

Литература

1. О.Б.Кунгурцев, С.М.Бородавкін. Застосування мереж фреймів для побудови моделі вилучення фактів з текстів на природній мові // Искусственный интеллект, 2009. – №4.– С. 202 - 207.
2. А.Б.Кунгурцев, И.В.Барыкина. Формирование словаря предметной области // Искусственный интеллект, 2006. - №1.
3. С.А.Шаров. Частотный словарь русского языка. – РосНИИ. – 2001
4. George K. Zipf. Human Behavior and the Principle of Least-Effort. – Addison-Wesley. – 1949
5. <http://cs.isa.ru:10000/dwarf/>. Программный пакет синтаксического разбора и машинного перевода.

УДК 004.3:004.492

Наводиться стислий опис глобальної інфраструктури моніторингу Internet – ATLAS. Проаналізовано характер сучасного Internet-трафіку та здійснене прогнозування його річного зростання

Ключові слова: ATLAS, ISP, AS, Comcast, SPSS, AR1

Приведено краткое описание глобальной инфраструктуры мониторинга Internet – ATLAS. Проанализирован характер современного Internet-трафика и выполнен прогноз его годового роста

Ключевые слова: ATLAS, ISP, AS, Comcast, SPSS, AR1

This article represents the short description of ATLAS global Internet infrastructure. The major findings of modern Internet-traffic are analyzed and the forecast of its annual growth is made

Key words: ATLAS, ISP, AS, Comcast, SPSS, AR1

ПРОГНОЗИРОВАНИЕ ГОДОВОГО РОСТА ТРАФИКА ИНТЕРНЕТ, НА ОСНОВЕ ПОДОБРАННОЙ МОДЕЛИ ВРЕМЕННОГО РЯДА

В. В. Шкарупило*

Контактный тел.: 066-129-73-45

E-mail: vadshkar@yandex.ru

К. Н. Касьян

Кандидат технических наук, доцент*

Кафедра компьютерных систем и сетей

*Запорожский национальный технический университет
ул. Жуковского, 64, г. Запорожье, 69063

1. Введение

На сегодняшний день наблюдаются существенные структурные изменения характера Internet-трафика. Хотелось бы особо акцентировать внимание на следующих ключевых аспектах (приведенные данные актуальны на четвертый квартал 2009 года):

- 150 автономных систем (AS) агрегируют в себе более 50% всего трафика глобальной сети Internet;

- около 30 мировых корпораций генерируют 30% всего трафика;

- подавляющий объем трафика циркулирует между клиентом и сервером с данными: за год (с 2008 по 2009 год) относительная доля контентного трафика в общем объеме трафика выросла с 35 до 55%; такой стремительный рост дает все основания полагать, что именно вышеназванный тип трафика следует рассматривать в качестве центрального элемента, с целью анализа и прогнозирования;

- значительно снизилась стоимость услуг поставщиков Internet-сервисов (ISP), что обусловлено, прежде всего, высоким уровнем конкуренции, стре-

мительным ростом пропускной способности каналов (активное внедрение оптоволоконных линий связи). ISP постоянно экспериментируют с тарифами и вариантами оплаты своих услуг.

Таким образом, актуальным становится вопрос прогнозирования годового роста трафика Internet. Этот прогноз позволит, в частности, количественно оценить эффект от стремительного внедрения оптоволоконных линий связи и существенного смягчения ценовой политики ISP.

2. Инфраструктура мониторинга Internet

ATLAS – глобальная инфраструктура мониторинга Internet [1]. Сбор данных осуществляется с более чем 110 ISP (Internet Service Providers). Отслеживание трафика осуществляется практически в режиме реального времени (14 Тб/с).

Приблизительное размещение точек сбора данных приведено на рис. 1. Рисунок фиксирует текущее состояние на 4-й квартал 2009 года.

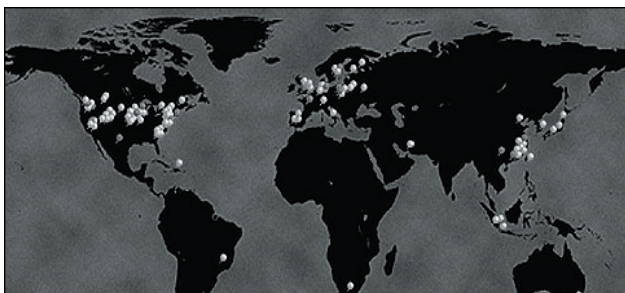


Рис. 1. Расположение серверов сбора данных ATLAS

На рис. 2 приведен график, отражающий рост объема контентного трафика, относительно объема всего трафика, инкапсулируемого североамериканской автономной системой (AS) Comcast, которая входит в число 50 крупнейших AS.

