International Journal of Modern Engineering Research (IJMER)
www.ijmer.com Vol.2, Issue.2, Mar-Apr 2012 pp-512-517 ISSN: 2249-6645

ISO-Failure in Web Browsing Using Markov Chain Model and Curve Fitting Analysis

Diwakar Shukla[1], Kapil Verma[2] and Sharad Gangele [3]
[1](Department of Mathematics and Statistics, Sagar University, Sagar M. P. 470003) India.
Associate Member, DST-Centre for Interdisciplinary Mathematical Sciences, Banaras Hindu University, Varanasi, U.P.
[2](Department of Computer Science, M.P. Bhoj (Open) University Kolar Road Bhopal, M.P) India.
B.T. Institute of Research and Technology, Seronja, Sagar, M.P.
[3](Department of Computer Science, M.P. Bhoj (Open) University, Kolar Road, Bhopal, M. P.) India.

ABSTRACT

There are many browsers available for internet surfing and user has options to choose anyone providing the maximum facility. Every browser has popularity and limitations. The popularity is related to number of users who prefer a browser over others. Similarly, the limitation relates to excess delay in searching websites, which is browser failure. The browser sharing is a phenomenon which relates to the number of users attracted for using a particular browser besides its limitations. This paper presents an analysis of browser sharing when limitations of two browsers are rated as equal value. The term iso-failure appears for equality of browsers failure probability. An attempt has been made to obtain the approximate linear relationship between find browser sharing and initial share. A straight line is fitted on the generated data using the method of least square. The coefficient of determination and confidence interval support the theoretical facts.

***Keywords* –** Transition Probability Matrix (TPM), Markov Chain Model (MCM), Coefficient of Determination (COD), Confidence Interval, Iso-Failure.

I. INTRODUCTION

A browser is used for accessing the different sites of internet. If a user takes into application only two browsers then he may has a favor for anyone depending upon the market popularity or the failure probability. If b_1 and b_2 are the two probabilities related to browser failure then both are suppose to be different. The notation iso-failure means the equality of failure level probability for both the browsers. If two browsers are in competitions then it is a matter of interest to know which one is having better share than others. This paper introduces the concept of iso-failure in the web browsing phenomenon. A stochastic model is proposed and analysis is performed. The basic model is proposed by Naldi (2002) and extended by Shukla and Gadewar (2007). Despande and Karpis (2004) discussed the Markov model application over web page access computing the prediction . A similar useful contribution is due to Pirolli (1996). A Markov chain model helps to establish interrelationship between process variables. Shukla and Jain (2007) have suggested stochastic model for multilevel queue scheduling and further extended by Shukla *et al.* (2010). In a useful contribution Shukla and Singhai (2011) discussed analysis of user web browsing behavior with the help of Markov chain model. This paper extends the same in light of iso-failure probability and establishes the linear relationship.

2. A REVIEW

Medhi (1991, 1992) presented a detailed discussion on the stochastic process and their applications. Chen and Mark (1993) discussed the fast packet switch shared concentration and output queueing for a busy channel. Humbali and Ramani (2002) evaluated multicast switch with a variety of traffic patterns. Newby and Dagg (2002) have a useful contribution on the optical inspection and maintenance for stochastically deteriorating system. Dorea *et al.* (2004) used Markov chain for the modeling of a system and derived some useful approximations. Yeian and Lygeres (2005) presented a work on stabilization of class of stochastic different equations with Markovian switching. Shukla *et al.* (2007 a) advocated for model based study for space division switches in computer network. Francini and Chiussi (2002) discussed some interesting features for QoS guarantees to the unicast and multicast flow in multistage packet switch. On the reliability analysis of network a useful contribution is by Agarwal and Lakhwinder (2008) whereas Paxson (2004) introduced some of their critical experiences while measuring the internet traffic. Shukla *et al.* (2009 a, b & c) presented different dimensions of internet traffic sharing in the light of share loss analysis. Shukla *et al*. (2010 a, b. c & d) have given some Markov chain model applications in view to disconnectivity factor, multi marketing and crime based analysis. Shukla *et al*. (2011 a, b, c, d, e & f) discussed the elasticity property and its impact on parameters of internet traffic sharing in presence blocking probability of computer network specially when two operators are in business competitions with each other in a market.

3. OBJECTIVES:

The contents of this paper is for

1. To examine the browser sharing when there exist iso failure of browsers.
2. To establish linear relationship using procedure of least square.

4. MODEL:

Let $\{X_n , n \geq 0\}$ be a Markov chain on state space $\{C, Q, B_1, B_2, S\}$ asper Shukla and Singhai (2011) where

State C: represents connecting state.

