REVIEW PAPER
International Journal of Recent Trends in Engineering, Vol 2, No. 4, November 2009

A Study on Long Term Behavior of Web Applications using Fuzzy Markov Model

R. Sujatha[1], B. Praba[2], and T.M. Rajalaxmi[3]

[1 2 3] Department of Mathematics, SSN College of Engineering, Chennai, India.
Email: [1 2]brprasuja@yahoo.co.in
Email: [3]laxmi.raji18@gmail.com

***Abstract*—In this paper we propose a Ping – Pong transitions on web applications. The numerical values of parameters in real time situations are not exactly predictable; to overcome this we are using a fuzzy criterion through possibilities which are represented as a triangular fuzzy number. We introduce a method to study the long term behavior (steady state) of Ping – Pong transitions using Fuzzy Markov Model. And we compare the steady state behavior of original transitions with Ping – Pong transitions by measuring the distance between two steady – state vectors.**

***Index Terms*— Fuzzy Markov Model, Possibility space, steady – state vector, Ping – Pong Transitions, Aggregate state, Distance Metric.**

I. INTRODUCTION

A web application is a self-contained sub tree of the web site. Web applications are inherently distributed and they require cooperation between a server and a client in order to accomplish their tasks and it offer cross platform universal access to web resources for immense user population. The goal of Web applications is to be accessible regardless of the platform a client is executing on. Because of the user focus and the large size of the web quality assertion for the web becomes progressively more significant. A mixture of analysis and quality assurance is already being performed on web applications. But the web environment presents many new challenges and requires new techniques. Web – technology is just opening completely new areas of application of fuzzy sets. A web application [7] is a web site where user input – navigation through the site and data entry affects the state of the business beyond, of course, access logs and hit counters. In essence, a web application uses web sites as the front end to a business application.

Henderson, Kotz and Abyzov [10] are interested in user mobility; i.e., how often, and how far, a user moves during a session. To study user's mobility they defined Ping - Pong transitions. Our study shows that there exist large personal differences in user's navigation as well as in their data transfer rates. In this paper we define Ping - Pong transitions for web applications i.e. users spend a large fraction of their time and long periods of time at some location (web page), which we call Ping - Pong transitions. Using these transitions we analyze the long term behavior of web applications.

Most of our conventional tools for proper modeling and computing are crisp, deterministic and exact in nature. Exactness assumes that parameters of a model represent precisely the features of the real system that has been modeled. Now as the complexity of a system increases our ability to make precise and yet significant statements about its behavior diminishes until a threshold is reached beyond which precision and significance become almost mutually exclusive characteristics.

One of the meanings attributed to the term 'uncertainty' is "vagueness". That is, the difficulty of making sharp or precise distinction. A mathematical frame work to describe these phenomena was suggested by Lotfi. A. Zadeh in his seminal paper entitled "Fuzzy sets". The underlying fuzzy principle is that 'Everything is a matter of degree'. Thus, the membership in a fuzzy set is not a matter of assertion or rejection but rather a matter of degree.

Real situations are very often uncertain or vague in a number of ways. Due to lack of information, the future state of the system might not be known completely for overcoming this problem we are using *fuzzy Markov model*. In view of the fact that web pages and link between the web pages have inexact data, for capturing the uncertainty we model a web application as a fuzzy Markov model (FMM) with set of states and transition possibility between the states.

This paper consists of five sections. In section two we discuss the preliminaries, in section three we describe Ping - Pong transitions, we express the illustration in section four and in section five we discuss the Distance metric. Finally we end up with some conclusions.

II. PRELIMINARIES

In this section we shall discuss some basic definitions.

A. Fuzzy Set

The concept of fuzzy sets is a generalization of the crisp sets.

A *fuzzy set* can be defined mathematically by assigning to each possible individual in the universe of discourse a value representing its grade of membership in the fuzzy set. This grade corresponds to the degree to which that individual is similar or compatible with the concept represented by the fuzzy set.

Formally let $\tilde{A}$ be a fuzzy set, then the *membership function* [5] $\mu_{\tilde{A}}(x) \in [0,1]$ is evaluated for $\tilde{A}$ at $x \in R$, where $[0,1]$ denotes the interval of real numbers from 0 to 1, including 0 and 1. Thus the fuzzy sets are the subsets of the real number system.

B. $\alpha - cut$

An $\alpha - cut$ [8] of A is denoted by $\tilde{A}_\alpha$ and is defined as $\{x / \mu_{\tilde{A}}(x) \geq \alpha, x \in X\}$.