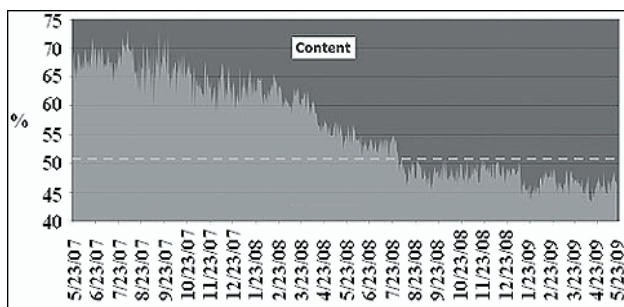


Рис. 2. Структура Internet-трафика AS Comcast

Из рис. 2 видно, что, за период с 5/23/07 по 5/23/09, относительная доля контентного трафика, приходившегося на долю Comcast, увеличилась с 33% до 53%. Поскольку эти цифры отражают общую тенденцию структурного изменения характера передаваемых данных по каналам глобальной сети Internet, а Comcast входит в число 50 крупнейших AS мира, можем выдвинуть гипотезу, что, спрогнозировав относительный рост контентного трафика на 2010 год, мы сможем

составить приблизительную картину о годовом процентном приросте объема передаваемых данных. Следует отметить, что довольно резкий «скачок» объема контентного трафика в 2008 году обусловлен, прежде всего, следующими факторами: активным наращиванием Японией и Южной Кореей пропускной способности своих каналов.

Следует отметить, что прогноз ATLAS относительно годового увеличения объема Internet-трафика, базирующийся на 9 экзбайтах (10^{15} байт) данных, обрабатываемых в месяц, составляет примерно 44%. Прогноз же Cisco, основывающийся на аналогичном объеме ежемесячных данных, составляет примерно 50%. Независимый опрос ISPs, в свою очередь, свидетельствует о годовом росте порядка 30 – 45 %.

3. Постановка задачи

Необходимо спрогнозировать относительное увеличение объема всего Internet-трафика на 2010 год. Анализ временного ряда и прогнозирование его будущих значений (экстраполяция) будем осуществлять с помощью статистического программного пакета SPSS 10.0. После построения интересующего нас графика в вышеуказанном программном пакете, необходимо подобрать математическую модель, выделить тренд, проанализировать случайную составляющую на коррелированность и выполнить экстраполяцию [2]. Далее предполагается сравнить объем контентного трафика за 5/23/2009 с прогнозируемым значением на 5/23/2010 и определить процентный прирост контентного трафика за год (его мы будем рассматривать в качестве ключевого фактора, стимулирующего рост глобального Internet-трафика). Для проверки корректности полученного результата, его предлагается сравнить с аналогичными прогнозами мировых лидеров в области мониторинга Internet.

4. Прогнозирование годового прироста объема трафика

Основываясь на рис. 3, для построения графика роста контентного трафика и получения соответствующей математической модели, будем использовать выборку из 100 значений.

Полученный график приведен на рис. 3.

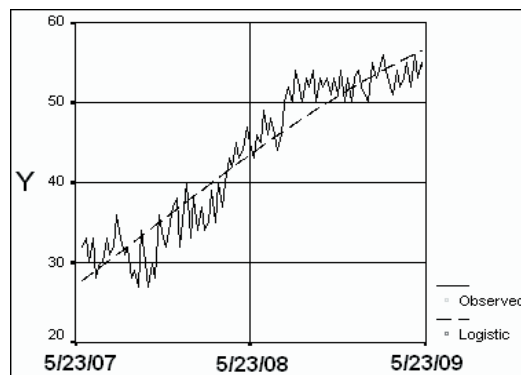


Рис. 3. Логистическая модель тренда

Подобранный тренд соответствует логистической модели. Y – доля контентного трафика в процентах. Полученный тренд, описываемый логистической моделью, показан пунктиром. В нашем случае коэффициент Пирсона равен 0.89: это свидетельствует о том, что подобранная модель на 89% соответствует характеру происходящего процесса.

Модель тренда описывается следующим выражением:

$$tr(t) = \frac{1}{(1/a) + b_0 \cdot b_1^t}, \tag{1}$$

где функция $tr(t)$ представляет изменение процентной доли контентного трафика во времени t ($t=1,2,\dots,100$); a, b_0, b_1 – коэффициенты. В нашем случае $a=80, b_0=0.0232, b_1=0.985$.

