

## **ИССЛЕДОВАНИЕ КОРРЕЛЯЦИОННОЙ ЗАВИСИМОСТИ МЕЖДУ КАНАЛОМ СВЯЗИ И АЛГОРИТМАМИ ПРИКЛАДНОГО ПО**

Рассматривается метод корреляционного анализа зависимости между качеством работы канала связи и алгоритмами прикладного программного обеспечения (ПО). Целью данного метода является отделение задачи повышения эффективности работы алгоритмов прикладного ПО от задач мониторинга и диагностирования компьютерных сетей, поскольку данные задачи решаются отдельно и независимо. Приводится практическая реализация этого метода и результаты его применения.

### **Введение**

Согласно [1], в компьютерных сетях (КС) часто встречается ситуация, когда необходимо определить, является ли причиной неудовлетворительного качества работы КС сама сеть или же влияние оказывают неэффективные алгоритмы работы прикладного ПО. В противном случае вероятность неоправданных затрат на модернизацию аппаратной платформы КС довольно высока, поскольку причина ее плохой работы может заключаться не в недостаточной пропускной способности сетевого оборудования или низкой производительности сервера, а в неэффективных алгоритмах самого прикладного ПО или его неправильной настройке. Под неэффективными алгоритмами работы прикладного ПО понимаются такие алгоритмы, применение которых не позволяет максимально полно использовать пропускную способность ресурсов сети (перекрытие запросов с ответами, передача данных пакетами минимальной длины и пр.).

Обратная ситуация, когда прикладное ПО неадекватно работает вследствие наличия узких мест в КС, причем возможность доказать это эмпирическим путем весьма сомнительна, тоже может иметь место.

Таким образом, следует включить в процессы мониторинга и диагностирования КС решение задачи отделения неудовлетворительного функционирования сети от неэффективных алгоритмов прикладного ПО, а разработка соответствующих методов и алгоритмов для корпоративных компьютерных сетей – весьма актуальна в настоящее время.

### **Постановка задачи**

Для решения задачи отделения гипотезы «неудовлетворительное функционирование сети» от гипотезы «неэффективные алгоритмы прикладного ПО» предлагается ввести два наблюдаемых параметра, один из которых характеризует качество работы сети на всех уровнях ее функционирования, включая прикладной, а второй характеризует только качество работы транспортной платформы КС без учета функционирования Software-уровней сетевой модели. Первым параметром,  $y_1$ , будет время реакции прикладного ПО на запрос клиента, вторым,  $y_2$ , – скорость работы сети.

Корреляционный анализ позволит выяснить, существует ли корреляционная линейная связь между параметрами «скорость работы сети» и «время реакции ПО» и, следовательно, определить носителя неисправности: если связь существует, то носитель неисправности – компонент сети, в противном случае – прикладное ПО.

### **Алгоритм отделения неисправностей ПО от неисправностей КС**

1) Выполняется запуск типовой задачи прикладного ПО, используемого в сети с сервера на любом из узлов-клиентов (так как по умолчанию принимается нуль-гипотеза: «прикладное ПО – носитель неисправности», инвариантная к структуре КС).

2) Выполняется генерация трафика с клиента, выполняющего типовую задачу на сервере, с последовательным увеличением нагрузки (шаг – 10%) на сеть (канал связи).

3) Параллельно с увеличением нагрузки на каждом шаге выполняется измерение каждого из наблюдаемых параметров в один момент времени. При этом скорость работы сети измеряется как время получения пакета от клиента сервером на уровне их сетевых

адаптеров. Например, анализатор протоколов Observer представляет эту информацию в виде параметра Station1-Station2 функции Time Interval Dynamics. Параметр «время реакции прикладного ПО» измеряется как “latency” – разность между запросом клиента и ответом сервера. Эта информация также представлена в Observer в виде параметра Station1/Port-Station2/Port функции Time Interval Dynamics. Так как интересует время реакции только от типовой задачи наблюдаемого прикладного ПО, то необходимо знать порты, которые используются этой типовой задачей. Это легко сделать с помощью функции Decode, встроенной в анализатор протоколов Observer. Все данные, полученные функцией Time Interval Dynamics, могут быть легко сохранены и переконвертированы в табличную форму для статистической обработки, в частности, для выполнения корреляционного анализа.

