

ДИСПЕРСИОННЫЙ АНАЛИЗ В ПРОЦЕССАХ ПАССИВНОГО МОНИТОРИНГА И ДИАГНОСТИРОВАНИЯ КОРПОРАТИВНЫХ КОМПЬЮТЕРНЫХ СЕТЕЙ

Рассматриваются аспекты применения дисперсионного анализа в процессах пассивного мониторинга и диагностирования корпоративных компьютерных сетей (ККС). Целью применения дисперсионного анализа является выявление фактора, оказывающего наибольшее влияние на критерий качества работы компьютерной сети, что входит в задачи диагностирования ККС. Описывается модель ККС как объекта мониторинга. Формулируются правила выполнения дисперсионного анализа в рамках модели ККС как объекта мониторинга. Выполняется практическая реализация и описываются экспериментальные результаты применения дисперсионного анализа в процессах пассивного мониторинга и диагностирования ККС.

Введение

Согласно [1], внедрение статистических методов анализа и мониторинга гетерогенных телекоммуникационных сетей, к каковым относятся корпоративные компьютерные сети (ККС), на основе перехвата и декодирования трафика с использованием доступного инструментария (анализаторы протоколов, встроенные средства операционных систем для мониторинга системы и сети и др.) является задачей, в настоящее время требующей решения. Это обусловлено необходимостью эксплуатации и надежного функционирования достаточно большого набора инфокоммуникационных сервисов, обеспечивающих эффективную работу пользователя с разнородной информацией в ККС. При росте числа упомянутых сервисов возрастает сложность обслуживания и обеспечения должного уровня качества работы компьютерной сети. Как следствие, возрастает трудоемкость процессов мониторинга и диагностирования ККС, и необходимость формализации упомянутых про-

цессов и задействования математических методов в ходе постановки диагноза представляется очевидной.

Исходя из [2], под мониторингом ККС будем понимать комплекс мер и процедур, включающий в себя сбор информации о работе основных компонентов и систем сети, анализ собранной информации и, как следствие, выполненного анализа – определение нормального профиля сети и выявление отклонений от него. Установление причин появления этих отклонений относится к задачам диагностирования ККС и непосредственно базируется на результатах работы процессов мониторинга.

Особенностью пассивного мониторинга является отсутствие активных воздействий на объект наблюдения со стороны наблюдателя, что дает возможность поставить в соответствие пассивному мониторингу функциональное диагностирование в терминах [3]. С одной стороны, такой режим диагностирования не требует прекращения штатной работы диагностируемого устройства, что является существенным преимуществом с точки зрения пользователей ККС. С другой стороны, для постановки достоверного диагноза по результатам пассивного мониторинга требуется гораздо больший объем собранных данных в сравнении с активным мониторингом, что ведет к нелинейному возрастанию трудоемкости и временных затрат на их анализ. В совокупности со сказанным это подтверждает актуальность решения задачи разработки формализованных методов и алгоритмов статистической обработки данных, собранных в ходе пассивного мониторинга, в целях снижения трудоемкости и повышения эффективности процессов мониторинга и диагностирования ККС.

Модель ККС как объекта мониторинга

ККС как объект мониторинга (ОМ) определяется соотношением (1), связывающим критерий качества работы ККС y с факторами, оказывающими на него влияние:

$$y = \varphi(S_p\{x_1, x_2 \dots x_k\}, S_a\{x_1, x_2 \dots x_l\}, S_{sys}\{x_1, x_2 \dots x_m\}, S_{nos}\{x_1, x_2 \dots x_n\}), \quad (1)$$

где y – время реакции прикладного ПО сервера на запрос клиента; факторы, характеризующие компоненты ККС как ОМ и влияющие на значение критерия качества работы КС, представлены следующими множествами: $S_p\{x_1, x_2 \dots x_k\}$ – совокупность характеристик кабельной системы и другого пассивного оборудования; $S_a\{x_1, x_2 \dots x_l\}$ – совокупность характеристик активного сетевого оборудования (сетевые платы, концентраторы, коммутаторы, маршрутизаторы); $S_{sys}\{x_1, x_2 \dots x_m\}$ – совокупность характеристик системных ресурсов сервера и рабочих станций; $S_{nos}\{x_1, x_2 \dots x_n\}$ – совокупность конфигурационных и сетевых настроек сетевой операционной системы.

