



Алгоритм прогнозирования динамики изменения трафика с использованием статистических данных

А.О. Игумнов, М.А. Сонькин

Национальный Исследовательский Томский Политехнический Университет

Аннотация: В статье представлен подход к прогнозированию изменения динамики событий, на примере анализа трафика, с использованием статистических данных. Проведён анализ систем обработки трафика как систем массового обслуживания. Представлена актуальность задачи прогнозирования изменения трафика. Алгоритм подбора шаблонов позволяет строить прогноз, как для регулярной динамики, так и для неожиданных всплесков. Алгоритм прогнозирования позволяет улучшить работу систем анализа трафика для периодов высокой нагрузки.

Ключевые слова: Мониторинг сети, анализ трафика, прогнозирование, система массового обслуживания, случайные события, обслуживание заявок.

Мониторинг сети

В наше время интернет (всемирная система объединённых компьютерных систем) используется повсеместно, для огромного количества функций: общение, реклама, передача данных и т.п. Мониторинг соответствующей активности использования сети позволяет определить конверсию или построить информационный тренд, определив количество и состав проходящей в сети информации [1]. Мониторинг сети может осуществляться множеством способов, одним из которых является анализ трафика, непосредственно проходящий в сети. Для анализа сопоставляется количественная характеристика применения (объём трафика или определённые соответствия в нём) относительно времени (Рис.1).

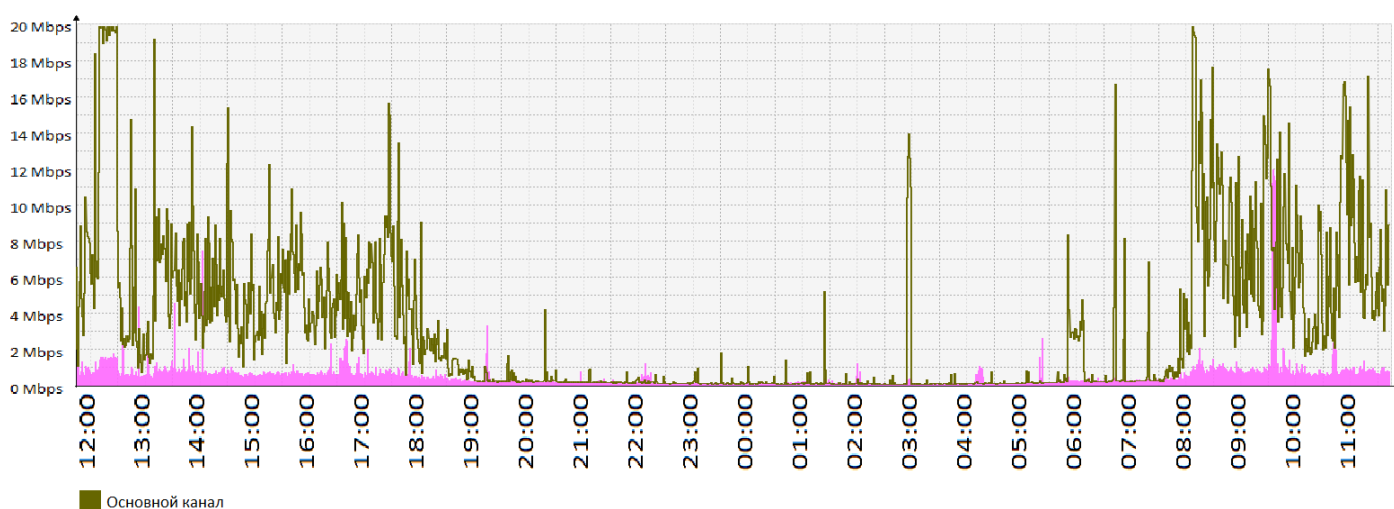


Рис. 1. Пример изменения трафика во времени

Системы массового обслуживания

Системы анализа трафика производят обслуживание поступающих требований, что по сути, делает их системами массового обслуживания (СМО) — системы, которые [2]. Системой массового обслуживания называется система, в которой:

- возникает массовые требования на выполнения каких-либо видов услуг
- происходит удовлетворение этих требований — обслуживание

При этом главной особенностью процессов обслуживания является их случайность, что приводит к случайному протеканию всего процесса обслуживания. Причины случайности заключаются в массовом характере потребностей, а также в случайности их возникновения [3, 4]. Таким образом, СМО можно рассматривать как очередь заявок которые поступают в обслуживающий модуль в результате работы которого формируется результат для конечного использования (Рис. 2):

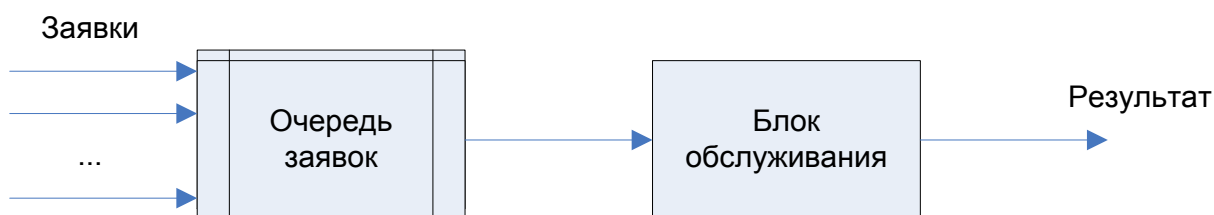


Рис. 2 Система массового обслуживания

Однако сам блок обслуживания СМО может состоять из множества модулей обслуживания входящей очереди. Каждый модуль может обрабатывать заявки за разное время, и сами модули могут обрабатывать заявки как последовательно, так и параллельно (Рис. 3).

