УДК 621. 391

УПРОЩЕННЫЙ РАСЧЕТ КОЭФФИЦИЕНТА ХЕРСТА МЕТОДОМ R/S-АНАЛИЗА

Ложковский А.Г.

Одесская национальна академия связи им. А.С. Попова,
65029, Украина, г. Одесса, ул. Кузнечная, 1.
aloshk@onat.edu.ua

СПРОЩЕНИЙ РОЗРАХУНОК КОЕФІЦІЄНТА ХЕРСТА МЕТОДОМ R/S-АНАЛІЗУ

Ложковський А.Г.

Одеська національна академія зв'язку ім. О.С. Попова,
65029, Україна, м. Одеса, вул. Кузнечна, 1.
aloshk@onat.edu.ua

SIMPLIFIED CALCULATION OF THE HERST COEFFICIENT BY THE R/S-ANALYSIS METHOD

Lozhkovskyi A.G.

O.S. Popov Odessa national academy of telecommunications,
1 Kuznechna St., Odessa, 65029, Ukraine.
aloshk@onat.edu.ua

Аннотация. Трафик пакетных сетей связи или распределение количества пакетов в единицу времени описывается самоподобным (*self-similarity*) случайным процессом с параметром Херста около 0,65…0,8. Обычно расчет коэффициента самоподобия Херста реального трафика выполняется методом абсолютных моментов. В статье предложен упрощенный метод расчета коэффициента самоподобности трафика пакетных сетей связи. Упрощение состоит в том, что предлагается производить расчет не по всем возможным значениям R/S-статистики (регрессию), а только по двум из них. Погрешность расчета при этом не превышает 2…5%.

Ключевые слова: телекоммуникационные системы и сети, методы расчета и проектирования, самоподобный трафик, коэффициент самоподобия Херста.

Анотація. Трафік пакетних мереж зв'язку або розподіл кількості пакетів в одиницю часу описується самоподібним (*self-similarity*) випадковим процесом з параметром Херста близько 0,65…0,8. Зазвичай розрахунок коефіцієнта самоподібності Херста реального трафіка виконується методом абсолютних моментів. У статті запропонований спрощений метод розрахунку коефіцієнта самоподібності трафіка пакетних мереж зв'язку. Спрощення полягає в тому, що пропонується здійснювати розрахунок не за усіма можливими значеннями R/S-статистики (регресія), а тільки за двомах із них. Похибка розрахунку при цьому не перевищує 2…5%..

Ключові слова: телекомунікаційні системи та мережі, методи розрахунку та проектування, самоподібний трафік, коефіцієнт самоподібності Херста.

Abstract. The traffic of packet communication networks or the distribution of the number of packets per time unit is described by a self-similar random process with a Hurst parameter of about 0,65…0,8. Usually, the calculation of the self-similarity coefficient of Hurst real traffic is performed by the method of absolute moments. The paper offers a simplified method for calculating the self-similarity coefficient of packet network traffic. Simplification is that it is proposed to perform a calculation not for all possible values of R/S-statistics (regression), but only for two of them. The error in the calculation does not exceed 2…5%.

Key words: telecommunications systems and networks, methods of calculation and design, self-similar traffic, self-similarity coefficient of Hurst.

Трафик в пакетных сетях или распределение количества пакетов в единицу времени принято считать, что хорошо описывается самоподобным (*self-similarity*) случайным процессом с параметром Херста около 0,65…0,8 [1]. Трафик пакетных сетей связи можно представить как процесс поступления пакетов, описываемый временным рядом количества пакетов по фиксированным отрезкам времени (мгновенная интенсивность поступления пакетов). Статистический *R*/*S*-анализ, созданный Г. Херстом – это метод анализа динамики временных последовательностей. *R*/*S*-анализом также считают совокупность статистических приёмов и методов анализа временных рядов (преимущественно финансовых), позволяющих определить некоторые важные их характеристики, такие как наличие непериодических циклов, памяти.