C. Triangular Fuzzy Number

Triangular fuzzy number [8] is a fuzzy number represented by three points as follows:

$$\tilde{A} = (a_1, a_2, a_3) \text{ where } (a_1 \leq a_2 \leq a_3)$$

and its membership function is given as follows:

$$\mu_{\tilde{A}}(x) = \begin{cases} 0, & x < a_1 \\ \dfrac{x - a_1}{a_2 - a_1}, & a_1 \leq x \leq a_2 \\ \dfrac{a_3 - x}{a_3 - a_2}, & a_2 \leq x \leq a_3 \\ 0, & x > a_3 \end{cases}$$

D. $\alpha - cut$ of Triangular *Fuzzy Number*

An $\alpha - cut$ [8] of the triangular fuzzy number is represented as closed and bounded interval of real numbers.

$$\tilde{A}_\alpha = [(a_2 - a_1)\alpha + a_1, -(a_3 - a_2)\alpha + a_3] = [\tilde{A}_L(\alpha), \tilde{A}_R(\alpha)]$$

which satisfies the following conditions.

(i) A_L is increasing on $[0,1]$,

(ii) A_U is decreasing on $[0,1]$,

(iii) $A_L(1) \leq A_U(1)$

E. Possibility measure

Let Γ be the universe of discourse and ψ be the power set of Γ, Then the possibility measure [6] σ is a mapping $\sigma : \psi \to [0,1]$ such that

(i) $\sigma(\phi) = 0, \sigma(\Gamma) = 1$

(ii) $\sigma(\bigcup_i A_i) = \sup_i(\sigma(A_i))$

for every arbitrary collection A_i of ψ. The triplet (Γ, ψ, σ) is called as *Possibility Space.*

F. Possibilistic variable

A possibilistic variable [6] X is a mapping from Γ to U i.e.

$$X : \Gamma \to U .$$

If U is countable, then X is called *discrete* possibilistic variable otherwise it is called *continuous* possibilistic variable.

G. Fuzzy Markov Model

A fuzzy Markov model (FMM) [9] is the Model which has a finite number of states $S_1, S_2, ..., S_r$ at each transition $n = 1,2,..., l$ together with the fuzzy possibilities $\tilde{p}_{ij}, i, j \in \{S_1, S_2, ..., S_r\}$ where S_r the total number of states in the FMM is and $\tilde{p}_{ij}$ is the *transition fuzzy possibility* from the state S_i to S_j. In notation,

$$\tilde{p}_{ij} = \sigma(X_{n+1} = S_j / X_n = S_i)$$

i.e., state S_j at step $n+1$*th* transition given the state S_i at n *th* transition, where $S_1 \leq i, j \leq S_r$ $n = 1,2,..., l$ These $\tilde{p}_{ij}$ are the *transition fuzzy possibilities* which do not depend on n.

H. Transition Fuzzy Possibility Matrix

The transition fuzzy possibility matrix $\tilde{\mathbf{P}} = (\tilde{p}_{ij})$ is an $S_r \times S_r$ matrix of the transition fuzzy possibilities and each $(\tilde{p}_{ij}) \geq 0$. In this paper we are using elements of $\tilde{\mathbf{P}}$ as a special type of fuzzy number for capturing the fuzziness called the triangle fuzzy number.

` (i) The possibility of being in state S_j at ($n+1$) *th* transition is given by $\tilde{p}_j^{(n+1)} = \sigma(X_{n+1} = S_j)$

(ii) $\tilde{p}_{ij}^{\,n}$ be the possibility of starting off in state S_i and ending up in S_j after n steps.

Define $\tilde{\mathbf{P}}^n$ to be the max-min product of $\tilde{\mathbf{P}} - n$ times and it is well known that $\mathbf{P}^n = (\tilde{p}_{ij}^{\,n}) \forall n.$ If $\tilde{p}^{(0)} = (\tilde{p}_1^{(0)}, ..., \tilde{p}_r^{(0)})$, where $p_i^{(0)}$ be the possibility of initially being in state S_i, and $\tilde{p}^{(n)} = (\tilde{p}_1^{(n)}, ..., \tilde{p}_r^{(n)})$ where $\tilde{p}_i^{(n)}$ be the possibility of being in state S_i after n steps, we know that $\tilde{p}^{(n)} = \tilde{p}^{(0)} \otimes \tilde{\mathbf{P}}^n = \tilde{p}^{(n-1)} \otimes \tilde{\mathbf{P}}$, As $n \to \infty$, $\tilde{p}^{(n)}$ is called the steady state vector denoted by $\tilde{\Pi}$ which can be obtained by $\tilde{\Pi} \otimes \tilde{\mathbf{P}} = \tilde{\Pi}$ where the elements of $\tilde{\Pi}$ such that $\tilde{\pi}_j$ is the possibility of remaining in state j after the long run.

