBL-0.1-surr-001 | 0.746 / 90.50 | 0.803 / 99.69 | 0.787 | 0.355 / 63.2 | 0.837 / 99.84 | 0.773 / 95 | 0.709 / 95 |
| 16 | LBL-0.1-LOG-surr-001 | 0.743 / 88.27 | 0.666 / 96.87 | 0.795 | 0.734 / 87.67 | 0.816 / 99.89 | 0.682 / 95 | 0.647 / 95 |
| 17 | TCP-0.1 | 0.832 / 97.07 | 0.820 / 99.55 | 0.978 | 0.450 / 55.12 | 1.020 / 99.83 | 0.956 / 95 | 0.872 / 95 |
| 18 | TCP-0.1-LOG | 0.873 / 98.83 | 0.617 / 94.42 | 0.880 | 0.842 / 98.50 | 0.940 / 99.80 | 0.801 / 95 | 0.742 / 95 |
| 19 | TCP-1 | 0.738 / 91.77 | 0.762 / 98.85 | 0.976 | 0.147 / 62.17 | 1.014 / 99.94 | 1.046 / 95 | 0.981 / 95 |
| 20 | TCP-1-LOG | 0.734 / 87.60 | 0.422 / 85.87 | 0.947 | 0.765 / 78.15 | 0.933 / 99.87 | 0.952 / 95 | 0.833 / 95 |
| 21 | TCP-2 | 0.715 / 92.09 | 0.721 / 98.25 | 0.941 | 0.018 / 63.44 | 1.015 / 99.95 | 1.052 / 95 | 0.974 / 95 |
| 22 | TCP-2-LOG | 0.703 / 88.30 | 0.323 / 79.59 | 0.955 | 0.743 / 76.92 | 0.938 / 99.89 | 1.007 / 95 | 0.890 / 95 |
| 23 | TCP-0.1-surr-001 | 0.845 / 97.93 | 0.809 / 99.72 | 0.977 | 0.455 / 55.86 | 0.940 / 99.97 | 0.946 / 95 | 0.860 / 95 |
| 24 | TCP-0.1-LOG-surr-001 | 0.816 / 94.02 | 0.671 / 96.30 | 0.870 | 0.816 / 92.02 | 0.917 / 99.75 | 0.800 / 95 | 0.738 / 95 |
| 25 | N | 0.483 / 99.45 | 0.560 / 99.87 | 0.541 | 0.134 / 80.28 | 0.483 / 99.63 | 0.546 / 95 | 0.5 / 95 |
| 26 | L | 0.629 / 94.36 | 0.529 / 94.50 | 1.581 | 0.496 / 94.46 | 0.893 / 94.56 | 2.673 / 95 | 0.999 / 95 |
| 27 | R | 0.024 / 99.01 | 0.149 / 79.91 | 0.660 | 0.081 / 99.36 | 0.102 / 60.73 | 1.058 / 95 | 0.999 / 95 |

Табл. 4.1. Оценка показателя H реализаций.

Строго говоря, оценка показателя Хэрста для хаотической реализации системы Лоренца, возможно, не имеет значения, поскольку последняя не обладает соответствующим видом АКФ: $C(k) = k^{2H-2}$, где $0 < H < 1$. Тем не менее, представляет определенный практически интерес изучение взаимоотношений свойств персистентности и хаотичности. Например в [12] изучается эффект аномального скэйлинга на графиках корреляционного интеграла при анализе фликкер-шума. То есть, показывается, что при определенных условиях фликкер-шум может быть неверно идентифицирован как хаотическая реализаций.

Изучая результаты измерений показателя H можно отметить, что для реализаций сетевого трафика $H > 0.5$, то есть трафик относится к классу персистентных процессов. В то же время для белого шума практически все методы показали $H \sim 0.5$, что подтверждает правильность измерений.

Заметим что показатель H связан с ранее рассмотренным при изучении АКФ параметром β следующим соотношением (4.1)

$$H = 1 - \frac{\beta}{2} \quad (4.1)$$

На рис. 4.1 приведены АКФ (4.2) фрактального гауссовского шума (ФГС), обладающего различными коэффициентами Хэрста.

$$r(k) = \frac{1}{2} [|k+1|^{2H} - |2k|^{2H} + |k-1|^{2H}], \quad k \in \mathbb{N}, \quad 0 < H < 1 \quad (4.2)$$

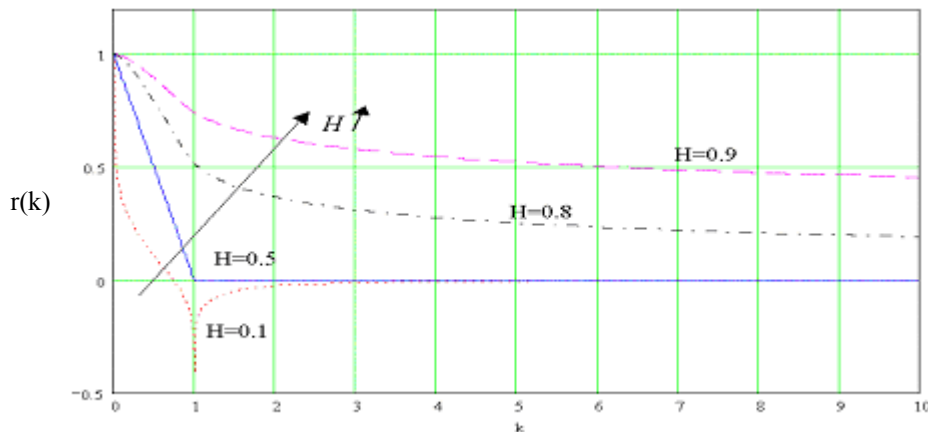


Рис. 4.1. АКФ гауссовского шума для различных значений H

Фрактальный гауссовский шум является самоподобным процессом и часто используется в качестве простейшей модели сетевого трафика.

