



***Information Technologies in Science,
Education, Telecommunication,
and Business***

**Труды международной
конференции “Информационные
технологии в науке,
образовании, телекоммуникации,
и бизнесе”**

Апробация технологии, выводы и рекомендации

Реализован КПК из гетерогенных компонент – модулей современных Oracle-систем финансового планирования и управления, унаследованной системы кадрового учета и слабоструктурированного архива мультимедиа. Реализация Интернет- и Интранет-порталов, манипулирующих компонентами гетерогенного КПК, обеспечила целый ряд внедрений в международной группе «ИТЕРА» с 10 тыс. сотрудников в 150 разнопрофильных компаниях. Комплексный характер технологии (модели, инструментарий, архитектура) обеспечивает сопряжение с широким диапазоном современных инструментальных средств разработки КПК. Функциональные преимущества подхода – манипулирование сложными разнородными и в разной степени структурированными объектами данных, объединение компонент с различной архитектурой – обусловлены ориентацией моделей и инструментальных средств на гетерогенные порталы КПК. По совокупной стоимости, срокам внедрения и возврату инвестиций результаты внедрения превосходят передовые аналоги в среднем на 30-40%. КПК внедрены в ряде корпораций («ИТЕРА», ИПУ РАН, МПЭ РФ и др.) [3,6].

ЛИТЕРАТУРА

1. Клещев А.С., Артемьева И.Л. Математические модели онтологий предметных областей.– Ч.2. Компоненты модели. // НТИ, сер. 2. 2001, № 3.– с.19-29.
2. Л.А. Калиниченко. СИНТЕЗ: язык определения, проектирования и программирования интероперабельных сред неоднородных информационных ресурсов.– М.: ИПИ РАН, 1993.
3. Зыков С.В. Проектирование корпоративных порталов. М.:МФТИ, 2005.– 258 с.
4. Вольфенгаген В.Э., Яцук В.Я. Аппликативные вычислительные системы и концептуальный метод проектирования систем знаний. — М: МО СССР, 1987.
5. Скотт Д.С. Области в денотационной семантике.– с.58-118.– в кн.: Математическая логика в программировании /Пер. с англ. – М.: Мир, 1991. – 408 с.
6. Zykov S.V. Enterprise Content Management: the Integrated Methodology. In: Proc. EISWT'07, Orlando, FL, U.S.A., July 9-12, 2007, p.p. 226-233.

УДК – 621.396

ПРИМЕНЕНИЕ СТАТИСТИКИ ЭКСТРЕМАЛЬНЫХ ЗНАЧЕНИЙ ДЛЯ АНАЛИЗА ТРАФИКА В КОРПОРАТИВНЫХ КОМПЬЮТЕРНЫХ СЕТЯХ

Репин Д.С., Филаретов Г.Ф.

APPLICATION OF EXTREME VOLUMES STATISTICS FOR TRAFFIC ANALYSIS IN CORPORATIVE COMPUTER NETS

Repin D.S., Filaretov G.F.

Extreme volumes statistics for traffic analysis in corporative computer nets is proposed. Possibilities of this approach for forecasting of critical situation, connected with limited channel capacity, are discussed.

1. Введение. Рассматривается задача анализа трафика в корпоративных компьютерных сетях. Предлагается для целей прогнозирования возникновения критических ситуаций в их функционировании, связанных с ограниченной пропускной способностью отдельных элементов сети, использовать математический аппарат статистики экстремальных значений.

Как известно, статистика экстремальных значений представляет собой определенную ветвь классической математической статистики, достаточно интенсивно развивавшуюся в 50 – 60 гг. прошлого века. В наиболее законченной форме ее основные результаты нашли свое отражение в монографии Э. Гумбеля [1].

В последнее время интерес к данной проблематике снова возрос – в первую очередь в прикладном плане применительно к решению ряда практических задач в метеорологии, гидрологии, анализе сейсмоопасности и сейсмостойкости, в материаловедении и строительстве. Представляется достаточно перспективной и возможность использования данного подхода и для анализа трафика в компьютерных сетях. Данное утверждение основывается в первую очередь на следующих общих соображениях.