Метод Херста позволяет выявить в статистических данных пакетного трафика такие его свойства, как кластерность, тенденцию следовать по направлению тренда (персистентность) и быструю перемежаемость последовательных значений интенсивности трафика (всплески интенсивности, приводящие к пачечности), сильное последействие, сильную память, фрактальность (самоподобность), наличие периодических и непериодических циклов (из-за особенностей используемых протоколов передачи).

Однако, существующие методы расчета коэффициента Херста являются достаточно трудоемкими, что затрудняет их использование в условиях реального процессорного времени обработки параметров трафика при выявлении его самоподобных ствойств [2].

Целью статьи является упрощение метода расчета коэффициента самоподобности трафика пакетных сетей связи.

Расчет коэффициента самоподобия Херста реального трафика обычно выполняется методом абсолютных моментов. В качестве значений случайного процесса рассматривается количество пакетов, поступающих в СМО за единицу времени. Исходная последовательность (ряд) количества пакетов длиной *N* делится на периоды длиной *m* (отдельные агрегированные процессы размером *m*). На непересекающихся временных интервалах или на границах каждого периода *k*-я последовательность имеет среднее значение:

$$X_k^{(m)} = \frac{1}{m}\sum_{j=1}^{m} X_{(k-1)m+1}, \qquad k = 1, 2, 3 \ldots, [N/m].$$

После расчета среднего значения $\overline{X}$ для всей последовательности, потом для каждого периода *k* рассчитается дисперсия D_k:

$$D_k^{(m)} = \frac{1}{N/m}\sum_{j=1}^{m}\left(X_k^{(m)} - \overline{X}\right)^2 .$$

Для самоподобного процесса дисперсия агрегированных процессов должна убывать медленнее, чем величина, обратная размеру выборки *m* [1]. Для выявления этого свойства строится дисперсионно-временной график зависимости дисперсий агрегированных процессов от степени агрегирования *m*. Поскольку Херстом было показано, что:

$$\log\left(\frac{\max D - \min D}{D_k^{(m)}}\right) \approx H \log\left(\frac{N}{2}\right),$$

то график этой зависимости строится тоже в логарифмическом масштабе. Выражение в левой части этого уравнения называется *R*/*S*-статистикой или нормированным размахом [2]. Из полученного графика определяется коэффициент β как тангенс угла наклона аппроксими-рующей кривой к построенной зависимости. Данная аппроксимация выполняется методом минимального среднеквадратичного отклонения от экспериментальных данных. Коэффициент β ($0 < \beta < 1$), задающий асимптотические свойства характеристик самоподобного случайного процесса, связан с параметром Херста следующим соотношением [1]:

$$H = 1 - \frac{\beta}{2}.$$

Для реальных процессов, не имеющих свойства самоподобия, $H < 0{,}5$, а для самоподобных процессов с долгосрочной зависимостью этот параметр изменяется в границах 0,65…0,8 (процесс имеет продолжительную память).

Подготовка ряда для расчета. Из основного временного ряда длиной N элементов (значений) создаются новые вспомогательные ряды, для которых рассчитываются определенные числовые характеристики. Каждый вспомогательный ряд делится на смежные периоды $I_{k,n}$ длиной m_k элементов, где k – номер вспомогательного ряда, а n – номер периода. Максимальное количество периодов вспомогательного ряда $n_{\max} = N / m_{\min}$, где $m_{\min}$ – минимальная длина периода. Для каждого натурального $2 \le n \le n_{\max}$ можно составить вспомогательный ряд, однако точность расчета существенно не ухудшится и при меньшем количестве рядов. Например, на рис. 1 показаны основной 0-й ряд длиной $N = 320$ элементов и вспомогательные ряды 1, 2 и 3, для которых $m_1 = 32$, $m_2 = 80$ и $m_3 = 160$ элементов соответственно.