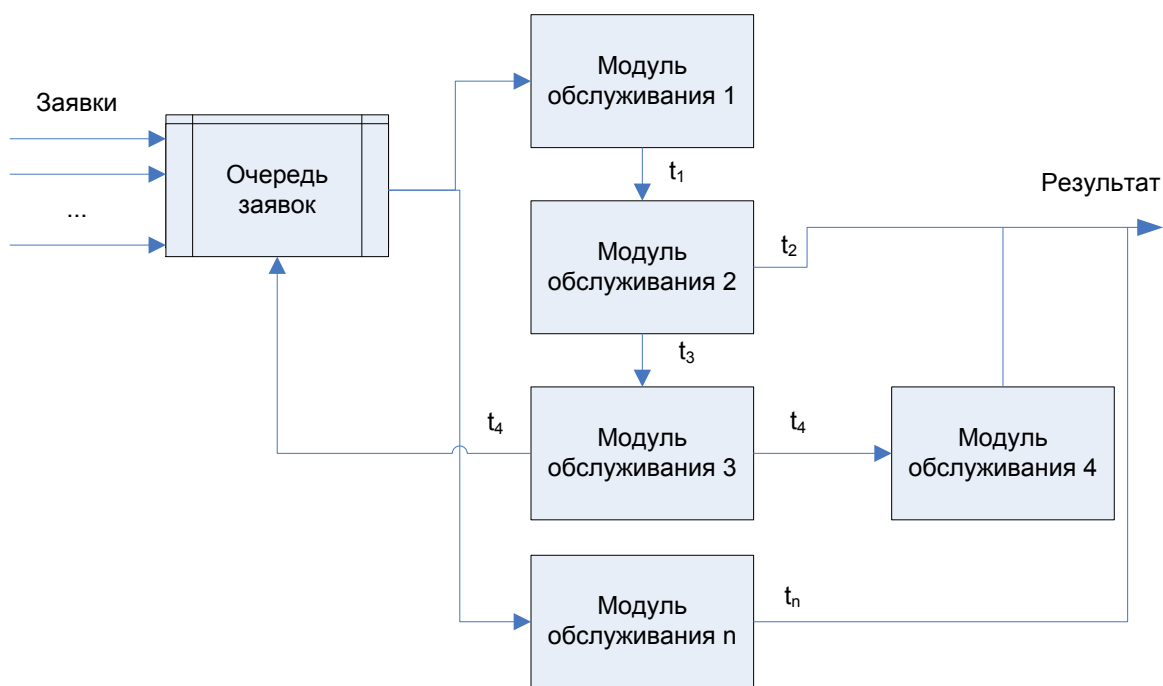


Рис. 3 Модульная система массового обслуживания

В связи с указанными особенностями модульной обработки заявок анализ трафика может осуществляться полностью только до определённого предела, за которым время обработки модулей может спровоцировать задержку обработки заявок в очереди привести к её переполнению и как следствие потери части заявок очереди (Рис. 4).

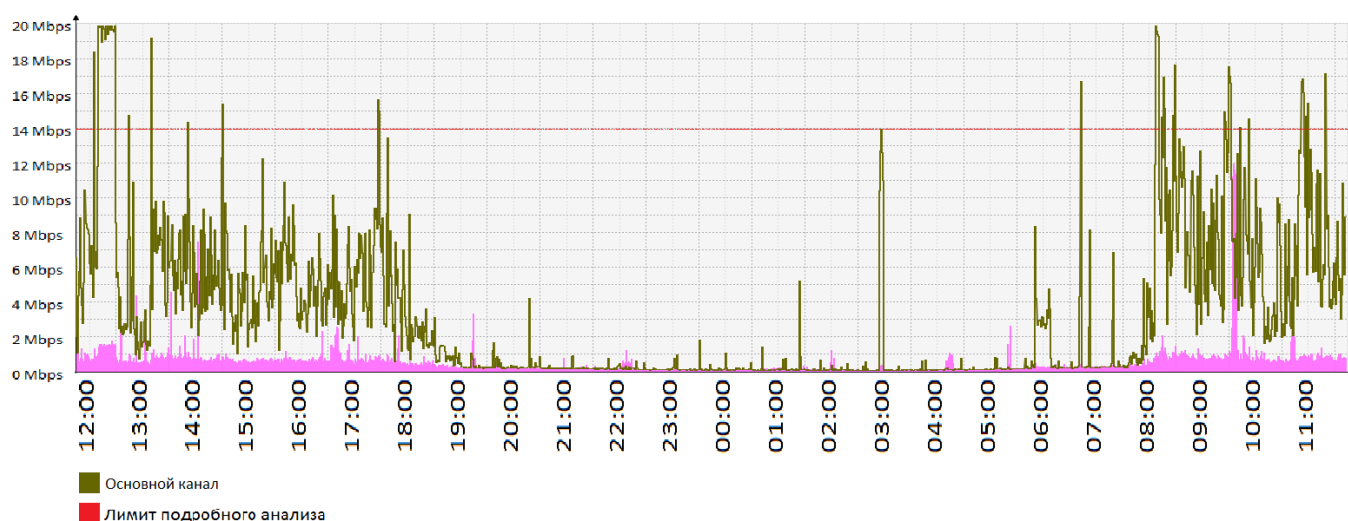


Рис. 4. Лимит подробного анализа трафика

В зависимости от алгоритмов обработки информации, модули, осуществляющие более продолжительную обработку трафика, могут освобождать задействованные ресурсы только по факту обработки заявки, и при неожиданном возрастании объёма трафика можно потерять часть информации. Поэтому в процессе работы актуально выстраивать прогноз, для регулировки приоритетов модулей заранее, до возрастания объёма до критических значений [5].

Прогнозирование

При анализе трафика по дням, можно заметить, что данным присуща определённая повторяемость. Связано это с наиболее закреплённым расписанием работы людей для анализируемой сети. Как следствие, можно предположить, что на основе предыдущей динамики, можно спрогнозировать изменение текущей [6, 7].