4) Количество измерений производится с учетом ограничений границ выборки для метода ранговой корреляции, а также с учетом варьирования нагрузки [2]:

$$5 < \frac{(N-n) \cdot g}{n} \leq 20, \quad (1)$$

где  $N$  – количество измерений;  $n$  – количество параметров;  $g$  – количество уровней варьирования нагрузки.

5) Временной интервал между измерениями (тестами) в ходе сбора статистической информации вычисляется по формуле:

$$I_t = \frac{Q_h}{N}, \quad (2)$$

здесь  $I_t$  – интервал времени между измерениями в ходе сбора статистической информации;  $Q_h$  – время, затраченное на сбор статистической информации на каждом уровне варьирования;  $N$  – количество измерений (тестов), вычисленных по формуле (1).

6) Корреляционный анализ. Между парами параметров необходимо вычислить коэффициент парной корреляции, который является общепринятой в математической статистике характеристикой связи между двумя случайными величинами. Если обозначить один параметр через  $y_1$ , а другой – через  $y_2$ , и число опытов, в которых они будут измеряться, – через  $N$  так, что  $u=1,2,\dots,N$ , где  $u$  – текущий номер опыта, то коэффициент парной корреляции  $r$  вычисляется по формуле:

$$r_{y_1 y_2} = \frac{\sum_{u=1}^N (y_{1u} - \bar{y}_1)(y_{2u} - \bar{y}_2)}{\sqrt{\sum_{u=1}^N (y_{1u} - \bar{y}_1)^2 \sum_{u=1}^N (y_{2u} - \bar{y}_2)^2}}. \quad (3)$$

Здесь  $\bar{y}_1 = \sum_{u=1}^N \frac{y_{1u}}{N}$  и  $\bar{y}_2 = \sum_{u=1}^N \frac{y_{2u}}{N}$  – средние арифметические соответственно для  $y_1$  и  $y_2$ .

Необходимо помнить, что коэффициент парной корреляции как мера тесноты связи имеет четкий математический смысл только при линейной зависимости между параметрами и в случае их нормального распределения. Для проверки значимости коэффициента парной корреляции нужно сравнить его значение с табличным (критическим) значением  $r$  [3,4]. Для пользования этой таблицей нужно знать число степеней свободы  $f=N-2$  и выбрать определенный уровень значимости, например, равный 0,05. Такое значение уровня значимости соответствует вероятности верного ответа при проверке гипотезы  $p=1-\alpha=1-0.05=0.95$  или 95%. Это значит, что в среднем только в 5% случаев возможна ошибка при проверке гипотезы.

Если экспериментально найденное значение  $r$  больше или равно критическому, то гипотеза о корреляционной линейной связи подтверждается, а если меньше, то нет оснований считать, что имеется тесная линейная связь между параметрами.

## Практическая реализация алгоритма

Генерация трафика (второй пункт алгоритма отделения неисправностей прикладного ПО от неисправностей КС) выполняется с помощью функции «Генератор трафика» анализатора протоколов Observer из пункта меню «Инструменты». Адрес назначения генерируемого трафика задается опцией Set Destination Address всплывающего меню поля содержимого генерируемого пакета либо параметром Destination Address (рис. 1, 2).