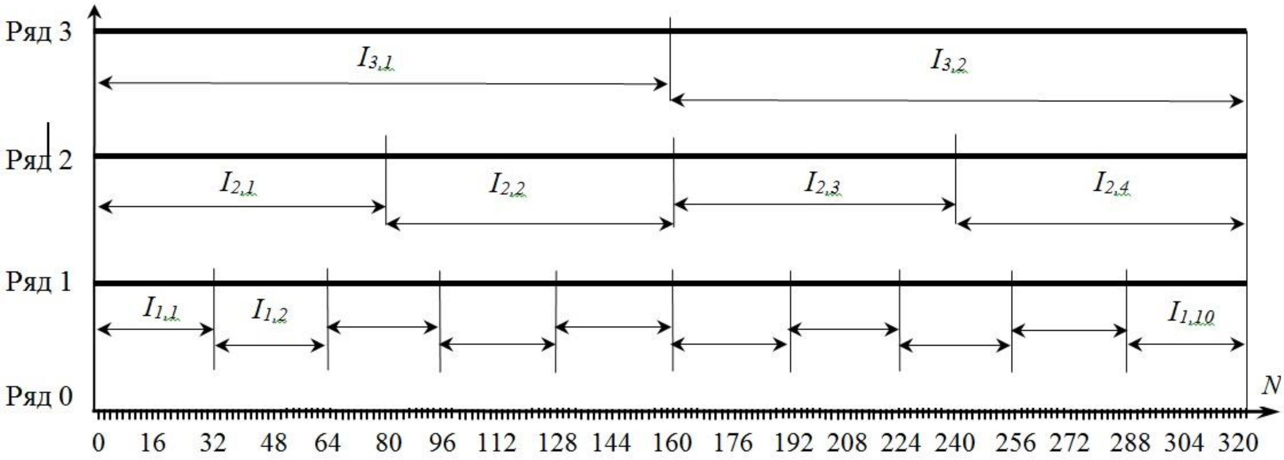

Рисунок 1 – Основной 0-й и вспомогательные ряды 1…3

Из рис. 1 видно, что 1-й вспомогательный ряд получен путем деления основного ряда на десять, 2-й вспомогательный ряд – на четыре и 3-й вспомогательный ряд – на два периода. Для получения достоверного результата обязательно $k_{\max} \ge 5$ при $m_{\min} \ge 10$. Поэтому, в данном случае для получения еще трех вспомогательных рядов можно, например, разделить основной ряд на 5, 8 и 20 периодов длиной $m_4 = 64$, $m_5 = 40$ и $m_6 = 16$ соответственно.

При количестве элементов в основном ряду $N = 2^i$ удобно определить минимальную длину периода $m_{\min} = 2^z$, где, например, $z = 3$, 4, 5 и более. При этом максимальное количество периодов вспомогательного ряда $n_{\max} = N / m_{\min} = 2^i / 2^z = 2^{(i-z)}$, а количество вспомогательных рядов $k_{\max} = i - z$ или $k_{\max} = \log_2(n_{\max})$. Например, при $N = 65536$ и $m_{\min} = 16$ получается $k_{\max} = 12$ вспомогательных рядов. В этом случае будет 12 значений $n = 2^{12}$, 2^{11}, 2^{10}, … , 2^1 и соответственно для каждого из них $m_k = N / 2^{12}$, $N / 2^{11}$, $N / 2^{10}$, … , $N / 2$.

Результаты расчета каждого вспомогательного ряда используются для расчета коэффициента Херста методом наименьших квадратов.

Расчет вспомогательных рядов. После подготовки в соответствии с приведенным алгоритмом вспомогательных рядов производятся следующие расчеты.

Для каждого n-го периода $I_{k,n}$ каждого ряда k:

1. Рассчитывается среднее арифметическое значение $E_{k,n}$ элементов N_j

$$E_{k,n} = \frac{1}{m_k} \sum_{j=1}^{m_k} N_j , \qquad (1)$$

где j – порядковый номер элемента в периоде n.

