

УДК 004.9

**П. М. Тишин**, канд. физ.-мат. наук,  
**А. С. Маковецкий**

### ПРИМЕНЕНИЕ НЕЧЁТКИХ МНОЖЕСТВ ДЛЯ ВЫЯВЛЕНИЯ НЕПОЛАДОК В РАСПРЕДЕЛЕННЫХ ИНФОРМАЦИОННЫХ СИСТЕМАХ

***Аннотация.** В работе для решения задачи выявления неполадок в распределенной информационной системе предлагается использовать отобранную совокупность диагностических параметров. Данная совокупность определяется с помощью MIB-деревьев и результатов мониторинга. По выбранному множеству параметров строится набор лингвистических переменных, который позволяет решать задачу выявления неполадок.*

***Ключевые слова:** выявление неполадок, нечеткие множества, нечеткие ситуации, лингвистические переменные, диагностические параметры, сетевые сервисы*

**P. M. Tishin**, Ph.D.,  
**A. S. Makovetskyi**

### APPLICATION OF FUZZY SETS FOR FAULT DETECTION IN A DISTRIBUTED INFORMATION SYSTEM

***Abstract.** In this paper, to solve the problem of fault detection in a distributed information system, it is proposed to use the selected set of diagnostic parameters. This is determined by MIB trees, and monitoring results. The set of linguistic variables which allows solving the problem of fault detection is based on the selected set of parameters.*

***Keywords:** fault detection, fuzzy sets, fuzzy situations, linguistic variables, diagnostic parameters, network services*

**П. М. Тишин**, канд. фіз.-мат. наук,  
**О. С. Маковецкий**

### ЗАСТОСУВАННЯ НЕЧІТКИХ МНОЖИН ДЛЯ ВИЯВЛЕННЯ НЕСПРАВНОСТЕЙ В РОЗПОДІЛЕНИХ ІНФОРМАЦІЙНИХ СИСТЕМАХ

***Анотація.** В роботі для вирішення задачі виявлення несправностей в розподіленій інформаційній системі пропонується використовувати відібрану сукупність діагностичних параметрів. Дана сукупність визначається за допомогою MIB-дерев і результатів моніторингу. По обраній множині параметрів будується набір лінгвістичних змінних, який дозволяє вирішувати задачу виявлення несправностей.*

***Ключові слова:** виявлення несправностей, нечіткі множини, нечіткі ситуації, лінгвістичні змінні, діагностичні параметри, мережеві сервіси*

**Введение.** Тенденция интенсивного роста использования облачных вычислений и online-сервисов оказывает сильное влияние на распределённые информационные системы (РИС). Стремительный рост данных систем вызывает необходимость резкого увеличения количества администраторов, которые их обслуживают. Увеличение количества серверов, администрируемых одним специалистом, позволит не только сэкономить средства на обслуживание новых мощностей, но и обеспечить высокий темп их прироста.

Такое увеличение может быть достигнуто только за счет большей автоматизации, что в свою очередь потребует средств для

более эффективного контроля ситуаций. Мониторинг ключевых показателей имеет большое значение для отслеживания поведения системы и сигнализирует об обнаружении неполадок в её работе.

На данный момент прилагаются усилия для решения проблемы выявления неполадок в поведении РИС. Решения промышленного сектора основаны на использовании пороговых методов [1]. Есть несколько исследовательских работ, предлагающих статистические методы для обнаружения неполадок [3 – 6; 8 – 9].

В настоящей работе для решения задачи выявления неполадок предлагаются методы, основанные на теории нечетких множеств.

Таким образом, практически в настоящей работе разрабатываются дополнительные методы к существующим системам об-

наружения неполадок, с помощью которых решается задача выявления наиболее значимых неполадок. Для решения задачи выявления неполадок предлагаются методы, основанные на теории нечетких множеств.

**Целью работы** является разработка новых методов диагностики РИС, позволяющих, по многократным наблюдениям диагностических параметров, производить выявление наиболее значимых неполадок.