Особенностью ККС как ОМ является то, что при исправном техническом состоянии каждого из сетевых компонентов в отдельности может иметь место ситуация, когда работа сети в целом не отвечает требуемому уровню качества и, следовательно, ККС с точки зрения пользователей и выполняемых в ней сетевых задач исправной не является. Это объясняется наличием ошибок на этапе проектирования и развертывания сети, несбалансированностью нагрузки и сетевых компонентов, использованием ПО с неэффективными алгоритмами реализации. Таким образом, становится актуальной задача определения степени влияния каждого из факторов на критерий качества работы ККС, что, при неудовлетворительных значениях критерия, позволит определить, какой из факторов с наибольшей вероятностью является потенциальным носителем причины их появления.

С другой стороны, модель ККС как ОМ можно представить в виде ациклической древовидной структуры $H = \langle V, W \rangle$ [4], где $V = D \cup E \cup SS \cup F$ – множество компонентов компьютерной сети; $W \subseteq D \times E \times SS \times F$ – иерархические взаимосвязи между ними. Здесь элемент d подмножества D – компонент как устройство, принадлежащее подмножеству высшего уровня; $e \in E$ – компонент как элемент устройства; $ss \in SS$ – компонент как подсистема элемента устройства; $f \in F$ – фактор, принадлежащий подмножеству низшего уровня ККС как ОМ.

Взаимосвязи между указанными подмножествами и их элементами организуются следующим образом (рис. 1):

1) для каждого $ss \in SS$ существует группа элементов $f_H(ss) \subseteq F$, для которых ss является общей вершиной;

- 2) для каждого $e \in E$ существует группа элементов $ss_H(e) \subseteq SS$, для которых e является общей вершиной;
- 3) для каждого $d \in D$ существует группа элементов $e_H(d) \subseteq E$, для которых d является общей вершиной.



Рис. 1. Иерархическая структура ККС как ОМ

Чем глубже вложенность компонентов, тем выше эффективность данных, собранных на начальном этапе мониторинга ККС, с точки зрения определения нормального профиля ККС и отклонений от него. Рис. 2 иллюстрирует пример построения иерархической структуры ККС как ОМ (рис.2, б) для тракта передачи данных из рис.2, а.

Дисперсионный анализ в процессах мониторинга и диагностирования

Дисперсионный анализ (ДА) – статистический метод, позволяющий анализировать влияние различных факторов на исследуемую переменную путем проверки значимости различия между средними с помощью сравнения дисперсий [5]. Таким образом, ДА является инструментом, с помощью которого непосредственно решается задача определения степени влияния факторов, описывающих модель ККС как ОМ в соотношении (1), на критерий качества работы сети y , о важности которой уже было сказано выше.

Для обеспечения возможности применения ДА к статистическим данным, собранным в процессе пассивного мониторинга, необходимо, чтобы в ходе сбора обеспечивалось варьирование значений факторов. Поскольку при пассивном мониторинге исследователь не оказывает никаких активных воздействий на характеристики сети, что гарантированно обеспечило бы требуемое варьирование по уровням, то необходимо проводить наблюдения за ККС как ОМ таким образом, чтобы варьирование обеспечивалось типовыми задачами, решаемыми в ККС. Для этого необходимо знать характер и периодичность интенсивности бизнес- и иных процессов, выполняемых на предприятии с применением ККС.

В процессах мониторинга и диагностирования ККС будет применяться как однофакторный, так и многофакторный ДА в соответствии со следующими правилами, базирующимися на иерархической структуре ККС как ОМ:

- 1) Если факторы принадлежат одной вершине $ss \subseteq SS$, – выполняется однофакторный ДА, при этом $p=n$, где p – число, определяющее, сколько раз будет выполняться однофакторный ДА, n – число факторов, принадлежащих заданной вершине ss . Таким образом, выполняется анализ каждого фактора на предмет обнаружения оказывающего наибольшее влияние на критерий качества работы сети. Данную задачу целесообразно решать в рамках оптимизации (или «тонкой» настройки) ККС.

