УДК 004.735:007.621.391

Савченко А. С., к.т.н. (Нац. авиационный унив-т. alina@inet.ua)

ИНФОРМАЦИОННО-ЭНТРОПИЙНЫЙ ПОДХОД К ОЦЕНКЕ ПРОИЗВОДИТЕЛЬНОСТИ КОМПЬЮТЕРНЫХ СЕТЕЙ С РАЗНОРОДНЫМ ТРАФИКОМ

Савченко А. С. Інформаційно-ентропійний підхід до оцінки продуктивності комп'ютерних мереж з різнорідним трафіком. Для оцінювання характеристик продуктивності комп'ютерних мереж з різнорідним самоподібним трафіком, необхідно застосовувати непараметричні методи. У якості нижнього порогу продуктивності запропоновано використовувати асимптотичні порівняльні оцінки, наприклад, інформаційно-ентропійні міри імовірнісних розподілів з "важкими хвостами". Показано, що розраховуючи диференціальну ентропію для розподілів Парето, Вейбулла і логарифмічно нормального розподілів інтенсивності різнорідного мережного трафіку, можна отримувати узагальнені порівняльні характеристики ефективності функціонування мереж в широкому діапазоні статистичних параметрів. Ці характеристики можна використовувати в тих випадках, коли отримати оцінки моментів або стаціонарних значень імовірності не представляється можливим.

Ключові слова: комп'ютерна мережа, самоподібний трафік, інформаційно-ентропійна міра, диференціальна ентропія, імовірнісний розподіл, статистичні параметри

Савченко А. С. Информационно-энтропийный подход к оценке производительности компьютерных сетей с разнородным трафиком. Для оценивания характеристик производительности компьютерных сетей с разнородным самоподобным трафиком, необходимо применять непараметрические методы. В качестве нижнего порога производительности предложено использовать асимптотические сравнительные оценки, например, информационно-энтропийные меры вероятностных распределений с "тяжелыми хвостами". Показано, что рассчитывая дифференциальную энтропию для распределений Парето, Вейбулла и логарифмически нормального распределений интенсивности разнородного сетевого трафика, можно получать обобщенные сравнительные характеристики эффективности функционирования сетей в широком диапазоне статистических параметров. Эти характеристики можно использовать в тех случаях, когда получить оценки моментов или стационарных значений вероятности не представляется возможным.

Ключевые слова: компьютерная сеть, самоподобный трафик, информационно-энтропийная мера, дифференциальная энтропия, вероятностное распределение, статистические параметры

Savchenko A. S. Information-entropy approach to performance evaluation of computer networks with heterogeneous traffic. Key parameters for the functioning of computer networks can be evaluated using information entropy measures. Comparative characteristics of the differential entropy of probabilistic distributions with «heavy tails» allow for the self-similarity of modern traffic networks. Counting differential entropy for Pareto distribution, Weibull and log-normal distributions of intensity of heterogeneous network traffic, we can obtain the generalized comparative performance characteristics of the networks in a wide range of statistical parameters. These characteristics can be used in those cases where the moments to obtain estimates or stationary probability values is not possible.

Keywords: computer network, self similar traffic, information and entropy measure, differential entropy, probabilistic distributing, statistical parameters

Введение. На сегодняшний день эффективная работа корпоративной компьютерной сети (ККС) является одной из важнейших составляющих успешной деятельности современного предприятия любой отрасли. Поддержание на заданном уровне качества обслуживания клиентов с использованием сетевой инфраструктуры предприятия – основная задача ККС. Ее решение невозможно без создания и внедрения эффективных систем управления наявными ресурсами. В работах [1, 2] предложена структура системы управления ККС. При системном подходе к проблеме повышения качества обслуживания (QoS) в ККС невозможно обойтись без надежных методов анализа качества обслуживания заявок на услуги в реальных условиях формирования их потоков и с учетом свойств циркулирующего трафика.