Как отмечалось еще Э. Гумбелем, использование статистики экстремальных значений оказывается наиболее продуктивным при выполнении следующих двух условий:

- наблюдаемый процесс обладает свойством цикличности, т.е. определенной повторяемостью своего поведения через некоторый фиксированный интервал времени; желательно также, чтобы этот интервал задавался естественным путем, определяясь самой природой процесса, а не вводился искусственно. В частности, в гидротехнике, метеорологии этот интервал, естественно, равен 1 году;

- наибольший интерес с точки зрения исследователя (пользователя) представляют именно экстремальные значения, имеющие место на каждом цикле, поскольку именно они могут привести к

серьезным последствиям, скачкообразно возрастающим потерям и т.п. (прорыв дамбы или плотины, с последующим катастрофическим наводнением, обрушение зданий, сооружений и др.).

Именно такого рода ситуация и имеет место, если речь идет о трафике, контролируемом в различных точках компьютерной сети. На рис. 1 приведены образцы графиков изменения суточного трафика, полученных в ходе мониторинга компьютерной сети Московского энергетического института.

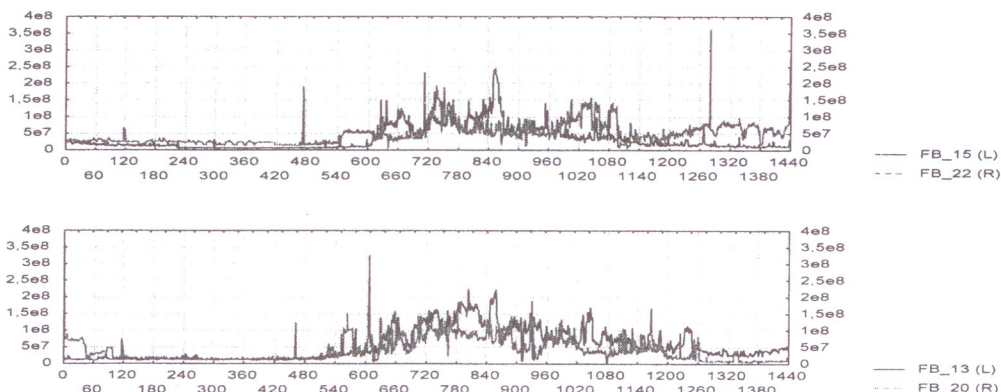


Рис. 1.

Очевидно, что, несмотря на индивидуальные различия, прослеживается общая повторяемость суточных графиков; кроме того, присутствуют и выбросы (экстремальные значения), которые могут привести к нарушениям в процессе передачи информации.

2. Основная идея и возможности метода анализа трафика, основанного на статистике экстремальных значений. Метод базируется на получении предварительной информации об интенсивности трафика, полученной путем наблюдения в течении ряда циклов повторяемости (например, ежесуточные наблюдения в течение месяца). Пусть, например, имеется N суточных реализаций, каждая из которых содержит L наблюдений x_{ln} ($l = 1, 2, \dots, L; n = 1, 2, \dots, N$) интенсивности трафика в бит/сек, полученной путем усреднения в течение 1 минуты ($L = 1440$). Зафиксируем для каждой реализации максимальное значение X_n . Известно, что существует всего два типа асимптотических (т.е. при N стремящемся к бесконечности) распределений для рассматриваемых максимальных значений. Очень важно, что эти распределения не зависят от функции распределения самих наблюдаемых величин x_{ln} , а область притяжения каждого из асимптотических распределений определяется только характером поведения «хвостов» функции плотности x_{ln} .