Для прогнозирования, можно использовать информацию о динамике изменения трафика в прошлом. Для этого можно осуществлять поиск соответствия динамики за последние S часов, соответствующему времени за предыдущие дни. Последние S часов можно разделить поминутно на участки



j относительно текущего времени t . Оценка соответствия при этом производится относительно объема трафика в минуту $V(t - j)$. Сравнение последовательности значений объема трафика позволят сформировать шаблон изменения трафика [8]. Для нивелирования изменчивости трафика во времени, предлагается искать шаблон соответствия относительно времени t_s учитывающее отклонение на некоторую дельту x относительно текущего времени t : $t - x \geq t_s \geq t + x$. Следовательно, выбор наиболее подходящего шаблона для конкретного дня p можно определить, найдя минимальное значение коэффициента:

$$\min_{t-x \geq t_s \geq t+x} \Delta(p, t, t_s) = \sum_{j=0}^{n_j} \left| \frac{1}{V(t-j)} - \frac{1}{V(i, t_s - j, p)} \right|,$$

где: $\Delta(p, t, t_s)$ – рассчитанный коэффициент приемлемости шаблона для t_s .

Минимальное значение коэффициента из интервала $t - x \geq t_s \geq t + x$ будет детерминировать наиболее приемлемое для текущей динамики изменения трафика для данного участка времени t [9].

Поскольку для трафика характерны факторы резкого изменения, предлагается ввести понятие нагрузки временного участка $t - j$:

$$w(i, j) = \frac{1}{[V(t - j, p)]^B},$$

где B это весовой коэффициент, определяющий степень актуальности текущей динамики изменений трафика. Соответственно конечный коэффициент соответствия шаблона определяется:



$$\min_{t-x \geq t_s \geq t+x} \Delta(p, t, t_s) = \sum_{j=0}^{n_j} w(i, j) \left| \frac{1}{V(t-j)} - \frac{1}{V(i, t_s-j, p)} \right|,$$

Для формирования конечного прогноза предлагается вычислить наиболее подходящие шаблоны для P количества дней относительно текущего дня u : $p \in (u - C, u - (C - 1), \dots, u - 1)$. Целесообразно использовать ограничить количество анализируемых шаблонов меньшим, чем C количеством H . Так же предлагается ввести показатель минимального соответствия E относительно которого производить отсев Δ шаблонов.

Таким образом, после расчёта подходящих шаблонов для C предыдущих дней и просеивания относительно минимальной степени соответствия E , если их остаётся более H то выбирается H наиболее подходящих. При условии полного отсутствия шаблонов удовлетворяющих E , выбирается день, для которого $\Delta(p, t, t_s)$ будет минимальной.

После выбора шаблонов для каждого исторического дня p , отбирается соответствующее значение объёма трафика $V(t_s, h)$ на основании которых можно вычислить прогнозируемое значение:

$$\hat{V}(t, p) = \frac{\sum_{h \in \Omega(t, p)} T(t_s, h)}{n_k},$$

где: $\hat{V}(t, p)$ – конечный прогноз, $\Omega(t, p)$ – набор подобранных шаблонов, n_k – количество шаблонов.

Итог

Используя архивные данные о загрузке сети непрерывно задействованного компьютера за месяц, можно проверить эффективность

работы алгоритма [10]. Для этого предлагается построить прогноз для последних актуальных часов, сопоставляя с предыдущими днями архива и сравнить с актуальными данными последнего дня (Рис. 5).