Изложение основных принципов оценки производительности вычислительной сети представлено в [3]. За основу анализа временных характеристик работы сети взяты модели простейшего потока. В рассматриваемых ККС указанные предположения выполняются не в полной мере. Использование неадекватных математических моделей трафика приводит к недооценке характеристик QoS сети. Трафик таких сетей, как правило, является разнородным (речь, видео, данные) и самоподобным по своей природе [4]. Его статистические характеристики уже не могут быть описаны распределениями экспоненциального семейства. В этом случае используются распределения с так называемыми “тяжелыми хвостами” (Парето, Вейбулла, логарифмически нормальное распределения). Случайный процесс поступления пакетов в систему на обслуживание, образующий поток пакетов (трафик), характеризуется законом распределения, устанавливающим связь между значением случайной величины (количеством пакетов) и вероятностью появления этого значения. В большинстве случаев для расчета параметров QoS достаточно знать о законе распределения только некоторые его числовые характеристики. Например, для расчета в условиях пуассоновского распределения достаточно математического ожидания M, а для нормального распределения – необходимо иметь значения M и дисперсии D.

Однако эти характеристики не всегда являются исчерпывающими, а иногда и бесполезными для прогнозирования значения случайной величины. Возможны варианты, когда случайные процессы характеризуются одинаковыми значениями математического ожидания и дисперсии, но внутренняя структура этих процессов различна. Одни могут иметь плавно меняющиеся реализации, а иные – ярко выраженную колебательную структуру при скачкообразном изменении отдельных значений случайной величины (например, резкое возрастание количества пакетов в сети, приводящее к “пачечности” *(burstness)* трафика). Для “плавных” процессов характерна большая предсказуемость реализаций, а для “пачечных” – очень малая вероятностная зависимость между двумя случайными величинами. В таких случаях говорят, что закон распределения, характеризующий процесс, несет в себе некоторую неопределенность и позволяет с большей или меньшей надежностью предсказать значение случайной величины. Например, при равномерном распределении все значения случайной величины равновероятны, а при экспоненциальном – наименьшие значения имеют наибольшую вероятность.

Числовой характеристикой распределения, которая может служить его мерой неопределенности, является энтропия закона распределения. Понятие энтропии связано с понятием количества информации, и является мерой неопределенности [5]. Приобретение информации сопровождается уменьшением неопределенности, поэтому количество информации можно измерять количеством исчезнувшей неопределенности, т. е. энтропии. Снятие неопределенности дает возможность принимать обоснованные решения при управлении сложными системами, например, ККС.

В случае дискретного источника сообщений, т.е. дискретной случайной величины, энтропия определяется формулой Больцмана

$$H(x) = -\sum_{x} f(x) \ln f(x) dx ,$$

где x – случайная величина, а $f(x)$ – ее распределение вероятностей.

Энтропия не зависит от значений, принимаемой случайной величиной, а только от их вероятностей. Обычно представляет интерес не абсолютное значение энтропии, а сравнение энтропий различных законов. Следовательно, для оценивания характеристик производительности ККС, в которых циркулирует разнородный самоподобный трафик, необходимо применять непараметрические методы. В качестве нижнего порога производительности можно получать некие асимптотические сравнительные оценки, например, информационно-энтропийные меры вероятностных распределений.

Анализ исследований и постановка задачи. В работах [6...8] рассмотрены информационно-энтропийные характеристики сетевого трафика беспроводных сетей. Предполагается, что временной интервал, на котором происходит передача данных, разбивается на виртуальные слоты. В каждом слоте может вообще не быть пакета (ни одна из станций сети не ведет передачу), или один пакет (одна и только одна станция ведет передачу), или имеет место коллизия, когда передавать пытаются две или более станций.

Пусть в начале каждого слота t_k j-я станция пробует отправить пакет. Вероятность попытки обозначим p_{kj}. Если попытка оказалась неудачной, после некоторого интервала отсрочки $\tau_d(n_{tr})$ она повторяется. Общая длительность интервала отсрочки не зависит от числа попыток передачи $n_{tr} = 0,1,2,\ldots$ вплоть до наступления события успешной передачи. Величина $\tau_d(n_{tr})$ выбирается из геометрического распределения с параметром $\tau_d(0)$, т.е. $\tau_d(n_{tr}) = 0,1,2,\ldots$ с соответствующими вероятностями $\tau_d(0)$, $\tau_d(0)\left[1-\tau_d(0)\right]$, $\tau_d(0)\left[1-\tau_d(0)\right]^2$, …

Шкала времени – дискретная, а каждый тип слота представляет собой целое число коротких (элементарных) интервалов. Поскольку станция делает попытку передачи в начале слота, вероятность коллизии и число повторных попыток не зависят от длительности пакета. Как отмечалось выше, наиболее общей мерой для вероятностных распределений, по крайней мере, принадлежащих к одному типу (в рассматриваемом случае – к дискретному типу), является энтропия. В работе представлены сравнительные энтропийные характеристики некоторых дискретных распределений и показано, что:

– геометрическое распределение неразрывно связано с биномиальным. Отличие состоит в том, что биномиальная случайная величина определяет вероятность m успехов в n испытаниях, а геометрическая – вероятность n испытаний до первого успеха (включая первый успех);

– равномерно распределённая на $[-a,a]$ случайная величина имеет наивысшую энтропию среди всех случайных величин, распределённых на $[-a,a]$;

– показательное распределение с параметром λ имеет наибольшую энтропию среди всех распределений, определённых на полуоси $[0,\infty]$ с математическим ожиданием λ;

– на всей прямой $[-\infty,\infty]$, среди всех распределений с фиксированными математическим ожиданием и дисперсией, наибольшей энтропией обладает нормальное распределение.

На практике, в ККС чаще встречаются источники информации, множество возможных состояний которых составляет континуум, т.е. непрерывные источники. Поэтому исследование информационно-энтропийных характеристик сетевого трафика ККС с самоподобными свойствами, описываемых непрерывными распределениями с "тяжелыми хвостами", является актуальной задачей.

Дифференциальная энтропия непрерывного источника сообщений. Оценка неопределенности выбора для непрерывного источника информации имеет свою специфику.

Во-первых, значения, реализуемые источником, математически отображаются непрерывной случайной величиной.

Во-вторых, вероятности значений этой случайной величины не могут использоваться для оценки неопределенности, поскольку в данном случае вероятность любого конкретного значения равна нулю.

Естественно связывать неопределенность выбора значения непрерывной случайной величины с плотностью распределения вероятностей этих значений. Учитывая, что для совокупности значений, относящихся к любому сколь угодно малому интервалу непрерывной случайной величины, вероятность конечна, существует формула для энтропии непрерывного источника информации [5]:

$$H(x) = -\int_{-\infty}^{+\infty} f(x)\log f(x)dx\,.$$

Поскольку для определения этой величины используется только функция плотности вероятности, т. е. дифференциальный закон распределения, она получила название относительной дифференциальной энтропии или просто дифференциальной энтропии непрерывного источника.

Ее можно трактовать как среднюю неопределенность выбора случайной величины с произвольным законом распределения по сравнению со средней неопределенностью выбора случайной величины изменяющейся в диапазоне, равном единице, и имеющей равномерное распределение.

Дифференциальная энтропия в отличие от энтропии дискретного источника является относительной мерой неопределенности. Ее значение зависит от масштаба случайной величины, а следовательно, и от выбора единицы ее измерения.

Энтропия дискретного источника всегда положительна. Дифференциальная энтропия $H(x)$ в отличие от энтропии источников дискретных сообщений может принимать положительные, отрицательные и нулевые значения, т.е. может быть отрицательной [9]. Дифференциальная энтропия в отличие от обычной энтропии дискретного источника не является мерой собственной информации, содержащейся в ансамбле значений случайной величины. Она зависит от масштаба x и может принимать отрицательные значения. Информационный смысл имеет не сама дифференциальная энтропия, а разность двух энтропий, чем и объясняется ее название.

Дифференциальная энтропия не меняется при изменении всех возможных значений случайной величины x на постоянную величину. Действительно, масштаб x при этом не меняется, и справедливо равенство

$$H(x+C) = -\int_{-\infty}^{+\infty} f(x+C)\log f(x+C)d(x+C) = -\int_{-\infty}^{+\infty} f(x)\log f(x)dx\,.$$

Из этого следует, что $H(x)$ не зависит от математического ожидания случайной величины, т.к. изменяя все значения x на C, мы тем самым изменяем на C и ее среднее, то есть математическое ожидание.

Дифференциальная энтропия аддитивна, то есть для объединения $x \cup y$ независимых случайных величин x и y справедливо равенство:

$$H(x \cup y) = H(x) + H(y)\,.$$

Доказательство этого свойства аналогично доказательству свойства аддитивности обычной энтропии.

Дифференциальная энтропия распределений с "тяжелыми хвостами". Как отмечалось выше, интерес представляет не абсолютное значение энтропии, а сравнение энтропий различных распределений. Рассмотрим, какие непрерывные распределения с "тяжелыми хвостами" (Парето, Вейбула, логарифмически нормальное) обладают наибольшей дифферненциальной энтропией.