2. Накапливается сумма $X_{k,n}$ рассчитанных отклонений каждого элемента N_j от среднего значения элементов

$$X_{k,n} = \sum_{j=1}^{m_k-1}(N_j - E_{k,n}). \tag{2}$$

Каждое накапливаемое в цикле (2) значение $X_{k,n}$ запоминается, образуя для *n*-го периода $I_{k,n}$ ряда *k* кумулятивный ряд X_j, где:

$X_1 = (N_1 - E_{k,n})$; $X_2 = X_1 + (N_2 - E_{k,n})$; $X_3 = X_2 + (N_3 - E_{k,n})$ и т.д.

Из этого следует, что значения кумулятивного ряда X_j рассчитываются по ниже приведенной формуле, где *j* также изменяется от 1 до (m_k – 1):

$$X_j = \sum_{i=1}^{j} N_i - jE_{k,n}. \tag{3}$$

В выражении (2) при замене предела суммирования на m_k вся сумма становится равной нулю, что является признаком правильного расчета.

3. Из кумулятивного ряда отклонений находится максимальное $X_{\max}(X_i)$ и минимальное $X_{\min}(X_i)$ значения, а затем рассчитывается размах (диапазон) отклонений $R_{k,n}$:

$$R_{k,n} = X_{\max}(X_j) - X_{\min}(X_j). \tag{4}$$

4. Рассчитывается сумма квадратов отклонения каждого элемента N_j от среднего значения элементов

$$Q_{k,n} = \sum_{j=1}^{m_k}(N_j - E_{k,n})^2. \tag{5}$$

По этим значениям рассчитываются стандартные отклонения $S_{k,n}$ (чаще σ – *Sigma*):

$$S_{k,n} = \sqrt{\frac{1}{m_k}\sum_{j=1}^{m_k}(N_j - E_{k,n})^2}. \tag{6}$$

5. Рассчитывается нормированный размах накопленных сумм (*the adjusted range of cumulative sums*) $RS_{k,n}$

$$RS_{k,n} = \frac{R_{k,n}}{S_{k,n}}. \tag{7}$$

В данном случае величина размаха $R_{k,n}$ нормируется значением эмпирического стандартного отклонения $S_{k,n}$ и, поэтому, иногда *R*/*S*-анализ называется методом нормированного размаха. (Обозначение *RS* нормированного размаха дало название этому методу – *R*/*S*-анализ или *R*/*S*-статистика).

6. Для всех *n* значений нормированного размаха отклонений $RS_{k,n}$ ряда *k* рассчитывается среднее значение

$$\overline{RS}_k = \frac{1}{n}\sum_{1}^{n} RS_{k,n}. \tag{8}$$

7. Финальный расчет для ряда *k* состоит в вычислении логарифма величин $\overline{RS}_k$ и m_k:

$$X_k = \log(m_k), \tag{9}$$

$$Y_k = \log(\overline{RS}_k). \tag{10}$$

Значения X_k и Y_k определяют координаты точек на графике зависимости $\log(\overline{RS}_k)$ от $\log(m_k)$ и используются для расчета коэффициента Херста *H* методом наименьших квадратов.

Логарифм может быть выбран по любому основанию, но для примера с $N = 2^i$ именно при двоичном логарифме оказывается более удобным для расчета вид графика этой зависимости, показанного на рис. 2.