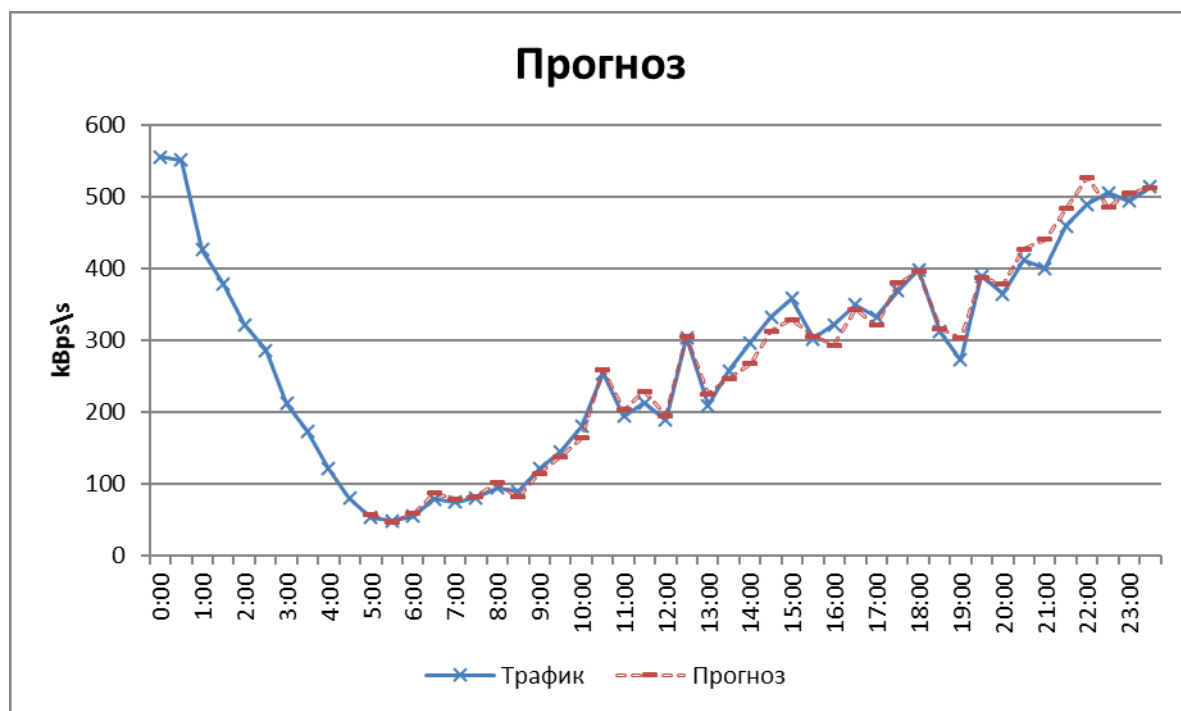


Рис 5. Сопоставление данных прогноза и актуальных данных

Как видно из графика в 80% случаев погрешность прогноза не превышает 4% погрешность относительно действительных данных, что позволяет определить резкие возрастания трафика. В оставшихся 20% погрешность возрастает до 7%, однако тренд прогноза всё равно совпадает с трендом реальных данных. Погрешность на более высоких показателях можно уменьшить, имея в распоряжении трафик за больший период времени. Для этого можно использовать приближённые алгоритмы, которые сократят выборку, и затем построить прогноз с использованием предложенного алгоритма.

Литература

1. S. McCreary, K. C. Claffy, "Trends in Wide Area IP Traffic Patterns", *ITC Specialist Seminar*, 2000-May. pp.3-10.



2. Система массового обслуживания // Cyclowiki.org URL:
http://cyclowiki.org/wiki/%D0%A1%D0%B8%D1%81%D1%82%D0%B5%D0%BC%D0%B0_%D0%BC%D0%B0%D1%81%D1%81%D0%BE%D0%B2%D0%BE%D0%B3%D0%BE_%D0%BE%D0%B1%D1%81%D0%BB%D1%83%D0%B6%D0%B8%D0%B2%D0%B0%D0%BD%D0%B8%D1%8F
 3. Матальцкий М., Хацкевич Г. Теория вероятностей, математическая статистика и случайные процессы. – Litres, 2016, 202 с.
 4. Розанов Ю. А. Стационарные случайные процессы. М.: ФИЗМАТЛИТ, 1990. 42-51 с.
 5. Каграманзаде А. Г., Каграманзаде С. Д. Прогнозирование трафика-основа прогнозирования современных сетей электросвязи. ЦНТИ, //ЦНТИ, “Информсвязь“, 1991 - №. 1, М., 44с.
 6. Бабенко Г. В., Белов С.В. Анализ трафика TCP/IP на основе методики допустимого порога и отклонения // Инженерный вестник Дона, 2011, №2 URL:
ivdon.ru/ru/magazine/archive/n2y2011/446.
 7. Ажмухамедов И.М., Марьенков А.Н. Поиск и оценка аномалий сетевого трафика на основе циклического анализа // Инженерный вестник Дона, 2012, №2 URL:
ivdon.ru/ru/magazine/archive/n2y2012/742
 8. Скуратов А. К., Безрукавный Д. С. Администрирование телекоммуникационной сети на основе статистического анализа трафика // Вестник ТГТУ. 2004. №4-1. URL:
cyberleninka.ru/article/n/administririvanie-telekommunikatsionnoy-seti-na-osnove-statisticheskogo-analiza-trafika
-