5. Исследование реализаций методами нелинейной динамики

Нелинейная динамика становится в последнее время все более популярнее при изучении временных рядов и предоставляет обширные методы для их анализа. Известно, что хаотические системы обладают следующими основными свойствами: нелинейностью, детерминированностью, чувствительностью к начальным условиям. Кроме того, хаотическая реализация выглядит как некоторый стохастический процесс. Также, аттрактор нелинейной хаотической системы часто фрактальный. Если удастся обнаружить признаки детерминированного хаоса в трафике, мы получим новую модель трафика и новый алгоритм по его предсказанию благодаря детерминированной природе хаоса. Интересная работа была

выполнена в 2000 г. Andras Veres и Miklos Boda из Ericsson Research [15], в которой с помощью имитационного моделирования на симуляторе ns-2 показано, что модель трафика протокола TCP (использовалась версия TCP Tahoe) может быть как простым периодическим процессом, так и при некоторых условиях обладать сложным поведением, согласующимся с понятием *детерминированного хаоса*. В частности исследователями была получена траектория системы в фазовом пространстве, отнесенная ими к классу странных аттракторов. Аттрактор – предельное множество траекторий в фазовом пространстве системы, к которому стремятся все траектории из некоторой окрестности этого множества. Также можно отметить существование хаотических моделей сетевого трафика [16]. В них изучаются свойства хаотических отображений (chaotic maps) типа

$$\begin{aligned}x_{n+1} &= f_1(x_n), y_n = 0, (0 < x_n < d) \\x_{n+1} &= f_2(x_n), y_n = 1, (d < x_n < 1)\end{aligned}$$

где $f_1(.)$ и $f_2(.)$ – некоторые функции, для которых выполняется условие чувствительности к начальным условиям. Кроме того, при построении модели предполагается, что источник трафика находится в активном или пассивном состоянии в зависимости от того x_n больше или меньше некоторого предела d . Адекватность данной модели подтверждается присущими ей свойствами медленно убывающей зависимости и распределением с тяжелым хвостом. Для одного из наборов параметров хаотического отображения корреляционная размерность аттрактора оценена как 0.91.

Автор полагает, что исследование сетевого трафика методами нелинейной динамики может оказать пользу при изучении его характеристик, а также разработке и улучшению алгоритмов его обработки.

5.1. Концепция суррогатных данных

Одной из основных задач настоящего анализа является проверка гипотезы о том, является ли сетевой трафик линейно коррелированным шумом или детерминированным процессом. Для этой цели воспользуемся концепцией, так называемых, суррогатных данных. Суть данного метода заключается в следующем: для исследуемой временной реализации процесса создается ансамбль (обычно пять или десять) “суррогатных” реализаций, являющихся случайными по своей природе, но обладающих точно такими же автокорреляционной функцией и дисперсией как у исходной реализации. На практике это достигается Фурье-преобразованием исходной реализации, изменением случайным образом фаз и обратным Фурье-преобразованием. Далее сравниваются результаты вычисления различных статистик (размерностей, корреляционных интегралов и пр.) для оригинального процесса и его ансамбля суррогатных

реализаций. В случае если результаты значительно различаются нулевая гипотеза о том, что исходная реализация является линейно-коррелированным шумом может быть отклонена.

Для каждой из реализаций трафика составим по 5 суррогатных реализаций и будем использовать их в дальнейшем анализе. Для примера на рис. 5.1 приведена одна из суррогатных реализаций для реализации TCP-1, а также ее гистограмма и АКФ.

Можно заметить, что при суррогатная реализация имеет такую же АКФ, что и у исходной, но отличное распределение, которое теперь более похоже на нормальное.

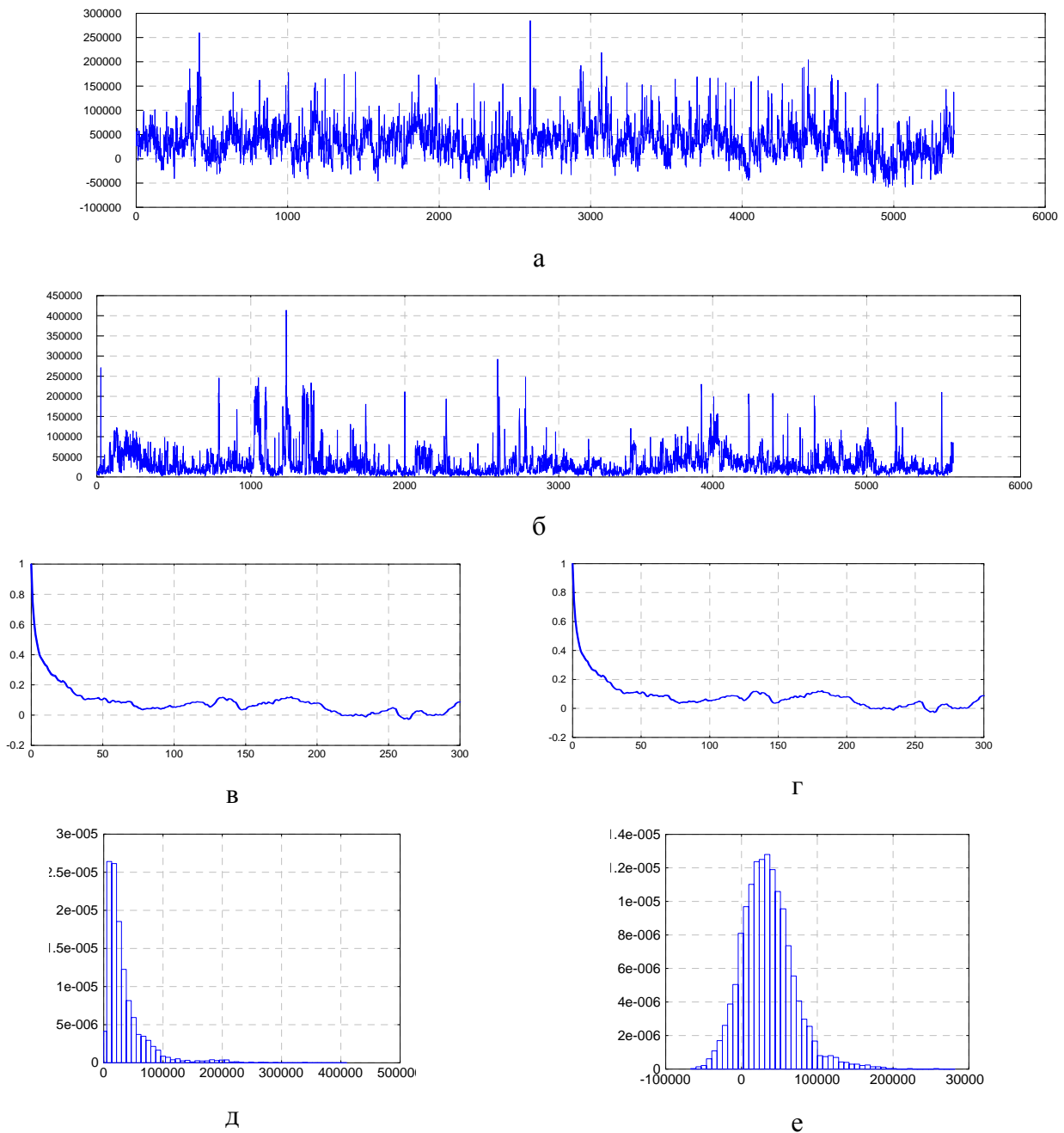


Рис. 5.1. а) Суррогатная реализация TCP-1-SURR-001;
 б) Исходная реализация TCP-1;
 в) АКФ исходной реализации TCP-1;
 г) АКФ суррогатной реализации TCP-1-SURR-001;
 д) Гистограмма исходной реализации TCP-1;
 е) Гистограмма суррогатной реализации TCP-1-SURR-001.