Плотность распределения Парето задается функцией:

$$f(x) = \frac{\alpha}{k}\left(\frac{k}{x}\right)^{\alpha+1},$$

где α – параметр формы; k – минимальное значение случайной величины x.

При $\alpha \le 2$ дисперсия бесконечна (что и требуется в качестве одного из условий самоподопобности).

Наличие в распределении так называемого "тяжелого хвоста" обеспечивает свойство пачечности трафика, поскольку в распределении существенно возрастают вероятности длинных интервалов между событиями (например, отсутствие пакетов на интервале) и для «поддержания» заданного среднего значения количества событий необходима их концентрация (увеличение) на других интервалах времени.

Параметр формы a распределения Парето и параметр Херста H находятся в такой зависимости [3]:

$$H = \frac{3-\alpha}{2}\,.$$

Дифференциальная энтропия распределения Парето рассчитывается таким образом.

$$H_P(X) = -\int_{-\infty}^{+\infty} \frac{\alpha}{k}\left(\frac{k}{x}\right)^{\alpha+1} \log\left(\frac{\alpha}{k}\left(\frac{k}{x}\right)^{\alpha+1}\right)dx =$$

$$= -\alpha k^{\alpha} \int_{-\infty}^{+\infty} \frac{1}{x^{\alpha+1}} \log\left(\frac{\alpha k^{\alpha}}{x^{\alpha+1}}\right)dx = \left| \begin{array}{ll} u = \log\left(\dfrac{\alpha k^{\alpha}}{x^{\alpha+1}}\right); & du = -\dfrac{\alpha+1}{x}dx \\ dv = \dfrac{1}{x^{\alpha+1}}dx; & v = -\dfrac{1}{\alpha x^{\alpha}} \end{array} \right| =$$

$$= -\alpha k^{\alpha}\left[\log\left(\frac{\alpha k^{\alpha}}{x^{\alpha+1}}\right)\cdot\left(-\frac{1}{\alpha x^{\alpha}}\right) - \int_{-\infty}^{+\infty}\frac{1}{\alpha x^{\alpha}}\cdot\left(-\frac{\alpha+1}{x}\right)dx\right] =$$

$$= \alpha k^{\alpha}\left[\frac{1}{\alpha x^{\alpha}}\log\left(\frac{\alpha k^{\alpha}}{x^{\alpha+1}}\right) + \frac{\alpha+1}{\alpha}\int_{-\infty}^{+\infty}\frac{1}{x^{\alpha+1}}dx\right] =$$

$$= \alpha k^{\alpha}\left[\frac{1}{\alpha x^{\alpha}}\log\left(\frac{\alpha k^{\alpha}}{x^{\alpha+1}}\right) + \frac{\alpha+1}{\alpha}\left(-\frac{1}{\alpha x^{\alpha}}\right)\right] = \frac{k^{\alpha}}{x^{\alpha}}\log\left(\frac{\alpha k^{\alpha}}{x^{\alpha+1}}\right) - \frac{(\alpha+1)}{\alpha}\frac{k^{\alpha}}{x^{\alpha}}.$$

Поскольку дифференциальная энтропия не зависит от конкретного значения случайной величины, распределенной по закону Парето, то x можно заменить на минимальное значение этой величины, т.е. k. $x \to x_{\min} \to k$. Тогда получим значение дифференциальной энтропии, определяющей нижний предел количества передаваемой информации.

$$H_P(X) = \frac{k^{\alpha}}{x^{\alpha}}\log\left(\frac{\alpha k^{\alpha}}{x^{\alpha+1}}\right) - \frac{(\alpha+1)}{\alpha}\frac{k^{\alpha}}{x^{\alpha}}\bigg|_{x\to k} = \frac{k^{\alpha}}{k^{\alpha}}\log\left(\frac{\alpha k^{\alpha}}{k^{\alpha+1}}\right) - \frac{(\alpha+1)}{\alpha}\frac{k^{\alpha}}{k^{\alpha}}\bigg| = \log\frac{\alpha}{k} - \frac{1}{\alpha} - 1.$$

Плотность распределения Вейбулла задается функцией вида:

$$f_W(x) = \begin{cases} \dfrac{k}{\lambda}\left(\dfrac{x}{\lambda}\right)^{k-1} e^{-\left(\frac{x}{\lambda}\right)^k}, & x \geq 0; \\ 0, & x < 0, \end{cases}$$

где $\lambda > 0$ – параметр масштаба, $k > 0$ – параметр формы.