Дальнейший анализ модели связан с исследованием остатков [2]. Необходимо подобрать модель случайной составляющей, что позволит нам, в последствии, выполнить адекватный прогноз.

Выполним анализ остатков. Зададим значение лага, равное 12. Построив частную автокорреляционную функцию (ЧАКФ), можно сделать вывод, что имеем дело со случайной составляющей типа «авторегрессия первого порядка (AR1)».

График ЧАКФ приведен на рис. 4.

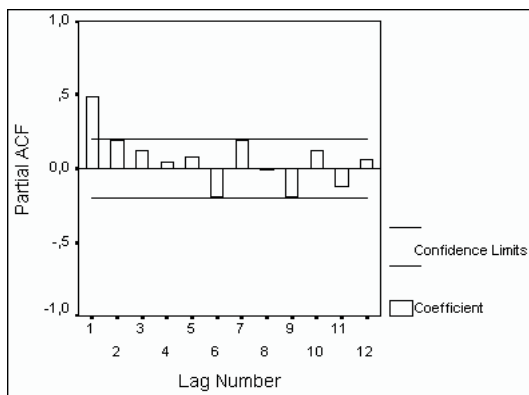


Рис. 4. График ЧАКФ

Случайная составляющая описывается следующим выражением:

$$\epsilon_t = \mu \cdot \epsilon_{t-1} + \alpha_t, \tag{2}$$

где ϵ_t – случайная составляющая типа AR1, в момент времени t ; ϵ_{t-1} – случайная составляющая типа AR1, в момент времени $(t-1)$; α_t – гауссовский белый шум; μ – числовой коэффициент ($\mu=0.52$).

Для описания нашего временного ряда, будем использовать аддитивную модель следующего вида:

$$x_t = tr_t + \epsilon_t, \tag{3}$$

где x_t – значения временного ряда; tr_t – значения подобранного нами тренда; ϵ_t – значения случайной составляющей типа AR1.

Основываясь на вышесказанном, выполним экстраполяцию имеющихся статистических данных. Полученный график приведен на рис. 5.

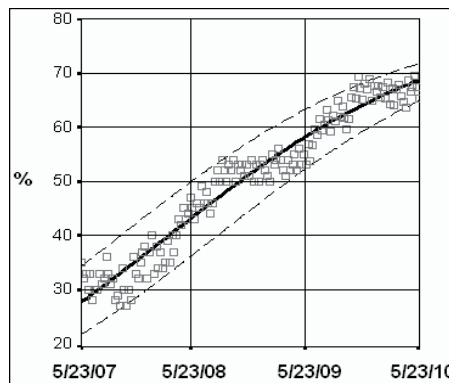


Рис. 5. Прогноз роста Internet-трафика на 5/23/2010

5. Результаты прогнозирования

Результаты прогнозирования свидетельствуют о том, что, с 5/23/2009 по 5/23/2010, относительная доля контентного трафика вырастет до 68%. Основываясь на этих цифрах, получим, что общий объем Internet-трафика увеличится приблизительно на 28%; однако, следует напомнить, что наш прогноз базируется лишь на данных одной из автономных систем (крупнейших). Видим, что полученный результат согласуется с аналогичными прогнозами мировых лидеров в области мониторинга Internet, соответственно, можем трактовать наши данные как достоверные.

Выводы

Таким образом, в данной статье было приведено краткое описание глобальной инфраструктуры мониторинга Internet – ATLAS, проанализирован характер современного Internet-трафика и выполнен прогноз его годового роста, который составил 28%, для чего была составлена соответствующая математическая модель. Полученный прогноз согласуется с аналогичными данными мировых лидеров в области мониторинга Internet и подтверждает адекватность подобранной модели.

Литература

1. C. Labovitz ATLAS Internet Observatory 2009 Annual Report / Labovitz C. [Электронный ресурс] – www.nanog.org/meetings/nanog47/presentations/Monday/Labovitz_ObserveReport_N47_Mon.pdf
2. Тюрин Ю. Н., Макаров А.А. Анализ данных на компьютере / Ю. Н. Тюрин, А. А. Макаров ; под. ред. В. Э. Фигурнова. – 3-е изд., перераб. и доп. – М. : ИНФРА-М, 2003. – 544 с. – ISBN 5-16-001316-4.