For obtaining the *steady state vector*, the idea is that, by simulating the underlying Markov chain for a sufficiently long time until it converges and we obtain a approximation of the steady-state probability distribution.

I. Process of finding the Steady state vector

Consider the transition fuzzy possibility matrix $\tilde{\mathbf{P}}$ and let $\tilde{\Pi}$ be a fuzzy set of the states S_i If the *Max - Min* composition of $\tilde{\mathbf{P}}$ and $\tilde{\Pi}$ gives $\tilde{\mathbf{P}}'$ a fuzzy set of S_i and when $\tilde{\mathbf{P}}'$ equals $\tilde{\Pi}$, then we say that $\tilde{\Pi}$ is an eigen [3] fuzzy set associated with the given transition matrix $\tilde{\mathbf{P}}$. There are so many methods for finding steady state vector [3]. Here we followed the below method.

(i). Find out the greatest elements in each column of the transition fuzzy possibility matrix $\tilde{\mathbf{P}}$. Then we get one row vector that is equal to say $\tilde{\Pi}$ then encircle the smallest element in that row vector by using the comparison of triangular fuzzy number [2] say that is equal to $\tilde{\pi}_0$.

(ii). Delete the columns containing that smallest of the greatest elements, and that of same rows from $\tilde{\mathbf{P}}$. Now we have get $\tilde{\mathbf{P}}'$ the first reduction of $\tilde{\mathbf{P}}$. It is important to note that we don't delete the rows passing through the positions of the value $\tilde{\pi}_0$.

(iii). Set in $\tilde{\Pi}$ the value of $\tilde{\pi}_0$, in the position of the deleted columns.

(iv). Return to the first step with $\tilde{\mathbf{P}}'$ instead of $\tilde{\mathbf{P}}$, but with the following restriction: let π_1 denoting the smallest of the greatest elements in each column of $\tilde{\mathbf{P}}'$, set in the *appropriate position* of $\tilde{\Pi}$,

$$\begin{cases} \tilde{\pi}_1, \tilde{\pi}_1 \geq \tilde{\pi}_0; \\ \tilde{\pi}_0, \tilde{\pi}_1 < \tilde{\pi}_0. \end{cases}$$

(*v*). Return to the first step until all reductions are exhausted.

This $\tilde{\Pi} = \{\tilde{\pi}_i\}$ is called as the steady state vector. This steady state vector is independent of the initial vector. Thus the possibilities are independent of the choice of time origin, and the process is stationary.

III. Ping-Pong Phenomena

By characterizing the web page users from the Association patterns *AP* with the help of web log access data, we found that some times users experience frequent re-associations between two or three web pages in a short period of time. This phenomenon is known as a *Ping - Pong* transition in the web applications.

In this section, we measure how often the ping-pong phenomenon occurs. We introduce a method to analyze the steady state behavior of ping-pong. We compare the steady state behavior of original transitions with the same of ping-pong transitions.

A. Ping – Pong Transitions

Ping - Pong transitions occur when a user is in nature located among some particular neighboring web page.

B. Example

(*i*) For web page (state) S_i and S_j, if at any time a user makes a sequence of transition $S_i \to S_j \to S_i \to S_j$, then these transitions are *Ping - Pong* transitions.

(*ii*) Also for web page S_i, S_j and S_k, if at any time a user makes a sequence of transitions $S_i \to S_j \to S_k \to S_i$, then these transitions are Ping - Pong transitions.

Once a Ping - Pong transition is identified in the *AP* of a user, we cluster the states incurred in the Ping -Pong transition into an *aggregate state AS*.

For example, S_i and S_j form an *AS* in case (*i*). $AS = \{S_i, S_j\}$ whereas S_i, S_j and S_k form an aggregate state in case (*ii*). $AS = \{S_i, S_j, S_k\}$.

C. Dominant state

Dominant state is a state with which the user has been associated most of the time among the *AP*s in the aggregate state *AS*.

