

УДК: 681.324

Разработка адаптивной модели для оптимального распределения нагрузки в корпоративных сетях

Р.К. АЛЕКПЕРОВ, канд. техн. наук, А.А. ГУСЕЙНОВА

Институт информационных технологий НАНА, г. Баку

E-mail: rased@iit.ab.az

В работе проанализирована нагрузка на каналы связи в корпоративных сетях и построена тригонометрическая модель нагрузки на каналы связи. Проведенный анализ показал, что статистические данные не полностью обеспечивают прогнозирование нагрузки на каналы связи. Построенная математическая модель сглаживает случайные выбросы в статистические данные и дает возможность прогнозирования количества будущих обращений. Имеется возможность проведения на основе данной модели практических работ по обеспечению равномерного распределения трафика.

Ключевые слова: каналы связи, корпоративная сеть, трафик сети, адаптивная модель, распределение нагрузки, ряд Фурье.

Известно, что нагрузка на каналы связи в корпоративных сетях, обеспечивающих выход в Интернет, зависит от количества пользователей, одновременно обращающихся в сеть, и от объема информации, которую требуется передать или принять. Основные процессы, связанные с отказами и прерываниями в работе корпоративных сетей, зависят от перегрузок в каналах связи. В разные рабочие часы нагрузка на каналы связи бывает различной. В некоторые моменты времени образуются пиковые накопления обращений. Неравномерное распределение обращений может привести к неэффективному использованию ресурсов корпоративной сети и Интернета в целом. Равномерное распределение информации в рабочие часы способствует повышению эффективности корпоративных сетей и удовлетворению требований пользователей. Для повышения эффективности широко используются такие технологии, как CDN (Content Delivery Network) [1–5], SAN (Storage Area Network) [6] и др. Применение этих технологий повышает пока-

затели качества услуг (QoS-Quality of Service), создает возможность для решения некоторых задач. Одним из мероприятий, повышающих качество услуг в корпоративных сетях, является осуществление контроля за трафиком. Посредством анализа трафика становится возможным осуществление соответствующих мероприятий. Работы, выполненные с целью повышения эффективности сетей, качества предоставляемых услуг, не теряют своей актуальности, как и контроль за трафиком и работой, осуществляемый для обеспечения равномерного распределения трафика.

Одной из основных задач в управлении корпоративной сетью является обеспечение рационального использования ресурсов [6]. К подобным задачам относится обеспечение равномерного распределения нагрузок в течение рабочих часов суток. Для проведения превентивных мероприятий необходимо прогнозирование объема трафика за сутки. Статистические данные, характеризующие трафик в корпоративной сети, в некотором смысле удовлетворяют требованиям контроля.

Однако статистическим данным свойственно подвергаться случайным воздействиям. Они относятся к прошлому и настоящему времени и малопригодны для прогноза. Аналитическая модель обеспечит адекватность прогнозирования будущих нагрузок в сети.

Известна модель зависимости объема трафика в канале связи от времени [7, 8], где модель построена как сумма moving average и стохастического компонента. Авторы построили модель, используя отрезок времени, где интенсивность обращений к сети высока.

Данная статья посвящена разработке аналитической математической модели зависимости объема трафика в канале связи от времени с использованием тригонометрических функций. Фактические данные для решения

задачи взяты из архивов корпоративной сети Национальной академии наук Азербайджана (НАНА) AzScienceNet.

Разработка аналитической модели зависимости объема трафика в канале связи от времени

Нагрузка в сети в рабочие дни недели дана в таблице. По данным таблицы создан двумерный массив, охватывающий пять рабочих дней и 24 ч в сутки $DH[1...5, 1...24]$. Мера измерений табличных данных и данных массива $DH[1...5, 1...24]$ — ГБ (GB). Из массива $DH[1...5, 1...24]$ создан массив $DM[1...5, 1...1440]$, отражающий минутные объемы трафика за рабочие дни недели. Массив охватывает пять рабочих дней и 1440 мин в каждом рабочем дне. Кроме того, создан одномерный массив $FDM[1...7200]$, отражающий пять рабочих дней в минутах и массив $ADM[1...1440]$ для среднего дня за неделю.