**Основная часть**

В основу работы взята диагностическая модель корпоративной сети, предложенная в работе [7].

В соответствии с данной работой, рассмотрим службу, которая будет работать на основе web-сервера Apache HTTP Server. В качестве источника диагностической информации в данном случае выберем сетевую карту. Совокупность диагностических параметров определяется с помощью МІВ-деревьев. МІВ-деревья представляют собой иерархическую организацию соответствующих данных для функциональных устройств, в то время как для отдельной задачи необходимы только некоторые переменные. Эти переменные обычно образуют подгруппы. С точки зрения задачи диагностики, значение каждой отдельной контролируемой величины определяется параметрами,

$$H, T, O, V, \quad (1)$$

где  $H$  – имя хоста;  $T$  – время получения данных;  $O$  – ОІD или имя переменной;  $V$  – значение переменной.

В соответствии с (1), одна МІВ-переменная имеет ряд значений (каждая МІВ-переменная может использоваться для выбранного контролируемого узла). Каждая из них имеет свой собственный ряд значений во времени. Контролируемая модель может быть описана множеством всех МІВ-переменных, которые наблюдаются во времени. При обращении к переменной её значение запоминается в оперативной памяти в ассоциативных структурах. Для повышения производительности производная от переменной величина рассчитывается только в процессе доступа к переменной. Основные производные величины, которые используются в мониторинге, приведены в табл. 1.

1. Основные производные величины

И	Формула	Описание
$pV$	$pV_i = V_{i+1}$	Предыдущее значение
$V$	$V_0 \text{ startvalue}$	Последнее значение
$S$	$S_i = S_{i-1} + V_i$	Общая сумма
$sT$	$T_0 \text{ starttime}$	Значение начала мониторинга
$M$	$M_i = \max(V_i, V_{i-1})$ $M_0 = V_0$	Максимальное значение
$m$	$m_i = \min(V_i, V_{i-1})$ $m_0 = V_0$	Минимальное значение
$T$		Время опроса последней переменной
$C$	$C_i = c_{i+1} \text{ if } D_i \neq 0$	Количество значимых изменений
$N$	$N_i = N_{i-1} + 1$	Количество прочтенных/переменных
$A$	$A_i = S_i / N$	Среднее значение
$tA$	$tA_i = \frac{tA_{i-1} * TT + dt * V_i}{TT + dt}$ $TT$ – time factor	Среднее значение временных задержек
$dT$	$dT_i = T_i - T_{i-1}$	Временной период между двумя последними опросами
$D$	$D_i = V_i - V_{i-1}$	Значение, на которое изменилась переменная
$Dp$	$Dp_i = D / V_{i-1} * 100$	Значение, на которое изменилась переменная (%)
$R$	$R_i = D_i / dT_i$	Градиент

Значения величин  $sT$ ,  $M$ ,  $m$ ,  $T$ ,  $C$ ,  $N$  запоминаются в оперативной памяти, а величины  $A$ ,  $tA$ ,  $dT$ ,  $D$ ,  $Dp$ ,  $R$  вычисляются при обращении к ним.

Диагностическими параметрами будут являться производные от измеряемых параметров. Измеряемыми параметрами для сетевой карты, очевидно, будут Object

Identifier (OID) из Management Information Base (MIB).

Эксперимент, проведенный в рамках данного исследования, представляет собой снятие показаний с помощью SNMP-протокола из MIB-базы для дальнейшего их анализа.

Общая схема эксперимента показана на рис. 1.