Экспоненциальное распределение – частный случай распределения Вейбулла, соответствующий значению параметра формы $k = 1$.

Дифференциальная энтропия распределения Вейбулла рассчитывается аналогично предыдущему случаю, и равна

$$H_W(X) = -\int_{-\infty}^{+\infty}\frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1} e^{-\left(\frac{x}{\lambda}\right)^k} \log\left(\frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1} e^{-\left(\frac{x}{\lambda}\right)^k}\right)dx = \gamma\left(1-\frac{1}{k}\right) + \left(\frac{\lambda}{k}\right)^k \log\left(\frac{\lambda}{k}\right).$$

Плотность логарифмически нормального распределения задается функцией вида:

$$f_L(x) = \begin{cases} \dfrac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}}, & x > 0; \\ 0, & x \leq 0, \end{cases}$$

где $\sigma > 0$ – среднеквадратическое отклонение, μ – математическое ожидание.

Дифференциальная энтропия логарифмически нормального распределения равна

$$H_L(X) = -\int_{-\infty}^{+\infty} \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln x-\mu)^2}{2\sigma^2}} \log\left(\frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln x-\mu)^2}{2\sigma^2}}\right) dx = \frac{1}{2} + \frac{1}{2}\log(2\pi\sigma^2) + \mu \ .$$

Анализ результатов. Для сравнительного анализа дифференциальных энтропий рассматриваемых непрерывных распределений с "тяжелыми хвостами" были рассчитаны зависимости энтропии от параметра, представляющего среднеквадратическое отклонение распределения.

На Рис. 1 представлен график зависимости дифференциальной энтропии распределения Парето от параметра α. Значение параметра находится в пределах $1 \le \alpha \le 3$, что соответствует параметру Херста $0 < H < 1$.

На графике можно наблюдать монотонный рост энтропии, следовательно, рост потребного ресурса обмена данными при увеличении параметра α.

На Рис. 2 представлен график зависимости дифференциальной энтропии распределения Парето от параметра H. Значение параметра находится в пределах $0{,}5 < H < 1$, что характеризует наличие самоподобных свойств. В таком случае также наблюдается монотонный рост энтропии (по модулю) в зависимости от коэффициента Херста. Это означает, чем выше степень самоподобия, тем выше степень неопределенности, соответственно, потребуется больший ресурс для обмена данными.

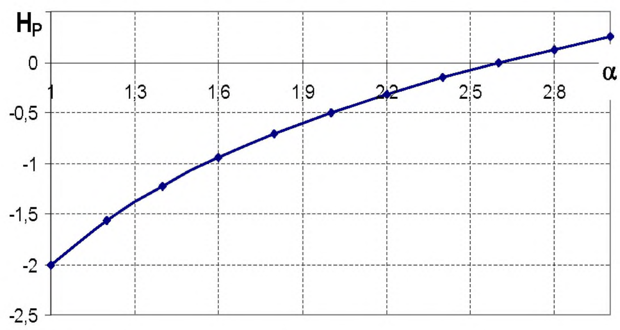

Рис. 1. Зависимость дифференциальной энтропии распределения Парето от параметра α

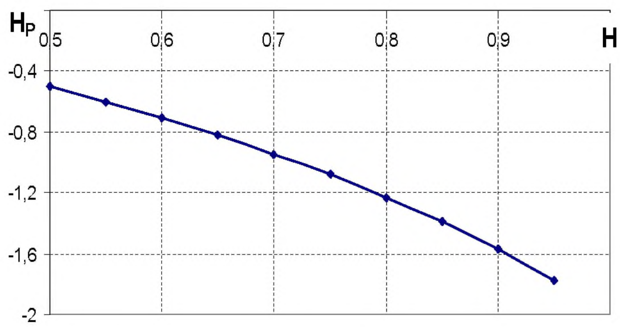

Рис. 2. Зависимость дифференциальной энтропии распределения Парето от параметра H

На Рис. 3 представлен график зависимости дифференциальной энтропии распределения Вейбулла от параметра k. Видно, что при росте параметра степень неопределенности уменьшается.

На Рис. 4 представлен график зависимости дифференциальной энтропии логарифмически нормального распределения от среднеквадратического отклонения. Показано, что чем сильнее разброс значений от математического ожидания, тем выше степень неопределенность, т.е. необходим больший ресурс для обмена данными.