Произведенная ранее (см. табл. 4.1) оценка показателя Херста показывает, что суррогатные реализации обладают примерно таким же уровнем персистентности, что и оригинальные реализации.

5.2. Идея реконструкции аттрактора

Анализ сетевого трафика фактически сводится к задаче обработки временного ряда. Как уже отмечалось теория нелинейной динамики в свою очередь предоставляет широкие возможности для изучения идентификации и прогнозирования временных рядов, обладающих некоторыми специфическими свойствами. Одной из ключевых концепций теории нелинейной динамики является использование теоремы Такенса о погружении аттрактора в пространствах различных размерностей. Под аттрактором здесь понимается предельное множество траекторий в фазовом пространстве системы, к которому стремятся все траектории из некоторой окрестности этого множества.

Данная методика позволяет восстановить параметры динамической системы по единственной реализации (временному ряду) с помощью изучения траектории системы в m -мерном фазовом пространстве, координатами которого являются компоненты следующего вектора: $z_i^m = \{X_i, X_{i+\tau}, \dots, X_{i+(m-1)\tau}\}$, где τ – временной сдвиг. Данная операция называется погружением аттрактора в пространство размерности m . Результатом успешного погружения является выявление определенных закономерностей в поведении траектории системы в пространстве данной размерности.

5.3. Ложные ближайшие соседи (FNN)

Метод False Nearest Neighbors (FNN) служит для определения минимальной приемлемой размерности пространства погружения. Его принцип понятен на интуитивном уровне. Допустим $z_i^{(m)}$ и $z_j^{(m)}$ – два близких соседа в реконструкции размерности m , а $z_i^{(m+1)}$ и $z_j^{(m+1)}$ соответствуют им в реконструкции $m+1$. Если мы имеем дело с действительно близкими соседями, то они, как правило, близки в обеих реконструкциях. В то же время если соседи, близкие в реконструкции m ($\|z_i^{(m)} - z_j^{(m)}\|$ – мало), становятся отдаленными в реконструкции $m+1$ ($\|z_i^{(m+1)} - z_j^{(m+1)}\|$ – велико), то они называются ложными близкими соседями или false nearest neighbors (FNN). Если теперь увеличивать m и оценивать количество FNN, то при достижении нужной размерности, при которой достигается правильная реконструкция, это количество резко уменьшается. Очевидно, что по графику зависимости количества “ложных соседей” от размерности пространства погружения (по его минимуму или убыванию в нуль) можно сделать вывод о минимально

возможной размерности фазового пространства. Построим графики FNN для исследуемых реализаций, суррогатных реализаций, а также для белого шума и детерминированного хаоса (в виде системы Лоренца).

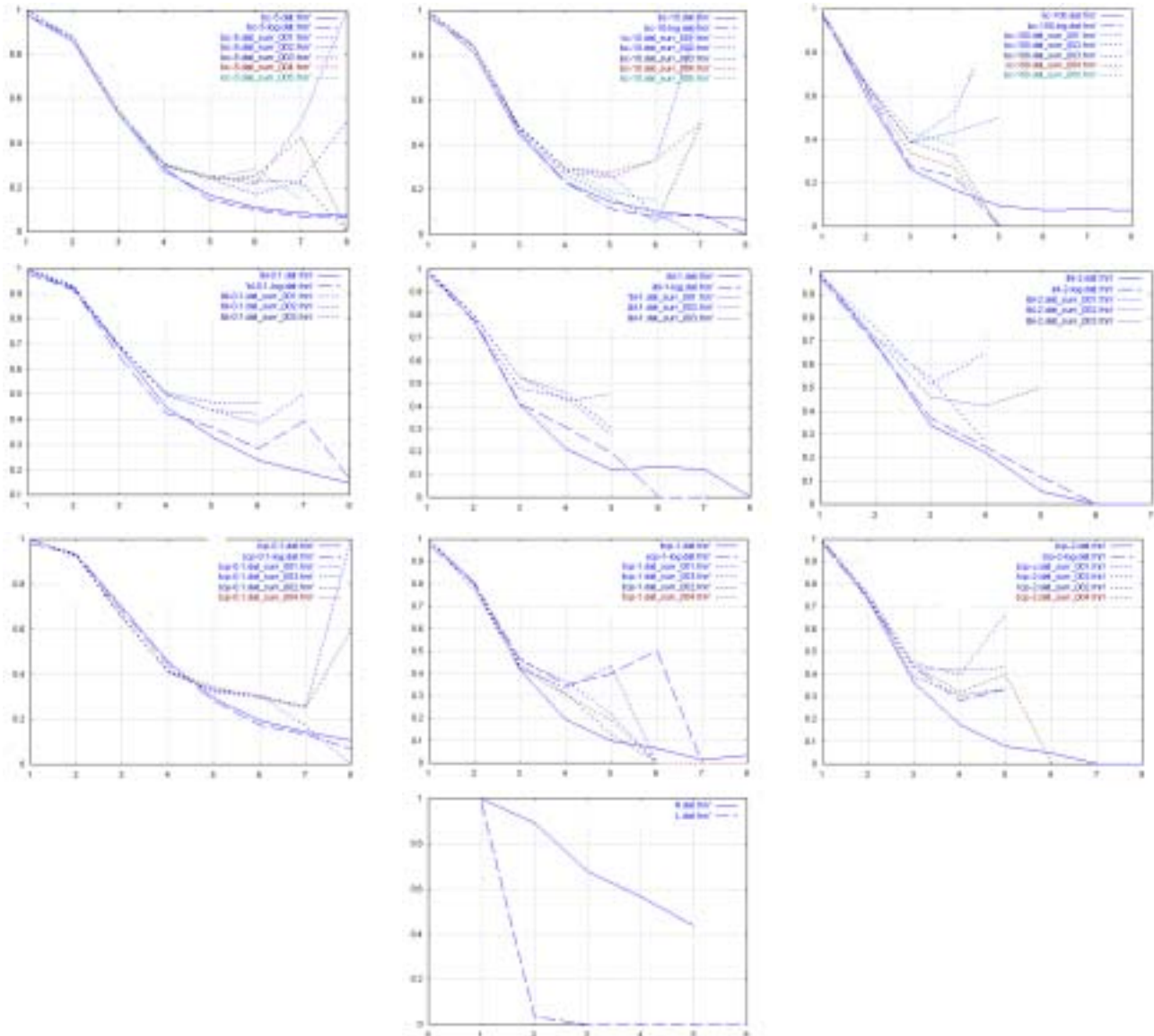


Рис. 5.2. Зависимости FNN (Для реализаций трафика временной сдвиг $d=5$, для ,белого шума $d=1$).