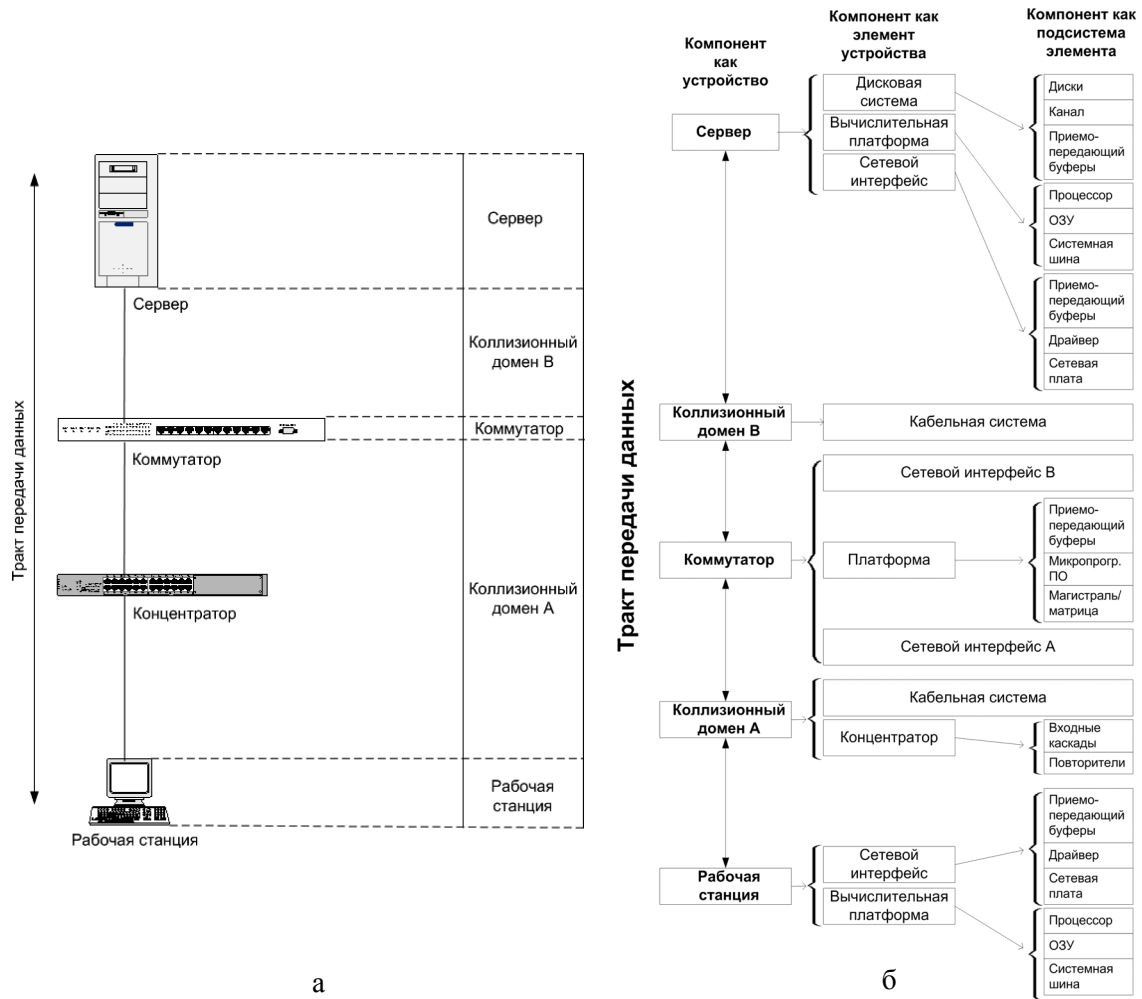


Рис.2. Построение иерархической структуры ККС как ОМ: а – тракт передачи данных (ТПД) ККС; б – иерархическая структура ТПД

2) Если факторы принадлежат разным вершинам $ss \subseteq SS$, но при этом каждая из этих вершин принадлежит такой группе элементов $ss_H(e) \subseteq SS$, для которой имеется общая вершина $e \in E$, – выполняется n -факторный ДА, где n – количество элементов $ss_H(e) \subseteq SS$ с общей вершиной e . Таким образом, решается задача обнаружения потенциального носителя причины отклонения от нормального профиля ККС на уровне компонента как подсистемы элемента. Данная задача решается в рамках диагностирования ККС (обнаружение неисправности и постановка диагноза).

3) N -факторный ДА применяется также и тогда, когда факторы принадлежат разным вершинам $ss \subseteq SS$ и для этих ss нет общей вершины e , но есть общая вершина $d \in D$. Здесь n – количеству элементов группы $e_H(d) \subseteq E$ с общей вершиной d . Решается задача обнаружения потенциального носителя причины отклонения от нормального профиля ККС на уровне компонента как элемента устройства. Данная задача так же, как и в п.2, решается в рамках диагностирования ККС, но с меньшей детализацией постановки диагноза.