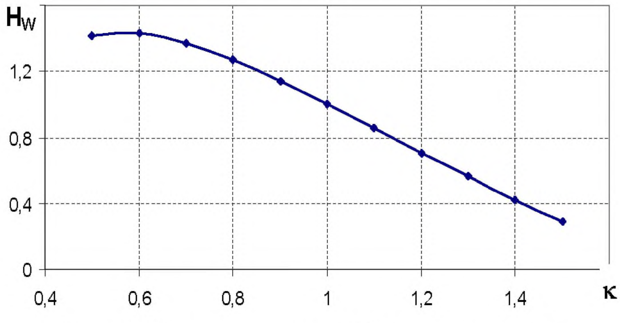

Рис. 3. Зависимость дифференциальной энтропии распределения Вейбулла от параметра формы k

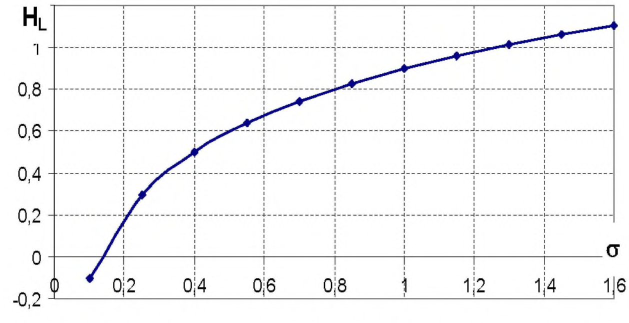

Рис. 4. Зависимость дифференциальной энтропии логарифмически нормального распределения от среднеквадратического отклонения

Заключение

Ключевые параметры эффективности функционирования ККС можно оценить с помощью информационно-энтропийных мер. Сравнительных характеристики дифференциальных энтропий непрерывных распределений с “тяжелыми хвостами” позволяют учитывать свойства самоподобия трафика современных сетей.

Рассчитывая дифференциальную энтропию для распределений Парето, Вейбулла и логарифмически нормального распределений интенсивности разнородного сетевого трафика, можно получать обобщенные сравнительные характеристики эффективности функционирования сетей в широком диапазоне статистических параметров. Эти характеристики можно использовать в тех случаях, когда получить оценки моментов или стационарных значений вероятности (например, через уравнения Колмогорова) не представляется возможным.

Целью дальнейших исследований является получение оценок производительности компьютерной сети в зависимости от параметров самоподобного трафика, выраженных через дифференциальную энтропию источника сообщений.

Литература

1. Віноградов М. А. Концепція управління корпоративною комп‘ютерною мережею на основі психофізіологічних механізмів професійної діяльності людини / М. А. Віноградов, А. С. Савченко // Наукові записки Українського науково-дослідного інституту зв‘язку. – 2013. – Вип.3(27). – С. 5-14.

2. Савченко А. С. Метод принудительного ввода системы управления в область устойчивости / А. С. Савченко // Наукові записки Українського науково-дослідного інституту зв‘язку. – 2012. – Вип. 2(22). – С. 100-105.

3. Вишневский В. М. Теоретические основы проектирования компьютерных сетей / В. М. Вишневский. – М.: Техносфера, 2003. – 512 с.

4. Столлингс В. Современные компьютерные сети. 2-е издание / В. Столингс. – СПб.: Питер, 2003. – 783с.

5. Стратонович Р. Л. Теория информации / Р. Л. Стратонович. – М.: Сов.радио, 1975. – 424с.

6. Bianchi G. Performance analysis of the IEEE 802.11 distributed coordination function // IEEE Journal on Selected Areas in Communications. – Nr. 18(3), March, 2000. – PP. 535-547.

7. Cali F., Conti M., Gregory E. Tuning of the IEEE 802.11 protocol to achieve a theoretical throughput limit // IEEE/ACM Transactions on Networking. – Nr. 8(6), December, 2000. – PP. 785-799.

8. Амирханов Э. Д. Оценивание производительности специализированных беспроводных сетей при произвольных статистиках длины пакетов данных / Э. Д. Амирханов // Наукові записки Українського науково-дослідного інституту зв‘язку. – 2012. – Вип.2(22). – С. 60-65.

9. Теория электрической связи: учебное пособие / К. К. Васильев, В. А. Глушков, А. В. Дормидонтов, А. Г. Нестеренко; под общ. ред. К.К. Васильева. – Ульяновск: УлГТУ, 2008. – 452 с.