Изучая зависимости FNN можно сделать вывод о том, что наиболее приемлемой размерностью реконструкции для реализаций BC-5, LBL-0.1, TCP-0.1 является размерность ~ 8 , а для реализаций BC-10, BC-100, LBL-1, LBL-2, TCP-1 и TCP-2 размерность ~ 5 . При этом характерно, что зависимости количества ложных соседей для суррогатных реализаций заметно отличаются от аналогичных для исходных реализаций и больше похожи на зависимость FNN для белого шума (см. последний график). Данное замечание говорит в пользу того, что возможно свойства изучаемых реализаций трафика обусловлены не только корреляциями, свойственными $1/f^\alpha$ – процессам.

В то же время графики FNN для оригинальных агрегированных реализаций трафика занимают некоторое промежуточное положение (по степени убывания) между чисто хаотическим процессом и число случайным процессом (белым шумом). Данное наблюдение может служить в пользу классификации сетевого трафика как некоторого детерминированного (возможно хаотичного) процесса.

5.4. Вычисление корреляционного интеграла

Одним из основных инструментов, используемых для изучения фазового портрета системы служит понятие корреляционного интеграла, который позволяет оценить размерность D_2 аттрактора, погруженного в пространство размерности m . Вычисление размерности аттрактора по временному ряду в свою очередь представляет интерес, поскольку позволяет оценить минимальное число существенных динамических переменных, необходимых для описания наблюдаемого процесса. Наиболее популярный алгоритм для расчета корреляционного интеграла был предложен П. Грассбергом и И. Прокаччиа и основывается на соотношении

$$C(\varepsilon, m) = \frac{\text{число пар } s \text{ с } \|z_i - z_j\| < \varepsilon}{\text{общее число пар } z_i, z_j}$$

Здесь z_i, z_j – векторы координат точек в фазовом пространстве размерности m . При этом справедливо соотношение

$$C(m, \varepsilon) \sim \varepsilon^{D_2}$$

позволяющее оценить размерность D_2 аттрактора по наклону наиболее линейного участка графика

$$\log C(\varepsilon, m) \cong -D_2 \log \varepsilon + const$$

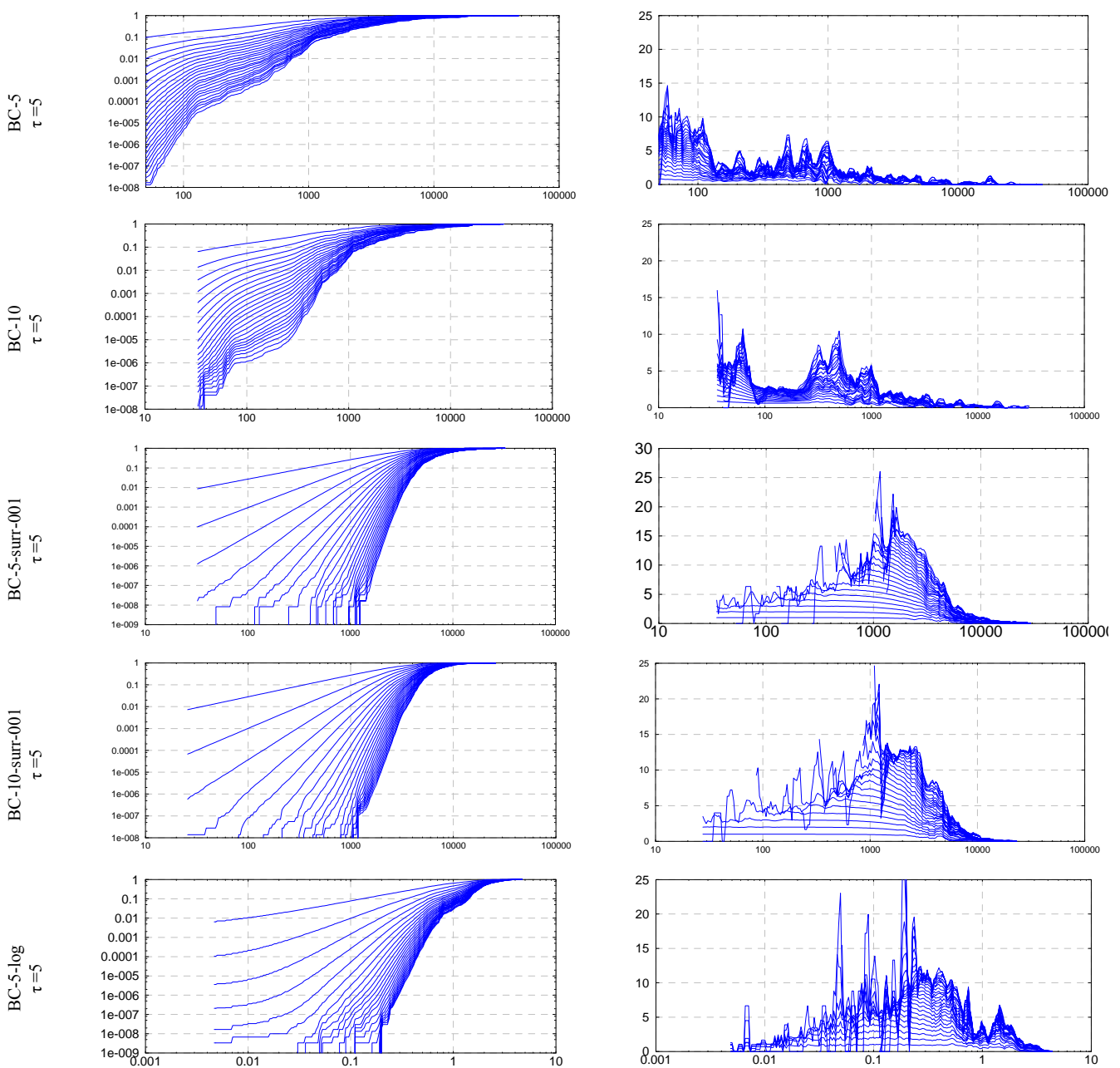
При увеличении m оценка D_2 должна стремиться к истинному значению корреляционной размерности аттрактора.