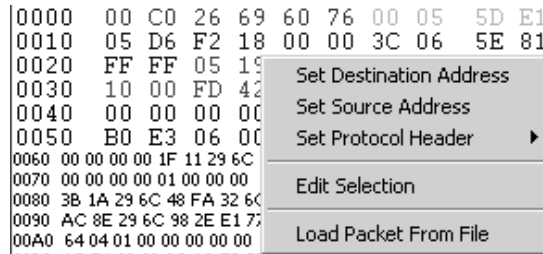


Рис. 1. Задание адреса назначения в теле генерируемого пакета

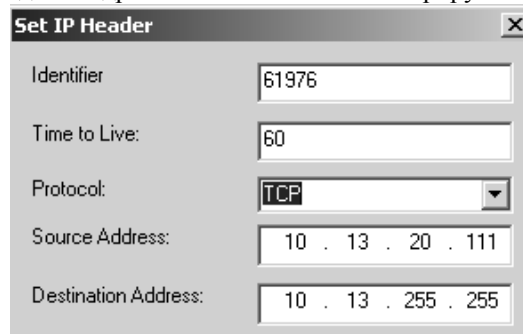


Рис. 2. Задание адреса назначения в IP-заголовке

Интенсивность трафика задается следующим образом (рис. 3).

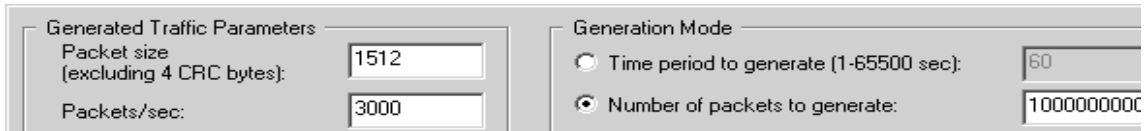


Рис. 3. Установки интенсивности трафика

Изменяемыми параметрами, влияющими на уровень интенсивности трафика, являются: размер пакета (максимально допустимый размер пакета совпадает с максимальным корректным значением размера кадра канального уровня), интенсивность передачи пак/с, период генерации трафика (временной интервал либо объем трафика в пакетах, который необходимо сгенерировать).

Выполняется генерация трафика с клиента, выполняющего типовую задачу на сервере, с последовательным увеличением нагрузки (шаг – 10%) на сеть (канал связи) (рис. 4).

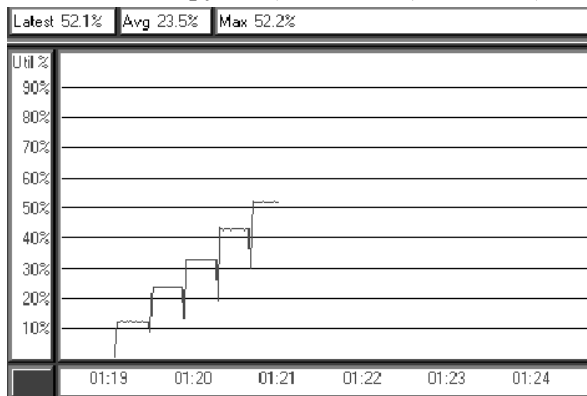


Рис. 4. Характер нагрузки при использовании метода отделения неисправностей прикладного ПО от неисправностей КС

Далее выполняется сохранение отчетов по контрольным точкам Station1-Station2 и по Station1/Port-Station2/Port (рис. 5) в одинаковом временном интервале, перекодирование результатов в табличную форму, их совмещение и выполнение корреляционного анализа.