4) Если факторы принадлежат разным вершинам $ss \subseteq SS$ и у них нет общих вершин ни в подмножестве E , ни в подмножестве D , – рекомендуется выполнять двухфакторный ДА, при этом выполнять его p раз, $1 \leq p \leq C_d^2$, где C_d^2 – количество сочетаний по 2 из всех

$$d \subseteq D : C_d^2 = \frac{d!}{2!(d-2)!}.$$

Здесь решается задача обнаружения потенциального носителя причины отклонения от нормального профиля ККС на уровне компонента как устройства, что является начальным этапом диагностирования ККС.

Практическая реализация и экспериментальные данные

Рассмотрим пример выявления компонента как устройства, оказывающего наибольшее влияние на критерий качества работы сети. Выполним дисперсионный анализ для статистических данных по факторам «утилизация процессора сервера, %» (фактор А) и «загрузка магистрального коммутатора, %» (фактор Б), где наблюдаемой переменной является критерий качества работы сети y «время реакции прикладного ПО сервера на запрос клиента, ms». В соответствии с приведенными выше правилами выбора типа ДА, а именно с п.4, здесь следует применить двухфакторный ДА. Согласно [5], двухфакторная дисперсионная модель имеет вид:

$$x_{ijk} = \mu + F_i + G_j + I_{ij} + \varepsilon_{ijk}, \quad (2)$$

где x_{ijk} – значение наблюдения в ячейке ij с номером k ; μ – общая средняя; F_i – эффект, обусловленный влиянием i -го уровня фактора А, где $i = \overline{1, m}$; G_j – эффект, обусловленный влиянием j -го уровня фактора Б, где $j = \overline{1, l}$; I_{ij} – эффект, обусловленный взаимодействием двух факторов, т.е. отклонение от средней по наблюдениям в ячейке ij от суммы первых трех слагаемых в модели (2); ε_{ijk} – возмущение, обусловленное вариацией переменной внутри отдельной ячейки ij .

Предполагается, что ε_{ijk} имеет нормальный закон распределения $N(0; \sigma^2)$, а все математические ожидания F, G, I_{ij} равны нулю. Групповые средние вычисляются следующим образом:

в ячейке: $\bar{x}_{ij} = \frac{\sum_{k=1}^n x_{ijk}}{n}$, по строке: $\bar{x}_i = \frac{\sum_{j=1}^l \bar{x}_{ij}}{l}$, по столбцу: $\bar{x}_j = \frac{\sum_{i=1}^m \bar{x}_{ij}}{m}$, общая средняя:

$$\bar{x} = \frac{\sum_{i=1}^m \sum_{j=1}^l \bar{x}_{ij}}{ml}, \text{ где } n \text{ – количество наблюдений в ячейке } ij.$$

В табл. 1 представлен общий вид вычисления значений с помощью дисперсионного анализа.

Таблица 1. Базовая таблица двухфакторного дисперсионного анализа

Компоненты дисперсии	Сумма квадратов	Число степеней свободы	Средние квадраты
Межгрупповая (фактор А)	$Q_1 = l \cdot n \sum_{i=1}^m (\bar{x}_i - \bar{x})^2$	$m-1$	$S_1^2 = \frac{Q_1}{m-1}$
Межгрупповая (фактор Б)	$Q_2 = m \cdot n \sum_{j=1}^l (\bar{x}_j - \bar{x})^2$	$l-1$	$S_2^2 = \frac{Q_2}{l-1}$
Взаимодействие	$Q_3 = n \sum_{i=1}^m \sum_{j=1}^l (\bar{x}_{ij} - \bar{x}_i - \bar{x}_j + \bar{x})^2$	$(m-1)(l-1)$	$S_3^2 = \frac{Q_3}{(m-1)(l-1)}$
Остаточная	$Q_4 = \sum_{i=1}^m \sum_{j=1}^l \sum_{k=1}^n (x_{ijk} - \bar{x}_{ij})^2$	$mln - ml$	$S_4^2 = \frac{Q_4}{ml(n-1)}$
Общая	$Q = \sum_{i=1}^m \sum_{j=1}^l \sum_{k=1}^n (x_{ijk} - \bar{x})^2$	$mln - 1$	

Проверка нулевых гипотез H_A, H_B, H_{AB} об отсутствии влияния на рассматриваемую переменную факторов А, Б и их взаимодействия АБ осуществляется сравнением отноше-

ний $\frac{S_1^2}{S_4^2}, \frac{S_2^2}{S_4^2}, \frac{S_3^2}{S_4^2}$ с соответствующими табличными значениями F–критерия Фишера. Если $F_{\text{эксп.}} < F_{\text{табл.}}$ – гипотеза об отсутствии влияния принимается, в противном случае отвергается.