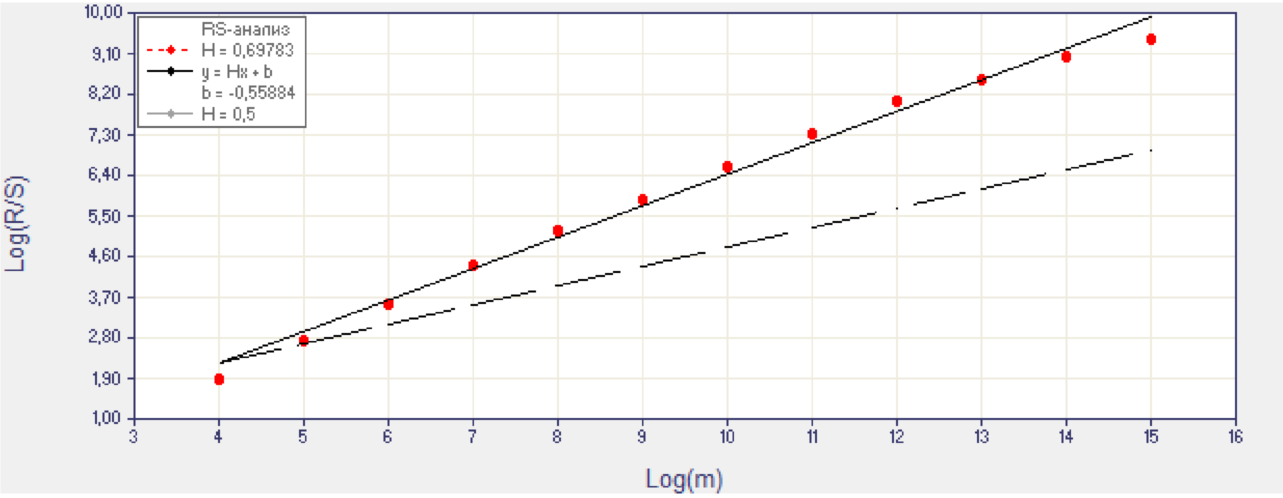

Рисунок 2 – График зависимости $\log(\overline{RS}_k)$ от $\log(m_k)$

Удобство графика состоит в том, что по значениям оси абсцисс легко определяется длина периода m_k в соответствующей точке графика. В данном случае слева направо последовательно $m_k = 2^4$, 2^5, 2^6,…, 2^{15}. На рис. 2 сплошной линией показана прямая, сглаживающая данные расчета X_k и Y_k, а штриховой – наклон прямой, соответствующей $H = 0{,}5$.

Расчет коэффициента Херста. Для оценки коэффициента Херста H анализируется зависимость нормированного размаха R/S от длины периода m. Для этого методом линейной регрессии рассчитанные значения Y_k и X_k (данные каждого вспомогательного ряда) аппроксимируются функцией вида $y = ax + b$. Эта регрессия называется методом наименьших квадратов, поскольку коэффициенты a и b вычисляются из условия минимизации суммы квадратов ошибок $|\, b + ax_i - y_i \,|$. При этом находятся наилучшие значения параметров a и b, максимально приближающие значения функции $y = aX_k + b$ к фактическим значениям Y_k.

Рассчитанные значения Y_k и X_k определяют координаты точек графика зависимости $\log(\overline{RS}_k)$ от $\log(m_k)$, показанного на рис. 2. Коэффициент Херста H соответствует угловому коэффициенту a для прямой, проходящей максимально близко к этим точкам или через них. Угловой коэффициент a линейной функции $y = ax + b$ рассчитывается так:

$$a = \frac{kg_1 - c_2 g_2}{kc_1 - c_2^2}, \tag{11}$$

где

$$c_1 = \sum_{i=1}^{k} X_i^2, \qquad c_2 = \sum_{i=1}^{k} X_i, \qquad g_1 = \sum_{i=1}^{k} X_i Y_i, \qquad g_2 = \sum_{i=1}^{k} Y_i, \tag{12}$$

а k – количество точек на графике (рассчитанных вспомогательных рядов). В этом случае линия построенной регрессии проходит через центр тяжести выборочных данных Y_k и X_k.

Смещение b линейной функции $y = ax + b$ рассчитывается как $b = \overline{y} - a\overline{x}$, где $\overline{y}$ и $\overline{x}$ – средние значения. Для Y_k и X_k получается:

$$b = \frac{1}{k}\sum_{i=1}^{k} Y_i - a\frac{1}{k}\sum_{i=1}^{k} X_i. \tag{13}$$

Таким образом, по результатам R/S-анализа, определяемым значениями Y_k и X_k, рассчитывается коэффициент Херста H, равный значению углового коэффициента a для прямой, сглаживающей результаты расчета Y_k и X_k.