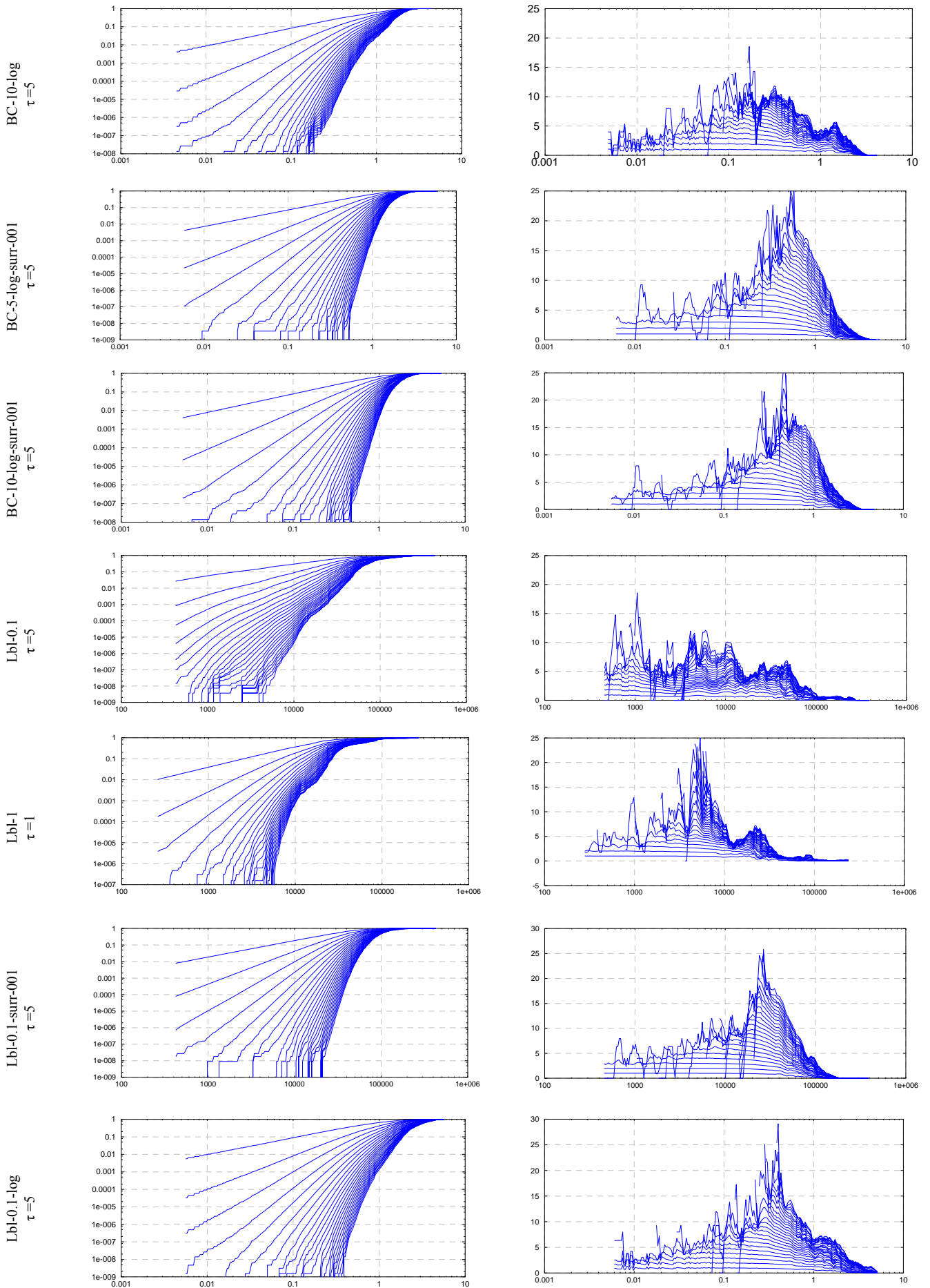
Если мы имеем дело с данными, которые по определению не должны давать достоверных результатов при такой обработке, то корреляционный интеграл вообще может не иметь линейного участка (хотя может и иметь), а его наклон обычно возрастает с увеличением m . В ряде случаев можно показать, что этот наклон будет линейным, т.е.

$$S(\varepsilon, m) \cong \varphi_0(\varepsilon)m + S_0(\varepsilon)$$

В других случаях, особенно при обработке данных ограниченной длины и небольших значениях m , зависимость чаще всего более сложная, однако в большинстве случаев она удовлетворительно аппроксимируется линейной зависимостью. Исследование зависимости наклона от m может

помочь разрешению вопроса о том, позволяет ли реконструкция судить о свойствах аттрактора или нет (по самым разным причинам – недостаточно данных, слишком сложная динамика, данные не детерминированные а случайные и пр.). Специальное исследование [17] показало, что $\varphi_0 = 0.05 \dots 0.1$ является некоторым пороговым значением. При меньших значениях методы нелинейной динамики обычно применимы, при меньших – нет. Таким образом, с помощью корреляционного интеграла можно решить вопрос об эффективности алгоритмов нелинейной динамики при обработке данного временного ряда или же использовать статистические подходы или как-либо комбинировать динамические и статистические методы. Ниже приведем результаты расчета корреляционного интеграла для исследуемых временных рядов.