Сбор и анализ статистических данных по заданным выше факторам выполнялся при следующих условиях: $m=2$, $A_1=20\%$, $A_2=75\%$ при пороговом значении фактора 90%; $l=3$, $B_1=10\%$, $B_2=15\%$, $B_3=25\%$ при пороговом значении фактора 35%; $n=4$.

Таблица 2. Двухфакторный ДА для выявления компонента как устройства

	B_1	B_2	B_3	средние в ячейке (ij):		ср. по стр. (i):	
A_1	10,00	8,00	13,00	x_{11}	8,50	x_1	9,50
	8,00	9,00	15,00	x_{12}	8,25	x_2	50,42
	7,00	7,00	10,00	x_{13}	11,75		
	9,00	9,00	9,00			ср. по стлб. (j):	
A_2	50,00	55,00	55,00	x_{21}	47,50	x_1	28,00
	45,00	50,00	60,00	x_{22}	51,25	x_2	29,75
	40,00	55,00	45,00	x_{23}	52,50	x_3	32,13
	55,00	45,00	50,00	общая средняя:			
				x	29,96		
суммы квадратов:			средние квадраты:		критерий Фишера (эксп.):		
Q_1	837,09	10045,04	S^2_1	10045,04	ф. А	517,87	
Q_2	8,57	68,58	S^2_2	34,29	ф. Б	1,77	
Q_3	4,02	16,06	S^2_3	8,03	ф. АБ	0,41	
Q_4		349,23	S^2_4	19,40	критерий Фишера (табл.):		
					ф. $A_{(0,05;1;18)}$	4,41	
					ф. $B_{(0,05;2;18)}$	3,55	
					ф. $AB_{(0,05;2;18)}$	3,55	

При указанных условиях для наблюдаемой переменной y были получены результаты, приведенные в табл.2, выполнен двухфакторный ДА, ход которого также отображен в табл. 2, получены экспериментальные значения критерия Фишера по каждому из факторов и их взаимодействию. Из табл. 2 видно, что нулевая гипотеза H_0 отвергается, нулевые гипотезы H_B и H_{AB} - принимаются. Более того, по существенному разрыву между экспериментальным и табличным значениям критерия Фишера можно однозначно утверждать, что в данном случае потенциальным критическим компонентом компьютерной сети является устройство, характеризуемое фактором А. Дальнейшие исследования показали, что сервер, после подключения его к гигабитному порту коммутатора, не справляется с экспоненциально возросшим уровнем входящего трафика в периоды осуществления массовых операций с прикладным ПО. Рекомендуется модернизация вычислительной платформы сервера.

Выводы

Рассмотрены модель и метод пассивного мониторинга и диагностирования ККС на базе дисперсионного анализа. Их научная новизна заключается в применении к задачам, обозначенным в статье и традиционно решаемым эмпирическим путем, аппарата математической статистики. Практическая значимость состоит в формализации процесса обработки статистических данных, собранных в ходе пассивного мониторинга, что приводит к снижению трудоемкости и повышению эффективности процессов мониторинга и диагностирования ККС. В перспективах исследования – дальнейшее развитие применения методов статистической обработки данных к процессам активного мониторинга и диагностирования ККС.

Список литературы: 1. *Скуратов А.К.* Использование временных рядов для целей статистического анализа телекоммуникационных сетей на основе исследования информационных потоков // Вестник Новгородского Государственного университета. 2005. №34. С. 112-117. 2. *Киселев Г.Д., Шпакаускас М.С.* Мониторинг мультисервисных компьютерных сетей средствами системы Nagios // Электроника и связь. Тематический выпуск «Электроника и нанотехнологии». 2010. №3. С. 178-181. 3. *ГОСТ 20911-75.* Техническая диагностика. Основные термины и определения. 4. *Губко М.В.* Математические модели оптимизации иерархических структур. М.: ЛЕНАНД, 2006. 264 с. 5. *Кремер Н.Ш.* Теория вероятности и математическая статистика. М.: Юнити – Дана, 2002. 343 с.

Поступила в редколлегию 04.11.2010

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.