Элементы этого массива имеют следующие соотношения:

$$DM[i, j] = \frac{DH\left[i, \left[\frac{j-1}{60}\right] + 1\right]}{60}; \quad (1)$$

$i = 1-5, j = 1-1440,$

$$FDM[j] = DM\left[\left[\frac{j-1}{1440}\right] + 1, \left\{\frac{j}{1440}\right\}\right]; \quad (2)$$

$j = 1-7200,$

$$ADM[j] = \frac{\sum_{k=1}^5 DM[k, j]}{5}; \quad j = 1-1440. \quad (3)$$

На рис. 1 представлен график объема часового трафика за неделю.

По оси OX — часы рабочих дней за выбранную отчетную неделю, по оси OY — объем трафика в ГБ. В окне показан часовой объем обмена информацией за пятый день. Из рис. 1 следует, что недельный трафик имеет некоторую периодичность.

На рис. 2 представлен график массива $ADM[1...1440]$, т.е. распределение трафика по минутам для среднего дня недели. Особенность этого графика заключается в том, что на 480-й мин рабочего дня (8-й час

Нагрузка в сети в рабочие дни недели

Часы \ Дни	I	II	III	IV	V
1	0	0	0,1	0	0
2	0,1	0	0,1	0,1	0
3	0	0	0,2	0	0
4	0,1	0	0	0	0
5	4	0	0	0	0
6	6	0	0	0	0
7	6,5	0,2	0	0,2	0,1
8	8,3	1,5	1	1	0,5
9	19	10	9	9	11
10	23	20	16	22	18
11	24	24	18	22	18,5
12	28	21	21	21	21,5
13	22,5	22	18	27	19
14	23	23	19	27	20
15	22,1	19	20	21	19
16	21	29	17,5	20	14,5
17	11	17	9	12	10
18	3	1	4	3	2
19	4	0,9	6	1,5	2
20	1	0,8	4,5	0,2	1
21	1,1	0,5	0,5	0,1	0,3
22	0,2	0,1	0,5	0	0,1
23	0	0,1	0,1	0	0
24	0	0	0,1	0	0

5.1.0
5.2.0
5.3.0
5.4.0
5.5.1
5.6.0
5.7.0
5.8.1
5.9.12
5.10.18
5.11.19
5.12.21
5.13.19
5.14.20
5.15.20
5.16.10
5.17.10
5.18.16
5.19.2
5.20.1
5.21.0
5.22.0
5.23.0
5.24.0