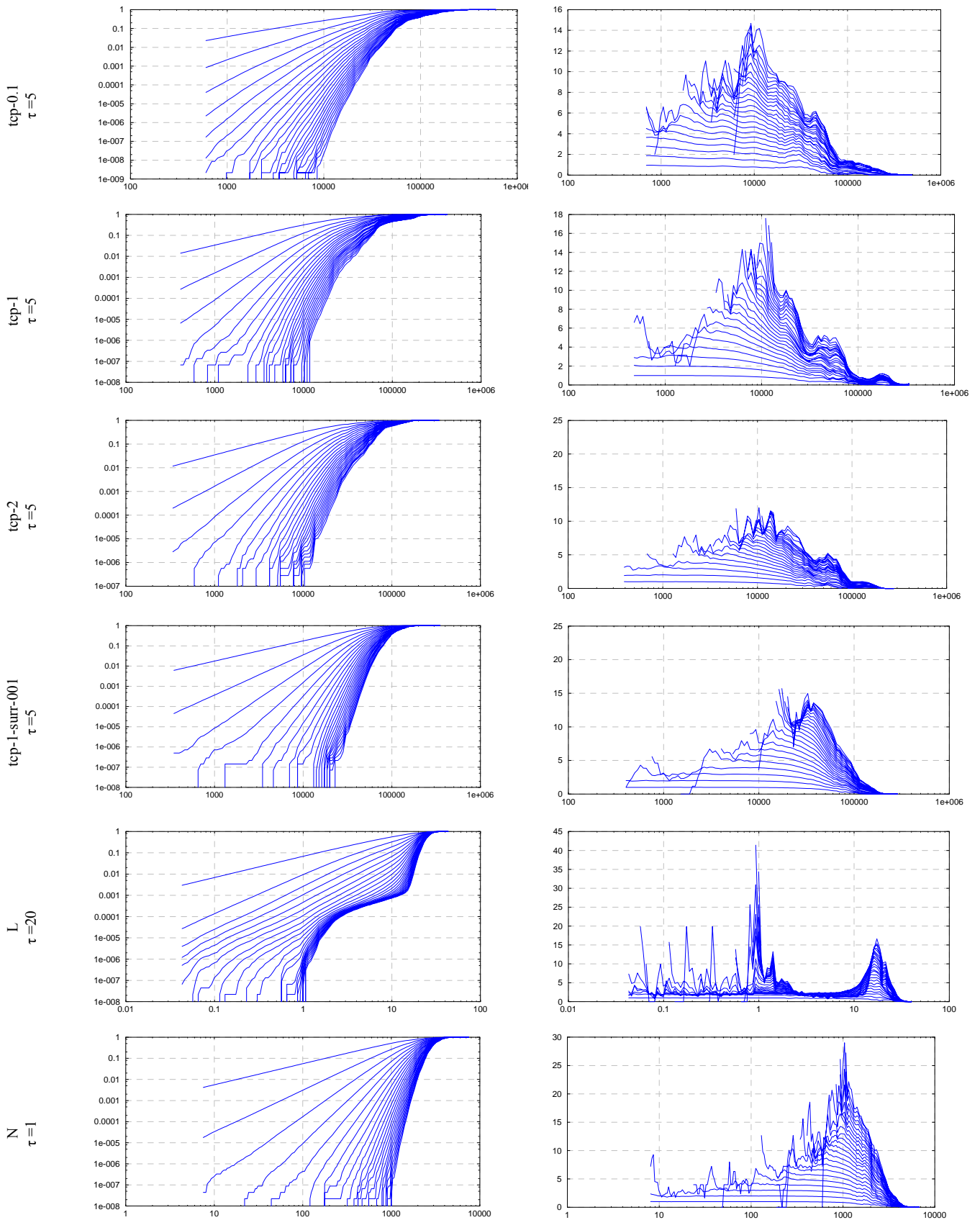


Рис.5.3. Справа: графики семейства зависимостей корреляционного интеграла. По оси ординат отложена величина $\log C(\epsilon)$, по оси абсцисс - $\log(\epsilon)$. Параметром семейства является размерность вложения m . Сверху – вниз m увеличивается от 1 до 25. Слева – семейство зависимостей локальных наклонов D_2 корреляционного интеграла от $\log(\epsilon)$. Параметром семейства является m (снизу вверх m увеличивается от 1 до 25).

Основные наблюдения, которые можно сделать изучая зависимости корреляционного интеграла состоят в следующем:

- для всех оригинальных временных реализаций трафика наблюдается характерный изгиб кривых корреляционного интеграла. Наиболее выражен изгиб у реализаций ВС-5 (на интервале 200-800 Байт/с), ВС-10 (на интервале 100-300 Байт/с), а также у хаотической реализации системы Лоренца;
- кривые корреляционных интегралов для суррогатных реализаций больше напоминают кривые корреляционного интеграла для белого шума, чем аналогичные для хаотической реализации. Это наблюдение отвергает гипотезу о классификации сетевого трафика как некоторого шумового процесса, обладающего коррелированной структурой. Другими словами, в трафике, возможно, присутствует некоторая детерминированная, однако достаточно слабая составляющая (не исключено, что с некоторой степенью хаотичности).
- если наблюдаемое насыщение наклона кривых корреляционного интеграла для реализации ВС-10 не случайно, оно может свидетельствовать о размерности соответствующего аттрактора ~ 2.5 .
- для хаотической реализации системы Лоренца наблюдается характерное насыщение наклонов корреляционного интеграла ($D_2 \sim 2.5$), в то время как для случая белого шума насыщения вообще не происходит, что хорошо согласуется с имеющейся теоретической базой.

5.5. Проверка гипотезы о статистической независимости (BDS-тест)

Для проверки гипотезы о статистической независимости последовательных членов временного ряда, соответствующего сетевому трафику, воспользуемся так называемым BDS – тестом. Данный тест основан на свойствах корреляционного интеграла и требует вычисления следующей статистики:

$$w_{m,n}(\varepsilon) = \sqrt{n-m+1} \frac{C(\varepsilon, m, n) - C^m(\varepsilon, 1, n)}{\sigma_{m,n}(\varepsilon)},$$

где n – объем анализируемого участка временного ряда, $C(\varepsilon, m, n)$ – корреляционный интеграл. Считается, что нулевая гипотеза о независимости и одинаковости распределения членов ряда отвергается при уровне значимости 5% если $|w_{m,n}(\varepsilon)| < 1.96$ для рекомендованного диапазона изменения ε . Для вычислений воспользуемся программной реализацией W. Davis Dechert. Поскольку программа позволяет работать только с числом отсчетов, не превышающих 7 500, проанализируем с ее помощью первые 7 500 отсчетов исследуемых реализаций. Обычно

рекомендуется использовать значения $\varepsilon = 1.5 \cdot \left(\frac{\sigma}{s}\right) \dots 2 \cdot \left(\frac{\sigma}{s}\right)$, где σ – среднее квадратическое отклонение, s – максимальный размах значений ряда. Некоторые результаты расчетов программы даны в приложении. Анализируя полученные численные результаты можно сделать вывод о том, что для хаотической реализации L, реализаций трафика BC-5, BC-10, LBL-0.1, LBL-1, TCP-0.1, TCP-1, их логарифмов, а также суррогатных реализаций гипотеза о статистической независимости членов ряда отвергается для всех рекомендованных значений ε и значений размерности погружения $m=2..10$. В то же время для белого шума нулевую гипотезу нельзя отвергнуть, поскольку $|w_{m,n}(\varepsilon)| < 1.96$ для всех изученных наборов параметров. Ряды BC-100, LBL-2, TCP-2 не анализировались, поскольку они имеют относительно небольшое количество отсчетов.

6. Исследование перепадов в реализациях сетевого трафика

Простейшее исследование предсказуемости можно провести следующим образом. Допустим, имеется $X=(X_1, X_2, \dots, X_n)$ – исследуемый агрегированный временной ряд. Построим вспомогательную последовательность $Y=(Y_1, Y_2, \dots, Y_n)$ такую, что $Y_i = (1 | X_{i+1} > X_i)$, $Y_i = (0.5 | X_{i+1} < X_i)$ и $Y_i = (2 | X_{i+1} = X_i)$. Теперь исследуем плотность распределения (гистограмму) последовательности Y . А именно рассмотрим коэффициент G , представляющий собой отношение частоты появления 0.5 к частоте появления 1 для всех изучаемых реализаций. Результаты вычислений сведены в табл. 6.1.

| № | Реализация | G |
|----|------------------|-------|
| 1 | BC-5 | 0.981 |
| 2 | BC-10 | 0.998 |
| 3 | BC-100 | 1.026 |
| 5 | LBL-0.1 | 0.977 |
| 6 | LBL-1 | 0.907 |
| 7 | LBL-2 | 0.969 |
| 8 | TCP-0.1 | 0.989 |
| 9 | TCP-1 | 0.989 |
| 10 | TCP-2 | 1.012 |
| 11 | BC-5-surr-001 | 1.006 |
| 12 | BC-10-surr-001 | 0.995 |
| 13 | BC-10-surr-001 | 1 |
| 14 | LBL-0.1-surr-001 | 0.997 |
| 15 | LBL-1-surr-001 | 1.035 |
| 16 | LBL-2-surr-001 | 0.969 |
| 25 | N | 1.006 |
| 26 | L | 0.963 |

Табл. 6.1. Значения коэффициента G для реализаций трафика

Изучая результаты по табл. 6.1. можно заметить, что количество в реализациях трафика количество перепадов $X_{i+1} > X_i$ примерно столько же, сколько перепадов $X_{i+1} < X_i$ и соответствует полученным значениям для белого шума и системы Лоренца. Каких-либо особенностей или тенденций выявить не удастся.