На рис. 2 изображены гистограммы реализаций, представленных на рис. 1. С их помощью

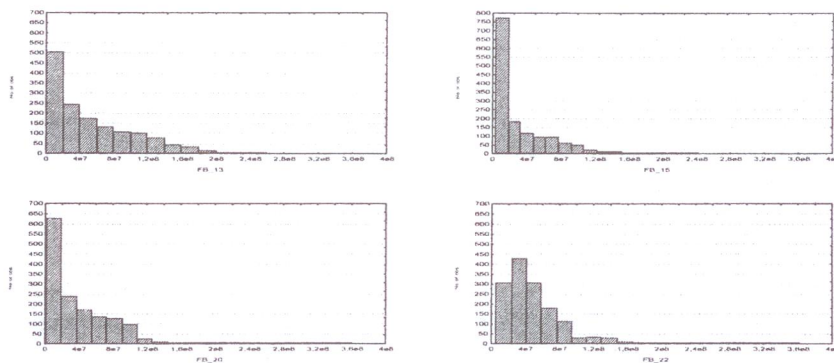


Рис. 2

достаточно обоснованно можно прийти к выводу об экспоненциальном характере затухания (поведения «хвоста») соответствующих функций плотности. Асимптотической функцией распределения вероятностей экстремальных значений для этого случая является, так называемое, первое предельное распределение или распределение Гумбеля: $F(x) = \exp[-e^{-\alpha(x-u)}]$. Здесь α и u - соответственно параметры формы и положения кривой интегральной функции распределения вероятностей.

Эмпирический анализ экстремальных значений и формирование различного рода заключений обычно проводится с помощью «вероятностной бумаги» [1]. На этой бумаге наблюдаемые значения X_n , размещенные в порядке возрастания в точках, соответствующих абсциссам $p_n = n/(N+1)$, будут лежать с некоторым разбросом около прямой линии $X = u + \frac{y}{\alpha}$, где y - приведенная переменная. Для

распределения Гумбеля $y = -\ln[-\ln p]$. Последующее изложение для наглядности будет вестись на модельном примере. Моделировался трафик, представляющий собой Пуассоновский поток с интенсивностью $3,8 \cdot 10^7$ бит/сек. Фиксировались усредненные за 1 минуту значения трафика; $L=1440$, $N=30$. Для каждой «суточной» реализации найдены максимальные значения X_n . Соответствующие точки нанесены на вероятностную бумагу для распределения Гумбеля (рис. 3) Как и ожидалось, все они группируются около прямой линии. С помощью метода наименьших квадратов можно оценить ее параметры. Получим: $X_m = 2,8861 \cdot 10^8 + 2,6767 \cdot 10^7 y$, т.е. $\hat{u} = 2,8861 \cdot 10^8$; $\hat{\alpha} = (2,6767 \cdot 10^8)^{-1}$.