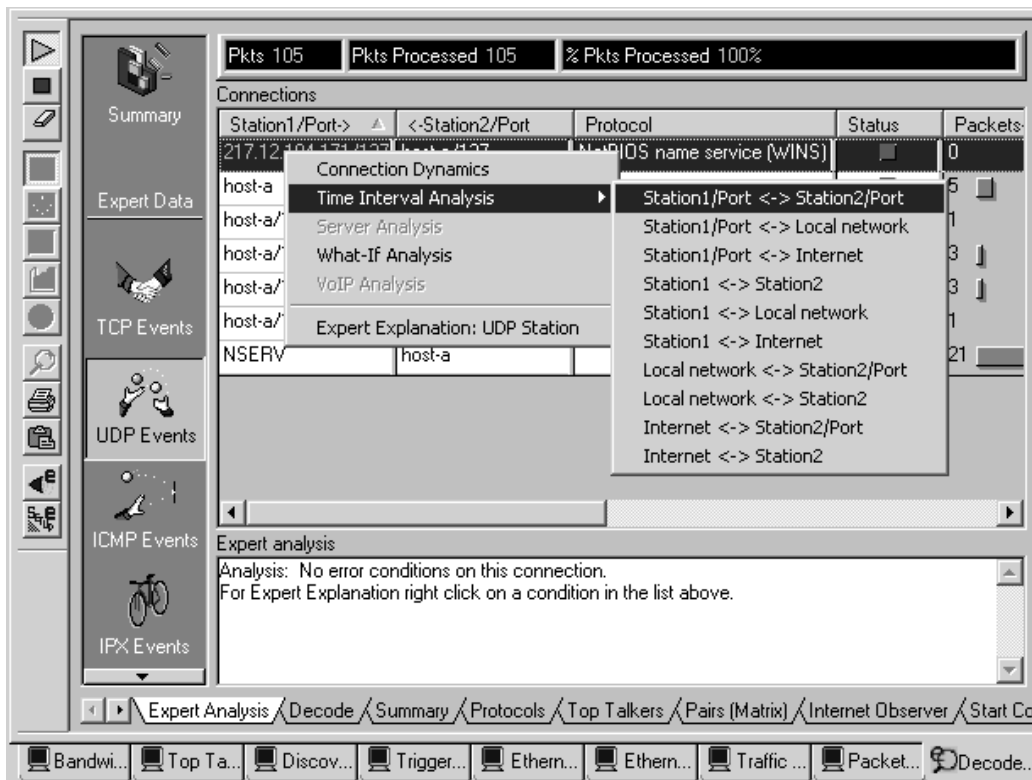


Рис. 5. Просмотр и сохранение результатов подачи ступенчатой нагрузки

Ниже приведена таблица статистических данных для корреляционного анализа, а также результаты корреляционного анализа (рис. 6). Здесь St1-St2 delay – время передачи пакета между сетевыми адаптерами Станции 1 и Станции 2. Этот параметр характеризует качество канала связи. St1/P-St2/P delay – время передачи и обработки пакета между Station1/Port и Station2/Port на прикладном уровне, этот параметр характеризует производительность прикладного ПО. Utilization – характеристика увеличения сетевой нагрузки. Количество измерений N на каждом уровне варьирования нагрузки, согласно (1), равно 11. Интервал между снятиями показаний  $I_t$  при  $Q_h = 30$  минут составляет 3 минуты.

D	E	F	G	H	I
0,20	0,31			Столбец 1	Столбец 2
1,10	5,00		Столбец 1	1	
0,20	1,00		Столбец 2	0,99229774	1
0,50	1,00				
1,10	7,00				
n 2n	1 5n				

Рис. 6. Результаты корреляционного анализа в Microsoft Excel

Здесь степень свободы  $f=11$ , уровень значимости  $=0,05$ ,  $r_{exp} = 0,993$ ,  $r_{tab}$  по указанным условиям равно  $0,553$  [4]. Так как  $r_{exp} > r_{tab}$ , то имеет место линейная корреляция наблюдаемых параметров, и можно сделать вывод о том, что причина неудовлетворительного качества работы КС заключается в самой сети, а не в прикладном ПО.

Дальнейшие эмпирические исследования показали, что причиной медленной работы компьютерной сети было узкое место: один из коммутаторов на тракте передачи данных не справлялся с входящим потоком и отбрасывал пакеты. После замены на коммутатор с большей буферной памятью проблема с замедлениями устранилась.