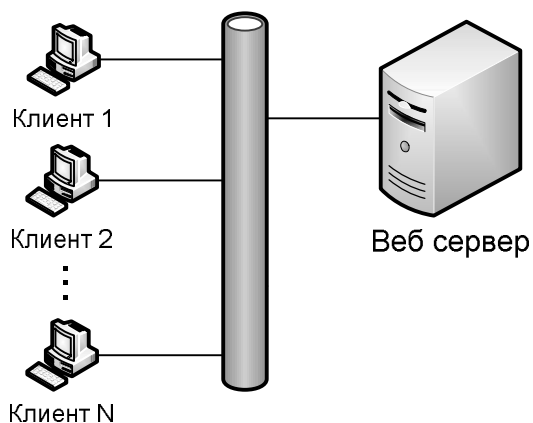


Рис. 1. Общая схема эксперимента

На множестве клиентов установлена программа Apache JMeter. Apache JMeter — инструмент для проведения нагрузочного тестирования, разрабатываемый Apache Software Foundation, и используется здесь для симуляции большой нагрузки на сервер.

Измеряемыми параметрами для сетевой карты, очевидно, будут Object Identifier (OID) из Management Information Base (MIB). Диагностическими параметрами будут являться производные от измеряемых параметров.

Из множества параметров экспертным путем были выбраны производные от MIB-переменных, которые образуют множество  $B(s)$ , введенное в работе [7], и представляют собой скорость изменения определенных OID'ов. OID'ы, отобранные в результате набора экспериментов, представлены в табл. 2.

Все множество значений  $D(b)$  некоторого выбранного параметра  $b$  разбивается на конечное число подмножеств,

$$D(b) = D_1(b) \cup \dots \cup D_{K(b)}(b), \quad (2)$$

где  $K(b)$  — общее количество подмножеств для параметра  $b$ . Причем, каждое подмно-

жество  $D_k(b)$ , включает значения параметра  $b$ , которые он принимает при одном и том же состоянии службы.

## 2. OID, отобранные в результате эксперимента

Название OID	Описание
ipInDelivers	Полное число входных дейтаграмм, успешно обработанных на IP-уровне
ipInReceives	Число полученных дейтаграмм
tcpInSegs	Полное число полученных tcp-сегментов

Для использования лингвистического описания значений параметра  $b$  построены полные ортогональные семантические пространства (ПОСП), которые служат областями лингвистических значений каждого из параметров [1].

Исходя из работы [7], можно сказать, что для построения ПОСП некоторого параметра  $b$  требуется определить множество нечетких термов,

$$D(b) = \{b^k\}_{k=1 \dots K(b)}, \quad (3)$$

где  $K(b)$  — количество нечетких термов, которые принимает соответствующая лингвистическая переменная. Нечеткие термы  $b^k \in D(b)$ , входящие в описание лингвистических переменных (3), определяются трапецеидальными функциями принадлежности  $\mu^k$ , которая положительно определена на некоторых интервалах  $(p_b^k(b), p_e^k(b))$ , зависящих от  $D_k(b)$ .

Для формализованного использования параметров системы, в соответствии с [7], определим для них лингвистические переменные (ЛП):

$$\begin{aligned} P_1 &= V_{cp}(ipInDelivers), \\ P_2 &= V_{cp}(ipInReceives), \\ P_3 &= V_{cp}(tcpInSegs). \end{aligned} \quad (4)$$

Нечеткой ситуацией называется нечеткое множество второго уровня [2],

$$\theta = \left\{ \mu(p_m^l) / p_m^l \right\}, \quad (5)$$

$$\mu(p_m) = \left\{ \frac{\mu(b_m^k)}{b_m^k} \right\}_{k=1}^{K(b_m)}, \quad (6)$$

где  $\mu(b_m^k)$  – значение функции принадлежности признака к определенному терму;  $b_m^k$  – конкретный терм лингвистической переменной  $m$ -го диагностического параметра;  $p_m$  – лингвистические переменные, определяемые формулой (4).

По результатам эксперимента были построены лингвистические переменные для описания штатных и нештатных ситуаций.

На рис. 2 изображены три термина [«Малая», «Средняя», «Большая»] лингвистической переменной  $P_1 = V_{cp}(ipInDelivers)$ , выражающие степень величины параметра ipInDelivers.

Каждый из трех термов представляется своим нечетким множеством, а нечеткое множество на множестве определения лингвистической переменной определяется функцией принадлежности этого значения данному нечеткому множеству.