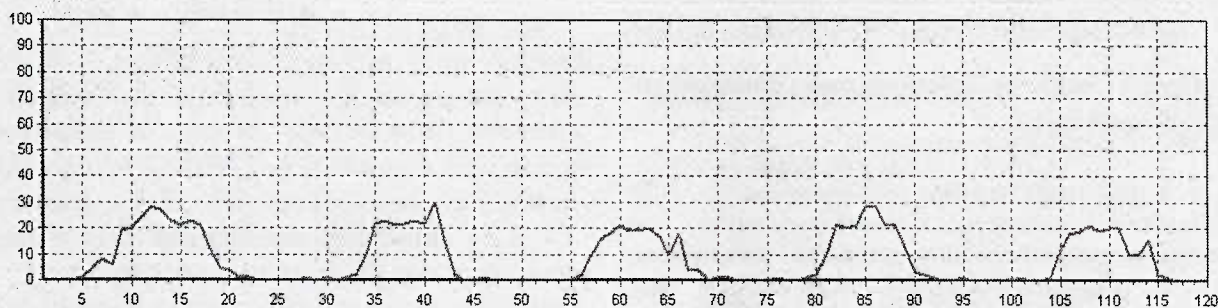


Рис. 1. Объем трафика в корпоративной сети НАНА за отчетную неделю

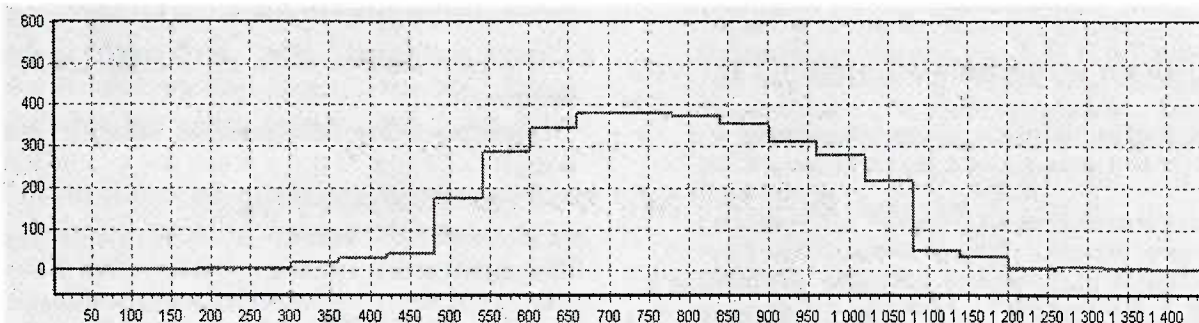


Рис. 2. Распределение трафика по минутам для среднего дня недели

дня) объем обращений к сети существенно поднимается, а на 1080-й мин (18-й час дня) объем обращений к сети снижается. Разница в границах часов показывает, что обращения к сети не распределены между минутами равномерно.

Аналитическая математическая модель зависимости объема трафика от времени с применением тригонометрических функций и разложением в ряд Фурье устраняют эти неравномерности.

Если функция $f(x)$ определена на некотором отрезке $[-l, l]$, частично монотонна в указанной области и выполняется условие Дирихле, она разложится в ряд Фурье [9]:

$$f(x) = \sum_{n=1}^{\infty} b_n \sin \frac{n\pi}{l} x, \quad (4)$$

где

$$b_n = \frac{2}{l} \int_0^l f(x) \sin \frac{n\pi}{l} x dx, \quad n = 1, 2, \dots \quad (5)$$

Известно, что при разложении нечетной функции по синусам сходимость ряда будет (n^{-3}) .

Допустим, что элементы массива ADM отражают значения функции $f(x)$, определенные на отрезке $[0; T]$ в точках $0; h; 2h; 3h; \dots$

При этом $f(0) = ADM[1]; f(h) = ADM[2]; f(2h) = ADM[3]; \dots$

Для обеспечения нечетности выполним следующие преобразования:

$g(t) = f(t) - (\alpha - \beta t)$. α и β выбираются из условия $g(0) = 0, g(T) = 0$.

Тогда выполняется $0 = f(0) - \alpha, 0 = f(T) - \alpha - \beta T$. Отсюда для α и β получим:

$$\alpha = f(0), \beta = \frac{f(T) - f(0)}{T}. \quad (6)$$

После чего преобразования принимают следующий вид:

$$g(t) = f(t) - f(0) - \frac{f(T) - f(0)}{T} \cdot t. \quad (7)$$

Для отрицательных значений t на отрезке $[-T; 0]$ можно определить $g(-t) = -g(t)$. Это обеспечит нечетность $g(t)$. Разложение в ряд Фурье $g(t)$ на промежутке $[-T; T]$ будет определяться выражением:

$$g(t) = b_1 \sin \frac{\pi}{T} \cdot t + b_2 \sin \frac{2\pi}{T} \cdot t + b_3 \sin \frac{3\pi}{T} \cdot t + \dots \quad (8)$$