Статистические данные для корреляционного анализа

Time	St1↔St2 delay,ms	St1/P↔St2/P delay,ms	Utilization, %
09:05:00:000	0.2	0.31	x...10
09:08:00:000	1.1	5	x...10
09:11:00:000	0.2	1	x...10
09:14:00:000	0.5	1	x...10
09:17:00:000	1	7	x...10
09:20:00:000	0.2	1.5	x...10
09:23:00:000	1	7	x...10
09:26:00:000	2.7	10	x...10
09:29:00:000	1	7	x...10
09:32:00:000	2.3	7	x...10
09:35:00:000	2.7	10	x...10
09:38:00:000	7.9	28	10...20
09:41:00:000	10.6	37	10...20
09:44:00:000	17.4	62	10...20
09:47:00:000	10.7	37	10...20
09:50:00:000	4.5	15	10...20
09:53:00:000	5.7	19	10...20
09:56:00:000	5.5	19	10...20
09:59:00:000	10.1	35	10...20
10:02:00:000	8.8	30	10...20
10:05:00:000	5.9	21	10...20
10:08:00:000	6.8	24	10...20
10:11:00:000	11.4	41	20...30
10:14:00:000	12.4	44	20...30
10:17:00:000	14.3	48	20...30
10:20:00:000	12.4	44	20...30
10:23:00:000	12.3	44	20...30
10:26:00:000	13.2	57	20...30
10:29:00:000	6.9	24	20...30
10:32:00:000	5.3	19	20...30
10:35:00:000	12.8	48	20...30
10:38:00:000	9.8	34	20...30
10:41:00:000	9.4	33	20...30
10:44:00:000	11	40	30...40
10:47:00:000	8.2	30	30...40
10:50:00:000	15.2	60	30...40
10:53:00:000	12.6	46	30...40
10:56:00:000	14.1	49	30...40
10:59:00:000	5.6	21	30...40
11:02:00:000	8.8	30	30...40
11:05:00:000	12.2	44	30...40
11:08:00:000	7.5	28	30...40
11:11:00:000	9	35	30...40
11:14:00:000	8.8	32	30...40

### Выводы

В данной статье рассмотрен метод корреляционного анализа зависимости между качеством работы канала связи и алгоритмами прикладного ПО. Научная новизна метода заключается в применении к задаче, обозначенной в статье и традиционно решаемой эмпирическим путем, математического аппарата. Практическая значимость состоит в формализации процесса отделения задачи верификации прикладного ПО от задач мониторинга и диагностирования компьютерных сетей, что приводит к снижению трудоемкости обслуживания компьютерных сетей и поиска сетевых неисправностей. В перспективах

исследования – дальнейшее развитие применения методов статистической обработки данных к процессам мониторинга и диагностирования компьютерных сетей.

**Список литературы:** 1. В. Подлазов, В. Борисенко, С. Юдицкий. Узкие места в локальных сетях - <http://www.osp.ru/lan/1998/09/133684/>. 2. Цейтлин Н.А. Процессы и аппараты производства основной химии: Труды НИОХИМ. Т. 56. Харьков, 1981. С. 29-38J. 3. Казаков Ю.Б. Электронный конспект лекций по предмету «Методы планирования эксперимента». <http://www.ispu.ru/portal/index.pl> 4. Хамханов К.М. Основы планирования эксперимента: Методическое пособие. Улан-Удэ: ВСГТУ, 2001. 94 с.

*Поступила в редколлегию 22.05.2010*

**Бабич Анна Витальевна**, канд. техн. наук, доцент кафедры АПВТ ХНУРЭ. Научные интересы: сетевые технологии, технологии дистанционного образования. Увлечения: активный отдых, путешествия, иностранные языки. Адрес: Украина, 61166, Харьков, пр. Ленина, 14, тел. 70-21-326. E-mail: babich@kture.kharkov.ua

**Емельянов Игорь Валерьевич**, аспирант кафедры АПВТ ХНУРЭ. Научные интересы: сетевые технологии, диагностика и оптимизация компьютерных систем и сетей. Адрес: Украина, 61166, Харьков, пр. Ленина, 14, тел. 70-21-326.

**Мурад Али А.**, аспирант кафедры АПВТ ХНУРЭ. Научные интересы: компьютерные сети следующего поколения. Адрес: Украина, 61166, Харьков, пр. Ленина, 14, тел. 70-21-326.