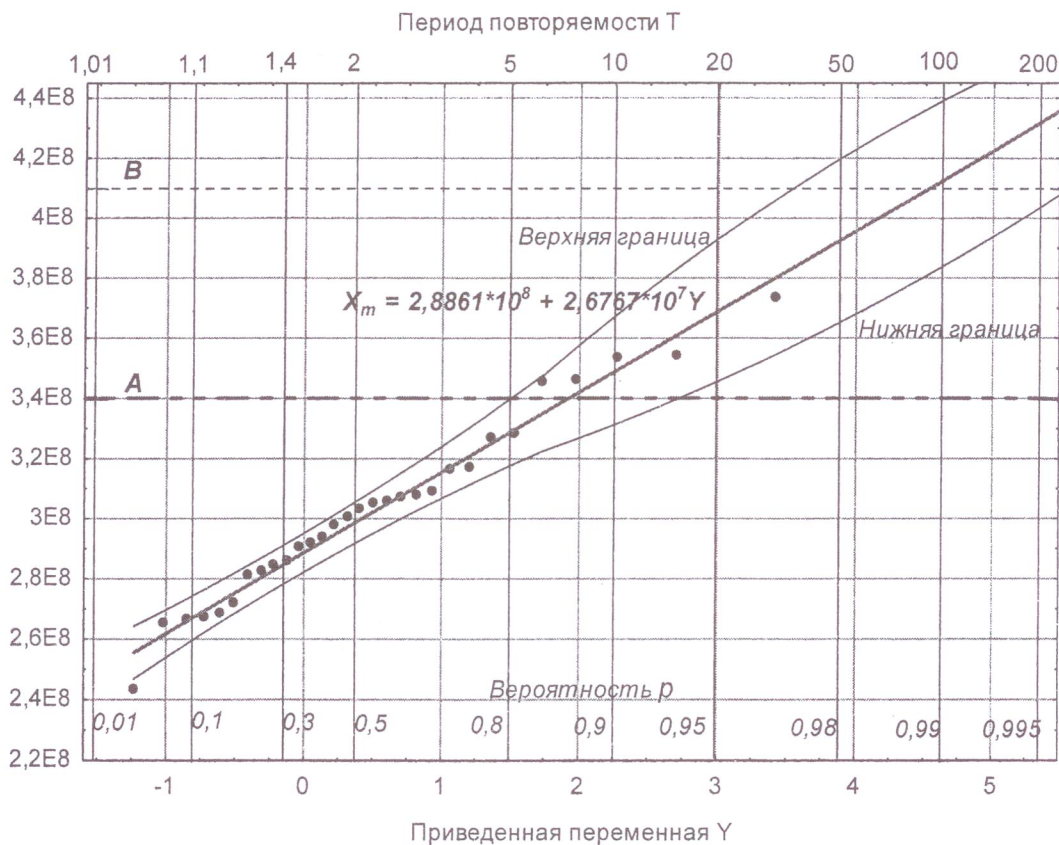


Рис.3.

Верхняя шкала на рис. 3 проградуирована в значениях периода повторяемости $T = 1/p$. Период T показывает, насколько часто в среднем будет происходить то или иное событие, связанное с превышением заданного уровня. Если, например, пользователя интересует уровень $3,4 \cdot 10^8$ (уровень **A**), то можно прогнозировать, что он будет превышаться в среднем примерно 1 раз в 7 суток. Аналогичным образом для уровня **B** получим $T=91$, т.е. превышение данного уровня будет происходить примерно 1 раз в 3 месяца.

На рис. 3, кроме аппроксимирующей прямой, указаны также контрольные граничные значения, построенные в соответствии с рекомендациями, изложенными в [1] для вероятности пребывания между указанными границами $P = 0,683$. Тот факт, что экспериментальные точки находятся внутри контрольной полосы наглядно свидетельствуют о том, что экстремальные значения в данном примере действительно подчиняются распределению Гумбеля. При этом сами границы можно использовать в качестве пессимистической T_n (верхняя граница) и оптимистической T_o (нижняя граница) оценок интервала повторяемости T . Для рассмотренных двух уровней получим соответственно $T_{nA} = 5$, $T_{oA} = 16$ и $T_{nB} = 28$, $T_{oB} = 258$. При увеличении интенсивности трафика значения периодов повторяемости будут уменьшаться. Интересно выяснить, насколько сильным будет это уменьшение. Для ответа на этот вопрос построены зависимости интервала повторяемости T от коэффициента увеличения интенсивности трафика (рис. 4)

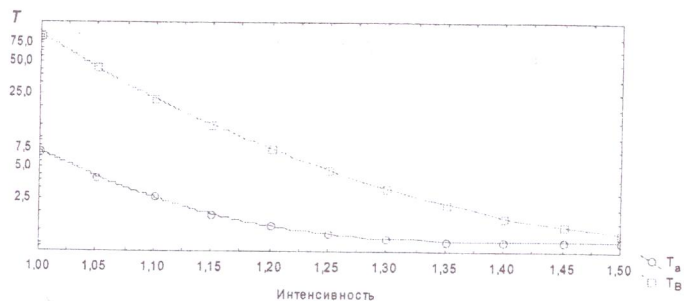


Рис. 4

Очевидно, что даже относительно небольшое увеличение трафика приводит к существенному уменьшению интервала повторяемости. В частности, для уровня В всего десятипроцентное увеличение интенсивности трафика приводит к более, чем трехкратному уменьшению T (с 91 до 25 суток), что необходимо учитывать при подключении новых пользователей к сети. Аналогично для уровня А имеет место примерно трехкратное уменьшение T .

ЛИТЕРАТУРА

1. Гумбель Э. Статистика экстремальных значений. Пер. с англ. — М.: Издательство «МИР», 1965. — 452 с.