7. Анализ полученных результатов и основные выводы

Подведем итоги произведенного анализа сетевого трафика и сформируем основные выводы:

- Произведен анализ трех реализаций сетевого трафика. Две из них (*LBL-PKT-5.TCP* и *LBL-TCP-3*) относятся к протоколу TCP, и одна (*BC-Oct89Ext.TL*) – к протоколу Ethernet.
- Анализ экспериментальных плотностей распределения (гистограмм) данных реализаций позволяет сделать вывод о том, что они обладают схожими особенностями. В частности, наблюдается плавное уменьшение частот появления дэйтаграмм, длительности которых находится в интервале 0 ... 512 байт, далее следует характерный всплеск в районе 512 байт практически на 2 порядка, после чего - резкое уменьшение (на 4 порядка) частот появления дэйтаграмм, длительность которых > 512 байт (> 600 для Ethernet-трафика). Так как дэйтаграммы протокола TCP инкапсулируются в Ethernet-кадры, можно сделать вывод, что распределение длин Ethernet-кадров определяются распределением TCP-дэйтаграмм, а не наоборот. Гистограмма Ethernet-кадров наследует особенности гистограммы TCP-дэйтаграмм, несмотря на то, что они относятся к разным реализациям.
- Для приведения исходных реализаций к эквидистантной шкале по оси времени необходимо произвести агрегирование. Агрегирование каждой из реализаций производится на трех уровнях 5, 10 и 100 сек для реализации *BC-Oct89Ext.TL* и 0.1, 1 и 2 сек для реализаций *LBL-PKT-5.TCP* и *LBL-TCP-3*. Меньших уровней агрегирования достичь не удастся, поскольку при этом значительно увеличиваются ошибки усреднения. С другой стороны большие уровни агрегирования не имеет смысла исследовать из-за сокращения длительности агрегированных временных рядов.
- В промежуток примерно с 11:00 до 18:00 в реализации *BC-Oct89Ext.TL* наблюдается минимальная неравномерность поступления Ethernet-кадров (см. рис. 2.3). Этот факт может объясняться высокой активностью использования сетевых ресурсов в указанной

- время суток, что в свою очередь, скорее всего, связано с продолжительностью рабочего дня.
- Увеличение уровня агрегирования реализаций приводит к сглаживанию. При этом СКО уменьшается на фоне практически неизменного среднего (см. табл. 3.1).
 - Ряды ВС-х, LBL-х а также TCP-х скорее всего подчиняются некоторому распределению с так называемым “тяжелым хвостом” (см. рис. 3.1).
 - Увеличение уровня агрегации приводит к изменению параметров “тяжелохвостого” распределения. Более точный ответ может дать исследование распределений методом максимального правдоподобия.
 - Логарифмирование рядов приводит к нормализации их распределений (см. рис. 3.1)..
 - Гистограмма хаотической реализации L имеет два характерных максимума (при значениях ~ -9 и ~ 9). Аналогичные максимумы наблюдаются в реализации ВС-10-LOG и (менее выраженные) в LBL-0.1-LOG (см. рис. 3.1).
 - АКФ рядов LBL-0.1 и LBL-0.1-LOG имеют отчетливую периодическую структуру (с частотой ~ 5 Гц). Менее выраженные периодичности (пульсации) АКФ имеются в рядах ВС-5-LOG и ВС-10-LOG, а также в рядах TCP-0.1 и TCP-0.1-LOG (см. рис. 3.2).
 - Результаты регрессионного анализа АКФ подтверждают присутствие МУЗ во всех исследуемых реализациях трафика.
 - Изучение энергетических спектров реализаций трафика (рис. 3.5) позволяет сделать вывод о присутствии сильной гармонической компоненты ~ 5 Гц в реализациях LBL-0.1, LBL-0.1-LOG. Данная гармоника также присутствует (хотя и менее существенна) в реализациях TCP-0.1 и TCP-0.1-LOG. При увеличении уровня агрегирования гармоника исчезает, так как верхняя частота спектра уменьшается и становится ниже 5 Гц. Очевидно, периодичность АКФ реализаций LBL-0.1, LBL-0.1-LOG обусловлена именно присутствием в их спектрах гармоники 5 Гц. Также в реализациях ВС-5-LOG и ВС-10-LOG наблюдается несколько сильных гармонических компонент. Причины появления подобных компонент в спектрах трафика неясны. Обнаруженное явление выявляет присутствие регулярной детерминированной составляющей в агрегированном сетевом трафике, что может быть интересным при решении задач прогнозирования телетрафика.
 - Можно отметить преобладание в полученных спектрах низкочастотных компонент, что согласуется с понятием фликкер-шума.
 - Изучая результаты измерений показателя H методами *Aggregate Variance*, *R/S*, *Periodogram*, *Absolute Moments*, *Variance of Residuals*, *Abry-Veitch Estimator*, *Whittle Estimator* можно отметить, что для всех реализаций сетевого трафика $H > 0.5$, то есть

трафик относится к классу персистентных процессов. В то же время для белого шума практически все методы показали $H \sim 0.5$, что подтверждает правильность измерений (см. табл. 4.1).

- Усредненное значение H для сетевого трафика $H \sim 0.8$. Зависимости коэффициента H от уровня усреднения не выявлено.
- Существуют модели сетевого трафика, использующие понятие детерминированного хаоса [15]. При эмуляции работы нескольких TCP-сессий по одному каналу обнаружены режимы детерминированного хаоса [16].
- Зависимости количества ложных соседей FNN для суррогатных реализаций заметно отличаются от аналогичных для исходных реализаций и больше похожи на зависимость FNN для белого шума (см. 5.2). Данное замечание говорит в пользу того, что возможно свойства изучаемых реализаций трафика обусловлены не только корреляциями, свойственными $1/f^\alpha$ – процессам.
- Графики FNN (рис. 5.2) для агрегированных реализаций трафика занимают некоторое промежуточное положение (по степени убывания количества ложных соседей) между чисто хаотическим процессом и число случайным процессом (белым шумом). Данное наблюдение может служить в пользу классификации сетевого трафика как некоторого детерминированного (возможно хаотичного) процесса.
- Кривые корреляционного интеграла для изучаемых реализаций трафика имеют характерный изгиб. Последний наиболее выражен у реализаций ВС-5 (на интервале 200-800 Байт/с), ВС-10 (на интервале 100-300 Байт/с), а также у хаотической реализации системы Лоренца.
- Кривые корреляционных интегралов для суррогатных реализаций больше напоминают кривые корреляционного интеграла для белого шума, чем аналогичные для хаотической реализации. Это наблюдение отвергает гипотезу о классификации сетевого трафика как некоторого шумового процесса, обладающего коррелированной структурой. Другими словами, в трафике, возможно, присутствует некоторая детерминированная, однако достаточно слабая составляющая (не исключено, что с некоторой степенью хаотичности).
- Наблюдаемое насыщение наклона кривых корреляционного интеграла для реализации ВС-10 может свидетельствовать о размерности соответствующего аттрактора ~ 2.5 .
- Для хаотической реализации системы Лоренца наблюдается характерное насыщение наклонов корреляционного интеграла ($D_2 \sim 2.5$), в то время как для случая белого шума насыщения вообще не происходит, что хорошо согласуется с имеющейся теоретической базой.