Коэффициенты ряда выразим как:

$$b_k = \frac{2}{N} \sum_{r=1}^{N-1} g(r \cdot h) \cdot \sin kr \frac{\pi}{N}. \quad (9)$$

Если в разложении оставим коэффициенты в количестве $m \ll N$, тогда функция $g(t)$ будет иметь сглаженный вид:

$$\bar{g}(t) = \sum_{k=1}^m b_k \sin k \cdot \frac{\pi}{T} \cdot t. \quad (10)$$

Для определения числа m анализируются коэффициенты b_k . Эти коэффициенты сначала имеют относительно высокие абсолютные значения, а потом по абсолютному значению постепенно уменьшаются. Индекс начала уменьшения можно выбрать как значения m . При этом $f(t)$ определяется как: $\bar{f}(t) = \bar{g}(t) + (\alpha + \beta t)$ и дает сглаженный массив ADM в аналитическом виде.

Таким образом, используя данные корпоративной сети НАНА из массива $DH[1 \dots 5, 1 \dots 24]$, можно построить аналитическую модель зависимости объема трафика от времени средних суток отчетной недели следующим образом:

$$DM[i, j] = \frac{DH \left[i, \left[\frac{j-1}{60} \right] + 1 \right]}{60}; \quad (11)$$

$$i = 1 - 5, j = 1 - 1440,$$

$$ADM[j] = \frac{\sum_{k=1}^5 DM[k, j]}{5}; j = 1 - 1440, \quad (12)$$

$$\alpha = ADM[1], \beta = \frac{ADM[1440] - ADM[1]}{1440}. \quad (13)$$

$$g(t) = ADM(t) - \alpha - \beta \cdot t; t = 1 - 1440. \quad (14)$$

$$h = 1; N = 1440; \quad (15)$$

$$b_k = \frac{2}{N} \sum_{r=1}^{N-1} g(r \cdot h) \cdot \sin kr \frac{\pi}{N}, \quad (16)$$

$$\bar{g}(t) = \sum_{k=1}^m b_k \sin k \cdot \frac{\pi}{T} \cdot t; m \leq \frac{N}{4}. \quad (17)$$

$$F(t) = \bar{g}(t) + (\alpha + \beta t); t = 1 - 1440. \quad (18)$$

График зависимости по модели представлен на рис. 3.

Программное обеспечение модели разработано в Delphi 7 [10].

Среднеквадратическое отклонение модели от массива ADM можно вычислить по следующей формуле:

$$\Delta M = \frac{\sum_{i=1}^N (ADM[i] - F[i])^2}{N}. \quad (19)$$

Для прогнозирования будущих нагрузок сети можно воспользоваться простым выражением:

$$F(pt) = K \cdot F(t), \quad (20)$$

где $F(t)$ — функция зависимости обращений от времени для опорной недели; $F(pt)$ — функция прогноза; K — отношение предполагаемого количества пользователей прогнозируемой недели к количеству пользователей опорной недели.