В случае, если регрессионная зависимость между величинами Y_k и X_k достаточно линейна, как, например, на рис. 2, то расчет коэффициента Херста можно выполнить упрощенно без использования выражений (11) и (12).

Угловой коэффициент прямой – коэффициент a в уравнении $y = ax + b$ прямой на координатной плоскости, численно равен тангенсу угла (составляющего наименьший поворот от оси x к оси y) между положительным направлением оси абсцисс и данной прямой линией.

Тангенс угла наклона прямой может рассчитываться как отношение противолежащего катета к прилежащему катету треугольника, образованного этой прямой в качестве гипотенузы и мнимыми отрезками, отложенными на величину относительного смещения двух произвольных точек этой прямой по осям x и y, как показано на рис. 3.

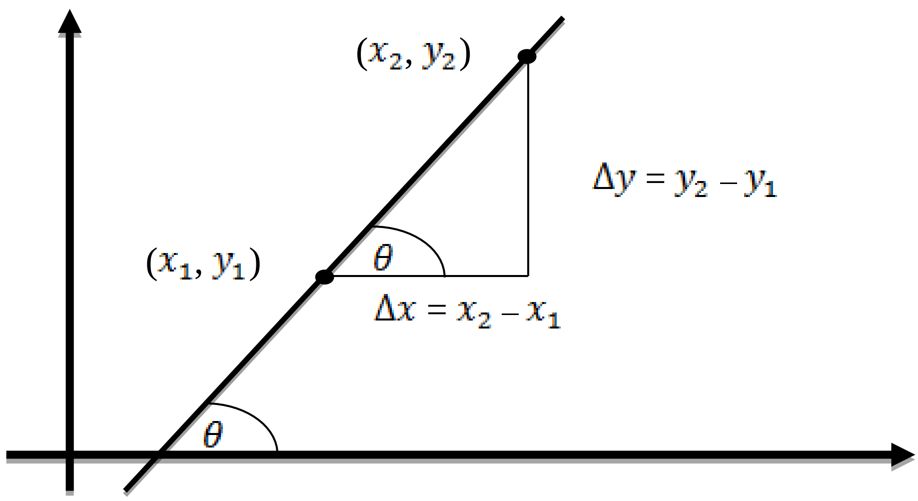

Рисунок 3 – Метод определения угла наклона прямой

Коэффициент a всегда равен отношению величин относительного смещения двух произвольных точек этой прямой по осям x и y:

$$a = \frac{\Delta y}{\Delta x} = \frac{y_2 - y_1}{x_2 - x_1}, \qquad (14)$$

т.е. он равен производной уравнения прямой по x.

Из рис. 2 видно, что крайние точки графика несколько отдалены от аппроксимирующей прямой, полученной в результате регрессии данных Y_k и X_k. Следовательно, из этих данных для расчета углового коэффициента a или коэффициента Херста H необходимо использовать координаты других точек. Например, если использовать из Y_k и X_k только результаты расчета при $k = 2$ и $k = k - 2$ (т.е. отступить от каждого края аппроксимирующей прямой на две точки), то результат упрощенного расчета будет приближаться к результату расчета по формуле (11):

$$H = \frac{Y_{k-2} - Y_3}{X_{k-2} - X_3}. \qquad (15)$$

На рис. 4 показан пример упрощенного расчета коэффициента Херста по значениям Y_k и X_k для третьего и десятого вспомогательных рядов из двенадцати (для третьей и десятой точки графика из двенадцати, показанных на рис. 2). В этих точках длина периода выбранных вспомогательных рядов m_k равна 2^6 и 2^{13} элементов соответственно. При этом точность этого расчета достаточно высока, поскольку полный расчет по формулам (11) и (12) дает значение $H = 0{,}69783$, а по формуле (15) коэффициент $H = 0{,}70986$, что только на 1,72% превышает результат полного расчета.