- Результаты вычисления BDS- статистики позволяют сделать вывод о том, что для хаотической реализации L , реализаций трафика BC-5, BC-10, LBL-0.1, LBL-1, TCP-0.1, TCP-1, их логарифмов, а также суррогатных реализаций гипотеза о статистической независимости членов ряда отвергается для всех рекомендованных значений ϵ и значений размерности погружения $m=2..10$. Из всех изучаемых случаев данная гипотеза не отвергается только для белого шума, что только подтверждает достоверность результатов анализа.

Таким образом, подводя черту, можно классифицировать сетевой трафик как некоторый сложный, похожий на случайный, однако предсказуемый процесс. В этой связи увеличение эффективности обработки трафика может основываться на алгоритмах прогнозирования трафика. Автор полагает, что использование техники прогнозирования позволит решить ряд задач в области обеспечения заданного уровня качества обслуживания QoS. Действительно, зная уровень трафика, который будет проходить через некоторый узел в сети в некоторый момент в будущем мы можем попытаться обработать его наиболее эффективно.

Дальнейшие свои исследования автор связывает с поиском наиболее подходящего алгоритма прогнозирования трафика. Поскольку алгоритмы нелинейной динамики дали некоторые результаты при анализе трафика, возможности использования методов теории хаоса при прогнозировании также представляется перспективным направлением. Хотя также представляет интерес изучение предсказательных свойств в отношении прогнозирования сетевого трафика и “классических” ARFIMA моделей.

Литература

1. <http://ita.ee.lbl.gov/> -Internet traffic archive
2. H. J. Fowler, W. E. Leland. [Local Area Network Traffic Characteristics, with Implications for Broadband Network Congestion Management](#), *IEEE JSAC*, 9(7), September 1991, pp. 1139-1149.
3. W. Leland, M. Taqqu, W. Willinger, and D. Wilson. On the Self-Similar Nature of Ethernet Traffic (Extended Version), *IEEE/ACM Transactions on Networking*, 2(1), February 1994, pp. 1-15.
4. V. Paxson, S. Floyd [Wide-Area Traffic: The Failure of Poisson Modeling](#), , *IEEE/ACM Transactions on Networking*.
5. R.Boel, Tite Kubushihi. Analysis, modeling and prediction of Internet traffic. SYSTeMS Group, Universiteit Gent (Belgium).
6. Пакет программ для анализа временных реализаций методами нелинейной динамики TISEAN v2.1: http://lists.mpipks-dresden.mpg.de/~tisean/TISEAN_2.1 -
7. Голяндина Н.Э., Некруткин В.В., Браулов К.А. Метод “Гусеница”-SSA: анализ временных рядов. 2002
8. Анищенко В.С. Знакомство с нелинейной динамикой. Лекции соросовского профессора. Саратов. ГосУНЦ “Колледж” 2000.
9. Кузнецов С. П. Динамический хаос (курс лекций).--М.: [Издательство Физико-математической литературы](#), 2001.--296 с
10. Программа FRACTAN v4.4, предназначенная для фрактального анализа временных реализаций: <http://impb.psn.ru/~sychyov/soft.shtml>
11. Малинецкий Г.Г., Потапов А.Б. Современные проблемы нелинейной динамики. М.: Едиториал УРСС, 2002. – 360с.
12. Theiler J. Some comments on the correlation dimension of $1/f^{\alpha}$ noise // *Phys. Lett. A*. **155** (1991). P480-493.
13. G. Olowoyeye, B. Kim, K. Chandra. Modelling Spectral Features in TCP Traffic
14. K. Chandra, C. You, G. Olowoyeye and C. Thompson, Non-linear Time-Series Models of Ethernet Traffic. Submitted to INFOCOM'99, July 1998.
15. A. Veres, M. Boda. The Chaotic Nature of TCP Congestion Control // *Proceedings of IEEE INFOCOM'2000*, March 2000.
16. A. Erramilli, R.P. Singh, P. Pruthi. An Application of Deterministic Chaotic Maps to Model Packet Traffic. *Queueing Systems*, 20(3):171-206, 1995.

17. Potapov A., Kurths J. Correlation integral as a tool for distinguishing between dynamics and statistics in time series data // *Physics D*. **120** (1998). P. 369-385.
18. Kugiumtzis D. Boudourides M. Chaotic Analysis of Internet Ping Data: Just a Random Generator? // SOEIS meeting at Bielefeld, March 27-28, 1998.
19. Beran J. Statistical Methods for Data with Long-Range Dependence. // *Statistical Science*, Volume 7, Issue 4 (Nov., 1992), 404-416.
20. W. A. Brock, W. D. Dechert, J.A. Scheinkman and B. LeBaron. A test for independence based on correlation dimension, *Econometric Reviews* 15: 197-235, 1996
21. M. B. Kennel, R. Brown and H. D. I. Abarbanel, Determining embedding dimension for phase-space reconstruction using a geometrical construction, *Phys. Rev. A* 45, 3403 (1992)
22. J.B. Gao, Y. Cao, J-M. Lee. Principal component analysis of $1/f^\alpha$ noise. *Physics Letters A* 314 (2003) 392-400.
23. R. Hegger, H. Kantz, T. Schreiber. Practical implementation of nonlinear time series methods: The TISEAN package. [ArXive:chaos-dyn/9810005v1](https://arxiv.org/abs/chaos-dyn/9810005v1) 30Sep 1998.
24. J.-P. Eckmann, S Kamphorst and D. Ruelle. "Recurrence plots of dynamical systems", *Europhys. Lett.* 4. 973-977 (1987)
25. Selfis v0.1b – программа для анализа экспоненты Хэрста разработки Thomas Karagiannis: <http://www.cs.ucr.edu/~tkarag/>
26. Программа Visual Recurrence Analysis (VRA), ver. 4.2 разработки Евгения Кононова.
27. Программа "Caterpillar" реализующая алгоритм сингулярного спектрального анализа SSA <http://www.gistatgroup.com/cat/>