9. Stephen Clark Traffic Prediction Using Multivariate Nonparametric Regression // Journal of Transportation Engineering. 2003 - №129. pp.161-168.
10. Дли М.И. Алгоритм экстраполяции случайного процесса с применением метода локальной аппроксимации // Программные продукты и системы. Тверь: Программные продукты и системы, 1999. 20-24 с.

References

1. S. McCreary, K. C. Claffy, "Trends in Wide Area IP Traffic Patterns", ITC Specialist Seminar, 2000May. pp.3-10
 2. Sistema massovogo obsluzhivaniya [Queuing system]. Cyclowiki.org URL:
cyclowiki.org/wiki/%D0%A1%D0%B8%D1%81%D1%82%D0%B5%D0%BC%D0%B0_%D0%BC%D0%B0%D1%81%D1%81%D0%BE%D0%B2%D0%BE%D0%B3%D0%BE_%D0%BE%D0%B1%D1%81%D0%BB%D1%83%D0%B6%D0%B8%D0%B2%D0%B0%D0%BD%D0%B8%D1%8F
 3. Matalytskiy M., Khatskevich G. Teoriya veroyatnostey, matematicheskaya statistika isluchaynye protsessy [Probability theory, mathematical statistics and random processes]. – Litres, 2016, 202 p.
 4. Rozanov Yu. A. Statsionarnye sluchaynye protsessy [Stationary random processes]. M.: FIZMATLIT, 1990. pp.42-51
 5. Kagramanzade A. G., Kagramanzade S. D. Prognozirovaniye trafika-osnova prognozirovaniya sovremennykh setey elektrosvyazi [Traffic forecasting is the basis for forecasting modern telecommunication networks]. TsNTI, TsNTI, "Informsvyaz". 1991 №. 1, M., 44 p.
 6. Babenko G.V., Belov S.V. Inženernyj vestnik Dona (Rus), 2011. №2 URL: ivdon.ru/ru/magazine/archive/n2y2011/446.
-



7. Azhmukhamedov I.M., Mar'enkov A.N. Inzhenernyj vestnik Dona (Rus), 2012, №2. URL: ivdon.ru/ru/magazine/archive/n2y2012/742
8. Skuratov A. K., Bezrukavnyy D. S. Administrirovanie telekommunikatsionnoy seti na osnove statisticheskogo analiza trafika [Administration of the telecommunications network based on statistical analysis of traffic]. Vestnik TGTU. 2004 №4-1. URL: cyberleninka.ru/article/n/administrirovanie-telekommunikatsionnoy-seti-na-osnove-statisticheskogo-analiza-trafika
9. Stephen Clark Traffic Prediction Using Multivariate Nonparametric Regression. Journal of Transportation Engineering. 2003. №129. pp.161-168
10. Dli M.I. Algoritm ekstrapolyatsii sluchaynogo protsessa s primeneniem metoda lokal'noy approksimatsii [An algorithm for extrapolating a random process using the local approximation method]. Programmnye produkty i sistemy. Tver': Programmnye produkty i sistemy, 1999. pp.20-24