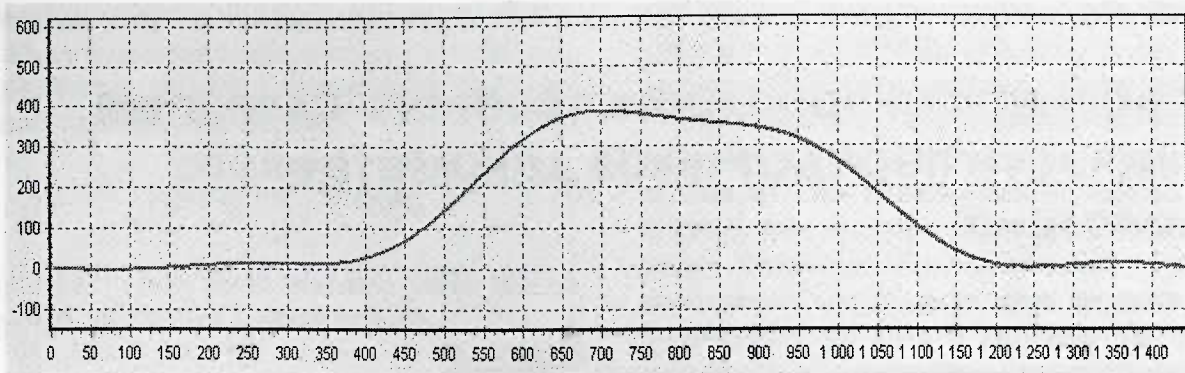


Рис. 3. График тригонометрической модели зависимости объема обращений к корпоративной сети

$$K = \frac{N_{pt}}{N_t},$$

где N_t — количество пользователей опорной недели; N_{pt} — количество пользователей прогнозируемой недели. Необходимо отметить, что при оценке N_{pt} учитывается количество пользователей, которые в прогнозируемой неделе будут в отпуске, командировке и т.д.

Если прогнозируемые значения будут выше пропускной способности корпоративной сети, можно планировать следующие действия.

1. Используя специальные интерфейсные программы, можно загрузить некоторые файлы (*.wav, *.avi, *.vob, *.bmp) и отодвинуть их в малозагруженные часы. Эти файлы определяются анализом архивов log-файлов трафика.

2. При недостаточности вышеуказанных операций с помощью мультиагентных технологий необходимо определить сайты, к информации которых часто обращаются пользователи сети. Таким образом, при обращениях к этой информации в рабочие часы запросы могут быть быстро удовлетворены.

Выводы

Разработана аналитическая тригонометрическая модель зависимости объемов обращений к корпоративной сети НАНА от времени. Аналитическая модель получена в виде ряда Фурье. Полученная модель сглаживает случайные выбросы в статистические данные

и дает возможность прогнозирования количества будущих обращений. Даны практические рекомендации для управления трафиком.

СПИСОК ЛИТЕРАТУРЫ

1. Lazar I., Terill W. Exploring content delivery network // IT Professional. 2001. V. 3. N. 4. P. 47—49.
2. Алгулиев Р.М., Алыгулиев Р.М., Гусейнова А.А. Повышение эффективности корпоративных сетей с применением CDN-технологии // Информационные технологии. 2008. № 7. С. 2—9.
3. Hablinger G., Hartleb F. Content delivery and caching from a network provider's perspective // Computer Networks. 2011. V. 55. N. 18. P. 3991—4006.
4. Choi J., Han J., Cho E., Kwon T., Choi Y. A survey on content-oriented networking for efficient content delivery // IEEE Communication Magazine. 2011. V. 49. N. 3. P. 121—127.
5. Adler M., Sitaraman R.K., Venkataramani H. Algorithms for optimizing the bandwidth cost of content delivery // Computer Networks: The International Journal of Computer and Telecommunications Networking. 2011. V. 55. N. 18. P. 4007—4020.
6. Алгулиев Р.М., Рагимов Э.Р. Синтез архитектуры защищенной корпоративной сети // Телекоммуникации. 2006. № 2. С. 19—23.
7. Бахвалов Л.А., Репин Д.С., Филаретов Г.Ф. Алгоритм анализа трафика в корпоративных компьютерных сетях на основе статистики экстремальных значений // Программные продукты и системы. 2008. № 3 (83). С. 8—10.
8. Бахвалов Л.А., Репин Д.С. Имитационное моделирование трафика в корпоративных компьютерных сетях // Материалы XXXV международной конференции «Информационные технологии в науке, образовании, телекоммуникации, бизнесе» — IT+SE'2008, Ялта—Гурзуф. 2008. С. 138—140.
9. Сергиенко А.Б. Цифровая обработка сигналов. Санкт-Петербург: «Питер». 2005. 603 с.
10. Хомоненко А. и др. Delphi 7. Санкт-Петербург: «БХВ Петербург», 2008. 1200 с.