D. Example

If a sequence of *AP* transitions $S_k \to S_i \to S_j \to S_i \to S_j \to S_l$ appears in the association patterns of a user, then the Ping - Pong transition $S_i \to S_j \to S_i \to S_j$ is identified with the aggregate state $AS = \{S_i, S_j\}$. If *AP* S_i is the dominant state in the $AS = \{S_i, S_j\}$, then the original sequence of AP transitions is replaced by the sequence of $S_k \to S_i \to S_l$.

Because of the frequent occurrence of ping pong transitions, the transition fuzzy possibility matrix at each *AP* derived under the original *AP* may not represent the exact navigation behavior of users for that we investigate the steady state analysis of original transition with the elimination of ping pong transition from the original transition.

IV. Illustration

We consider the website www.ssn.edu.in as FMM with state space $\{IT, ECE, CSE, EEE, BME\}$. From its log files we extract data to get the transitions between them. And its transition diagram without the ping pong transitions is given below

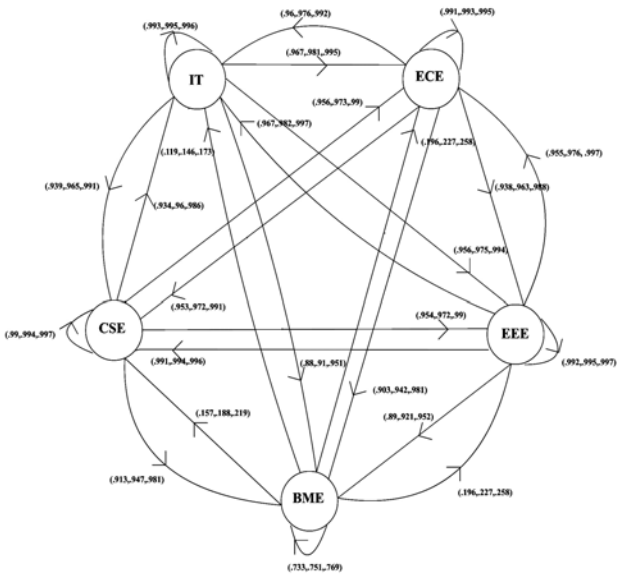

Figure 1. Transition diagram of without elimination of Ping – Pong Transitions

The transition possibility values between the states are calculated using α - *cut* method [8]. Its transition fuzzy possibility matrix is $\tilde{\mathbf{P}}_{\mathbf{O}} = (\tilde{p}_{ij}) = [C_1, C_2, C_3, C_4, C_5]$ where the columns C_1, C_2, C_3, C_4, C_5 are,

$$C_1 = \begin{pmatrix} (0.993, 0.995, 0.996) \\ (0.960, 0.976, 0.992) \\ (0.934, 0.960, 0.986) \\ (0.967, 0.982, 0.997) \\ (0.119, 0.146, 0.173) \end{pmatrix} \quad C_2 = \begin{pmatrix} (0.967, 0.981, 0.995) \\ (0.991, 0.993, 0.995) \\ (0.956, 0.973, 0.990) \\ (0.955, 0.976, 0.997) \\ (0.196, 0.227, 0.258) \end{pmatrix}$$

$$C_3 = \begin{pmatrix} (0.939, 0.965, 0.991) \\ (0.953, 0.972, 0.991) \\ (0.990, 0.994, 0.997) \\ (0.991, 0.994, 0.996) \\ (0.157, 0.188, 0.219) \end{pmatrix} \quad C_4 = \begin{pmatrix} (0.956, 0.975, 0.994) \\ (0.938, 0.963, 0.988) \\ (0.954, 0.972, 0.990) \\ (0.992, 0.995, 0.997) \\ (0.196, 0.227, 0.258) \end{pmatrix}$$

$$C_5 = \begin{pmatrix} (0.880, 0.910, 0.951) \\ (0.903, 0.942, 0.981) \\ (0.913, 0.947, 0.981) \\ (0.890, 0.921, 0.952) \\ (0.733, 0.751, 0.769) \end{pmatrix}$$

With the help of the above mentioned way for finding the steady state vector, we find the steady state vector for original path as ,

$$\tilde{\Pi}'_O = \begin{pmatrix} (0.967, 0.982, 0.997) \\ (0.955, 0.976, 0.997) \\ (0.990, 0.994, 0.997) \\ (0.992, 0.995, 0.997) \\ (0.913, 0.947, 0.981) \end{pmatrix}$$

where $\tilde{\Pi}'_O$ is the transpose of the row vector. During the extraction of data, we discover that the states namely {*IT, ECE*}, {*CSE, EEE*} are Aggregate states in Ping – Pong transitions i.e., $AS = \{IT, ECE\}$ and another one is $AS = \{CSE, EEE\}$. The FMM for the Ping – Pong transition is depicted in " Fig. 2".