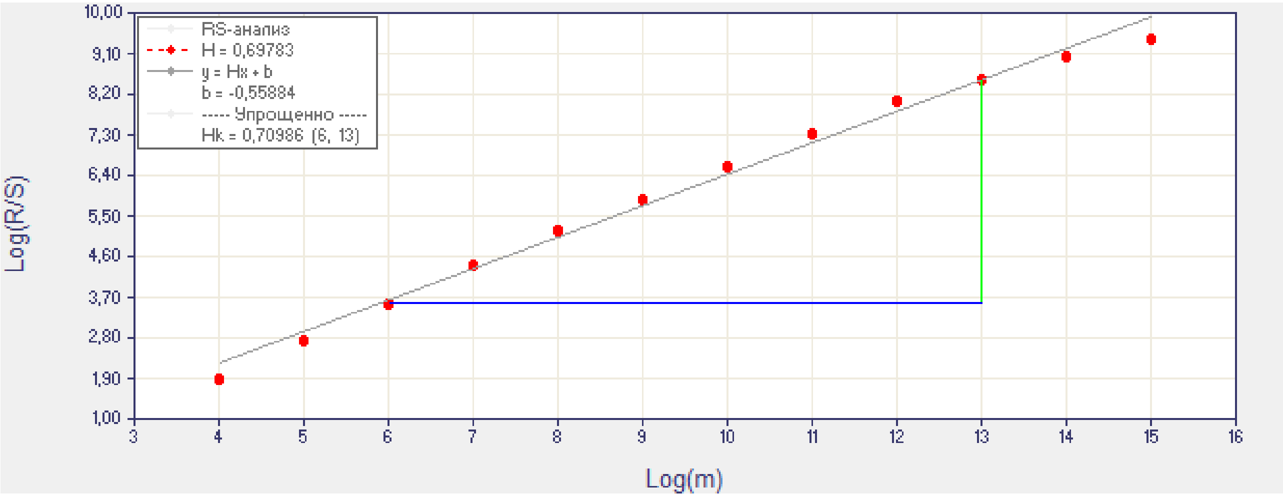

Рисунок 4 – Пример упрощенного расчета углового коэффициента

Отсюда делаем вывод, что погрешность расчета упрощенного метода не превышает 2…5%, что позволяет использовать его в условиях реального процессорного времени обработки, так как существенно сокращено количество операций расчета.

ЛИТЕРАТУРА:

1. Крылов В.В. Теория телетрафика и её приложения / В.В. Крылов, С.С. Самохвалова – СПб.: БХВ-Петербург, 2005. – 288 с.: ил.
2. Ложковский А.Г. Теория массового обслуживания в телекоммуникациях / А.Г. Ложковский // Одеса: ОНАЗ ім. О.С. Попова. – 2012. – 112 с.
3. Ложковський А.Г. Підвищення точності розрахунку характеристик якості обслуговування при самоподібному трафіку мережі / А.Г. Ложковський, О.В.Вербанов // Наукові праці ОНАЗ ім. О.С. Попова. – 2015. – № 1. – С. 36-41.

REFERENCES:

1. Kryilov V.V. Teoriya teletrafika i eyo prilozheniya. / V.V. Kryilov, S.S. Samohvalova // SPb.: BHV-Peterburg, – 2005. – 288 s.: il.
2. Lozhkovskii A.G. Teoriia massovogo obsluzhivaniia v telekomunikatsiiah / A.G. Lozhkovskii. – Odesa: ONAZ im. O.S. Popova. – 2012. – 112 s.
3. Lozhkovskyi A.G. Pidvyshchennia tochnosti rozrakhunku kharakterystyk yakosti obsluhovuvannia pry samopodibnomu trafiku merezhi / A.G. Lozhkovskyi, O.V.Verbanov // Naukovi pratsi ONAZ im. O.S. Popova. – 2015. – № 1. – S. 36-41.