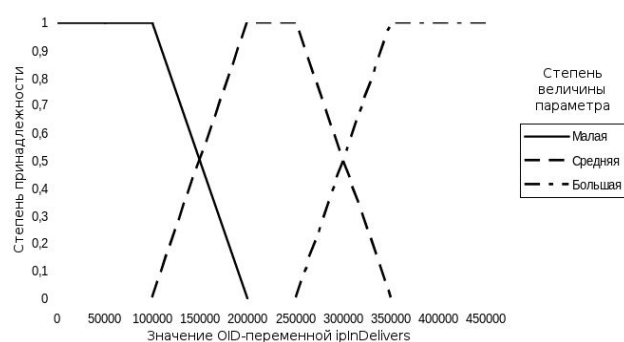


Рис. 2. Термы лингвистической переменной  $P_1$

**Выводы.** Разработанный подход позволяет дать качественное описание ситуации, при которой возникает неполадка в поведении системы.

Полученное описание штатных и нештатных ситуаций в РИС позволяет решать задачу выявления наиболее значимых непо-

ладок в поведении сервисов, работающих на основе веб-сервера Apache HTTP Server.

#### Список использованной литературы

1. Нестеренко С. А. Разработка модели онтологии диагностики сервис-ориентированных сетевых структур на основе много-сортного языка прикладной логики / С. А. Нестеренко, П. М. Тишин, А. С. Маковецкий // *Электротехнические и компьютерные системы*. – К. : Техника. – 2012. – № 07(83). – С. 102 – 108.
2. Рыжов А. П. Элементы теории нечетких множеств и ее приложений / А. П. Рыжов // *Диалог-МГУ*. – М. : 1998.
3. Bahl P., Chandra R., Greenberg A., Kandula S, Maltz D., and Zhang M., (2007), Towards Highly Reliable Enterprise Network Services via Inference of Multi-level Dependencies, *In Proc. 2007 Conf. on Applications, technologies, architectures, and protocols for computer communications, SIGCOMM*.
4. Buzen J., and Shum A., (1995), Masf – multivariate Adaptive Statistical Filtering, *In Int. CMG Conf.*
5. Cohen I., Goldszmidt M., Kelly T., Symons J., and Chase J., (2004). Correlating Instrumentation Data to System States: a Building Block for Automated Diagnosis and Control, *In Proc. 6th Conf. on Symp. on Operating Systems Design & Implementation, Vol. 6*.
6. Margineantu D., Bay S., and Chan P., (2005), Data Mining Methods for Anomaly Detection kdd-2005 Workshop Report, *Sigkdd Explorations*.
7. Viswanathan K., Lakshminarayan C, Talwar V., Wang C., Macdonald G., and Satterfield W., (2012), Ranking Anomalies in Data Centres, *In Proceedings of IEEE/IFIP Network Operations and Management Symposium, NOMS 2012, April 2012*.
8. Wang C., Talwar V., Schwan K., and Ranganathan P., (2010), Online Detection of Utility Cloud Anomalies using Metric Distributions, *In IEEE/IFIP Network Operations and Management Symposium*.
9. Wang C., Viswanathan K., Lakshminarayan C., Talwar V., Satterfield W., and Schwann K., (2011), Statistical Techniques for

Online Anomaly Detection in Data Centers, *In Integrated Network Management*.

Получено 15.03.2015

### References

1. Nesterenko S.A., Tishin P.M., Makovetskiy A.S. Razrabotka modeli ontologii diagnostiki servis-orientirovannykh setevykh struktur na osnove mnogosortnogo yazyka prikladnoy logiki [Development of Ontology Diagnostic Model of Service-oriented Network Structures Based on the Applied Polysort Logic Language], (2012), *Elektrotekhnicheskie i Komp'yuternye Sistemy*, Kiev, Ukraine, No. 07 (83), pp. 102 – 108 [In Russian], url: <http://www.etks.opu.ua/core/getfile.php?id=218>.