After we aggregate the states, the transition fuzzy possibility matrix $\tilde{\mathbf{P}}_{\mathbf{P}} = (\tilde{p}_{ij})$ for Ping - Pong transitions is

$$\tilde{\mathbf{P}}_{\mathbf{P}} = \begin{pmatrix} (0.993, 0.996, 0.998) & (0.962, 0.973, 0.984) & (0.910, 0.952, 0.994) \\ (0.967, 0.977, 0.986) & (0.994, 0.997, 0.999) & (0.875, 0.911, 0.947) \\ (0.190, 0.213, 0.236) & (0.220, 0.244, 0.268) & (0.610, 0.630, 0.650) \end{pmatrix}$$

and its steady state is calculated as

$$\tilde{\Pi}_P = ((0.993, 0.996, 0.998)(0.994, 0.997, 0.999)(0.910, 0.952, 0.994))$$

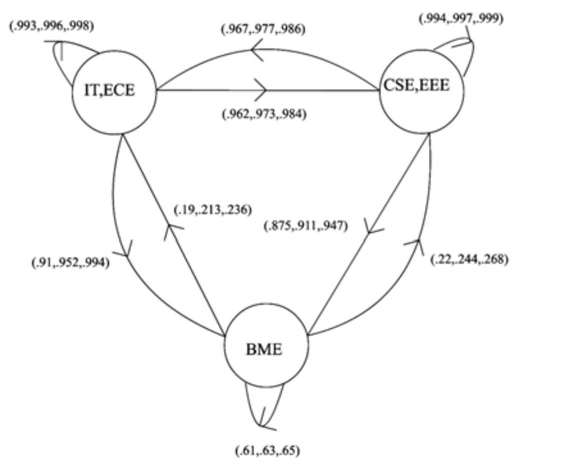

Figure 2. Transition diagram of elimination of Ping – Pong method

The Steady – state by eliminating the Ping – Pong transition is given below

$$\tilde{\Pi}_P^T = \begin{pmatrix} (0.993, 0.996, 0.998) \\ (0.993, 0.996, 0.998) \\ (0.994, 0.997, 0.999) \\ (0.994, 0.997, 0.999) \\ (0.910, 0.952, 0.994) \end{pmatrix}$$

where *T* is the transpose of row vector. The steady – state possibility of each state in an *AS* is taken as the corresponding steady – state possibility of members of that *AS*, i.e., steady – state possibility of $AS = \{IT, ECE\}$ is same as steady – state possibility of *IT* as well as ECE.

V. Distance between two steady state vectors

The concept 'distance' is designated to describe the difference. We know how to find the distance between two real numbers x, y. The distance is $|x - y| = \mathbf{d}(x, y)$. We also know how to find the distance between two points in $\mathbf{R}^2$. Now the metric **D** for the fuzzy sets $\tilde{A}, \tilde{B}$ is defined as follows.

A. Distance Measure

Let $\tilde{A}, \tilde{B}$ be the fuzzy sets and their corresponding $\alpha - cut$ [5] are $\left[\tilde{A}_L(\alpha), \tilde{A}_R(\alpha) \right]$ $\left[\tilde{B}_L(\alpha), \tilde{B}_R(\alpha) \right] 0 \le \alpha \le 1$. Define

$$L_\alpha = \left| \tilde{A}_L(\alpha) - \tilde{B}_L(\alpha) \right| \;\; \& \;\; R_\alpha = \left| \tilde{A}_R(\alpha) - \tilde{B}_R(\alpha) \right|$$

L_α is the absolute value of the difference between the lower bound and R_α is the absolute value of the difference between the upper bound, then

$$\mathbf{D}(\tilde{A}, \tilde{B}) = \max \{ \max (L_\alpha, R_\alpha) \} / 0 \le \alpha \le 1$$

This **D** is a metric.

Hence we find the distance between the two steady state vectors of the FMM corresponding to the example given in Sec.IV as follows.