2. Ryzhov A.P. Elementy teorii nechetkikh mnozhestv i ee prilozheniy [Elements of the theory of Fuzzy Sets and its Applications], (1998), *Dialog-MGU*, Moscow, Russian Federation [In Russian], url: <http://www.intsys.msu.ru/staff/ryzhov/FuzzySetsTheory&Applications.pdf>.

3. Bahl P., Chandra R., Greenberg A., Kandula S., Maltz D., and Zhang M., (2007), Towards Highly Reliable Enterprise Network Services via Inference of Multi-level Dependencies, *In Proc. 2007 Conf. on Applications, Technologies, Architectures, and Protocols for Computer Communications, SIGCOMM* (In English), url: [http://research.microsoft.com/pubs/73112/sherlock\\_sigcomm\\_07.pdf](http://research.microsoft.com/pubs/73112/sherlock_sigcomm_07.pdf).

4. Buzen J., and Shum A., (1995), Masf – multivariate Adaptive Statistical Filtering, *In Int. CMG Conf.* (In English), url: [http://www.researchgate.net/publication/221446750\\_MASF\\_Multivariate\\_Adaptive\\_Statistical\\_Filtering](http://www.researchgate.net/publication/221446750_MASF_Multivariate_Adaptive_Statistical_Filtering).

5. Cohen I., Goldszmidt M., Kelly T., Symons J., and Chase J., (2004), Correlating Instrumentation Data to System States: a Building Block for Automated Diagnosis and Control, *In Proc. 6th Conf. on Symp. on Operating Systems Design & Implementation*, Vol. 6 (In English), url: [https://www.usenix.org/legacy/event/osdi04/tech/full\\_papers/cohen/cohen.pdf](https://www.usenix.org/legacy/event/osdi04/tech/full_papers/cohen/cohen.pdf).

6. Margineantu D., Bay S., and Chan P., (2005), Data Mining Methods for Anomaly De-

tection kdd-2005 Workshop Report, *Sigkdd Explorations* (In English), url: <http://www.kdd.org/sites/default/files/issues/7-2-2005-12/DMMAD-KDD2005-report.pdf>.

7. Viswanathan K., Lakshminarayan C., Talwar V., Wang C., Macdonald G., and Satterfield W., (2012), Ranking Anomalies in Data Centers, *In Proceedings of IEEE/IFIP Network Operations and Management Symposium, NOMS 2012*, April 2012 (In English), url: [http://www.hpl.hp.com/people/vanish\\_talwar/2012\\_NOMS\\_RankingAnomalies-CR.pdf](http://www.hpl.hp.com/people/vanish_talwar/2012_NOMS_RankingAnomalies-CR.pdf).

8. Wang C., Talwar V., Schwan K., and Ranganathan P., (2010). Online Detection of Utility Cloud Anomalies Using Metric Distributions, *In IEEE/IFIP Network Operations and Management Symposium* (In English), url: [http://www.hpl.hp.com/people/vanish\\_talwar/2010\\_NOMS\\_AnomalyDetection.pdf](http://www.hpl.hp.com/people/vanish_talwar/2010_NOMS_AnomalyDetection.pdf).

9. Wang C., Viswanathan K., Lakshminarayan C., Talwar V., Satterfield W., and Schwan K., (2011), Statistical Techniques for Online Anomaly Detection in Data Centers, *In Integrated Network Management* (In English), url: <http://www.hpl.hp.com/techreports/2011/HPL-2011-8.pdf>.



Тишин  
Петр Металлинович,  
канд. физ.-мат. наук,  
доц. каф. Одесского  
нац. политехнич. ун-та,  
м/т.: +38(098)805-04-  
48.  
E-mail: tik88@mail.ru



Маковецкий  
Александр Сергеевич,  
ст. преподаватель каф.  
Одесского нац.  
политехнического ун-та,  
м/т.: +38(066)356-09-  
87.  
E-mail:  
drnewman@mail.ru