The first element of the steady state vector of the original transition is $(0.967, 0.982, 0.997)$. An $\alpha - cut$ of this triangular fuzzy number is,

$$\tilde{A}_\alpha^{\,1} = [(0.015)\alpha + 0.967, -(0.015)\alpha + 0.997]$$

After eliminating the Ping – Pong transitions the first element of the steady state vector is $(0.993, 0.996, 0.998)$. And $\alpha - cut$ of this triangular fuzzy number is,

$$\tilde{B}_\alpha^{\,1} = [(0.003)\alpha + 0.993, -(0.002)\alpha + 0.998]$$

the absolute value of lower bound $L_\alpha^{\,1}$ and the upper bound $R_\alpha^{\,1}$ is given below

$$L_\alpha{}^1 = |(0.012)\alpha + 0.026| \ \& \ R_\alpha{}^1 = |(0.013)\alpha + 0.001|$$

Thus the distance between them is given by

$$\mathbf{D}\left(L_\alpha{}^1, R_\alpha{}^1\right) = \max\{\max[(0.012)\alpha + 0.026, (0.013)\alpha + 0.001]\}$$
$$= \max\{(0.012)\alpha + 0.026\} = 0.012 + 0.026$$
$$= 0.038$$

Hence the distance between the first entries of the two steady state vectors is $\mathbf{D}(L_\alpha{}^1, R_\alpha{}^1) = 0.038$. The remaining entries can be calculated in a similar manner. And the Distance vector is,

$$\mathbf{D} = \begin{bmatrix} 0.038 \\ 0.056 \\ 0.005 \\ 0.002 \\ 0.093 \end{bmatrix}$$

The distance between the steady state vectors of the state *IT* is depicted below.

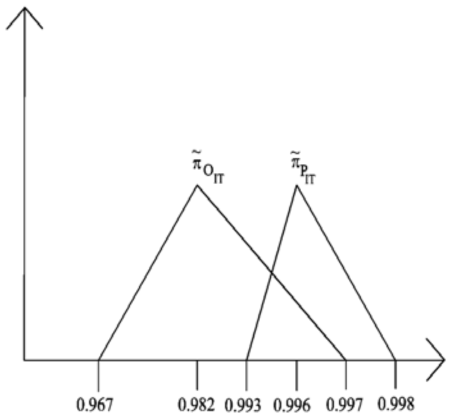

Figure 3. Distance between the steady state possibility for IT

CONCLUSTION

By comparing the steady state vector of original transitions with the Ping – Pong transitions, we conclude that the elimination of Ping - Pong transition in the sequence of transitions will make no effect and are very closer to original transitions. By using the Ping – Pong transitions the state spaces are reduced and calculation is easy for large systems. Hence for analyzing the long term behavior of large systems we can use Ping – Pong transitions.

ACKNOWLEDGMENT

We thank the management of SSN Institutions and Principal for providing necessary facilities to carryout this work. This work is a part of SSNCE funded project “Applications of fuzzy techniques in Web Testing”

REFERENCES

[1] Chris Cornelis, Peter De Kesel, Etienne E. Kerre, “Shortest Paths in Fuzzy Weighted Graphs”, *International journal of intelligent systems*, 19, 1051-1068, 2004.
[2] Dorohonceanu.D and Marin.D, “A Simple Method for Comparing Fuzzy Numbers”, 2002
[3] Elie Sanchez, “Eigen fuzzy sets and fuzzy Relations”, *Journal of Mathematical analysis and applications*, 81 (1981), 399 -421.
[4] James J. Buckley, “Fuzzy Markov Chains: Uncertain Probabilities”, *Mathware and Soft Computing*, Vol 9, No. 1 (2002).
[5] James J. Buckley and Esfandiar Eslami, *An introduction to fuzzy logic and fuzzy sets*, Springer, 2007.
[6] Kai-Yuan Cai, *Introduction to fuzzy reliability*, Kluwer Academic Publishers, USA, 1996.
[7] S. Jablonski, L. Petrov, C. Meiler, and U. Mayer, “Guide to web application and platform Architecture”, Springer, 2004
[8] Kwang H. Lee, *First Course on Fuzzy Theory and Applications*, Springer 2005.
[9] Sujatha.R. and Praba.B., “A fuzzy Markov model for web testing”, *Ganita Sandesh*, 21 (2007), 111 - 120.
[10] Tristan Henderson, David Kotz and Ilya Abyzovm, “The changing usage of a mature campus – wide wireless network”, *In Proceedings of ACM MobiCom’04*, Philadelphia, PA, 2004.