State Q: user quitting from the process

State B_1: user attempts to surf through browser B_1.

State B_2: user attempts to surf through browser B_2.

State S: success for connectivity and surfing.

The $X^{(n)}$ denotes the position of random variable X in the state space at the n^{th} browser connectivity attempt made by a web user.

5. ASSUMPTIONS FOR USER BROWSING BEHAVIOR [see Shukla and Singhai (2011)]

(1) The user attempts for dial up connection to use Internet. If the connection is not established, user quits with the probability Pc.

(2) When connection is made user chooses any one of browsers B_1, B_2 with the probability p and 1–p respectively.

(3) User navigates to any one browser at a time when successfully opened.

(4) Browser B_i *(i=1, 2)* failure occurs due to non-opening of any site through browser B_i .Then user either quits (with probability p_q) or switches to the next browser.

(5) Switching between browsers is on attempt by attempt basis *(n=1, 2, 3….)*.

(6) Initial preference for a browser is based on quality of services and variety of facility features are contained in both browsers.

(7) Failure probability of a browser B_1 is b_1 and of B_2 is b_2.

(8) Transition probability of surfing through B_1, being completed in a single attempt is $(1 – b_1)$.

(9) Absorbing state (transition from a state to itself) probability is 1. No further transition from this state occurs.

Under these assumptions user's browsing behavior has a Markov Chain Model (see fig.5.1) in which the transition probabilities are on the arcs connecting the circles and representing the chain states.

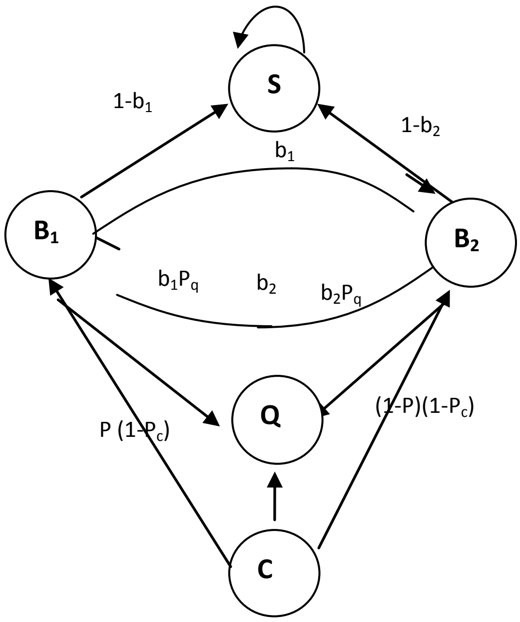

Fig.5.1 [Transition diagram of user browsing]

The initial conditions n=0, (state probability before the first surf attempt) are:

$$\left.\begin{aligned} P\left[X^{(0)}=C\right] &= 1\\ P\left[X^{(0)}=B_1\right] &= 0\\ P\left[X^{(0)}=B_2\right] &= 0\\ P\left[X^{(0)}=S\right] &= 0\\ P\left[X^{(0)}=Q\right] &= 0 \end{aligned}\right\} \quad (5.1)$$

The unit-step transition probability matrix is:

	B_1	B_2	S	Q	C
B_1	0	$b_1(1-b_1)$	$(1-b_1)$	b_1P_q	0
B_2	$b_2(1-b_2)$	0	$(1-b_2)$	b_2P_q	0
S	0	0	1	0	0
Q	0	0	0	1	0
C	$P(1-P_c)$	$(1-P)(1-P_c)$	0	P_c	0

Table: 5.1 [Transition Probability Matrix]

By using Shukla and Singhai (2011we write

$$P\left[x^{(2n)}=B_1\right]=b_1^{\,n-1}b_2^{\,n}(1-p)(1-pc)(1-pq)^{2n-1};n>0...(5.1)$$

$$P\left[X^{(2n+1)}=B_1\right]=\left(b_1b_2\right)^nP\left(1-P_C\right)\left(1-P_q\right)^{2n};n>0...(5.2)$$

Similarly, for browser B_2

$$P\left[X^{(2n)}=B_2\right]=b_1^nb_2^{n-1}P\left(1-P_C\right)\left(1-P_q\right)^{(2n-1)};n>0...(5.3)$$

When n is odd

$$P\left[X^{(2n+1)} = B_2\right] = \left(b_1 b_2\right)^n (1-P)\left(1-P_C\right)\left(1-P_q\right)^{2n};n>0...(5.4)$$

6. BROWSER SHARING: As per Shukla and Singhai (2011) browser sharing by B_1 is

$$P_1 = \lim_{n\to\infty} \overline{P}_1^{(2n)} = \left(1-b_1\right)\left(1-P_C\right)\left[\frac{P+(1-P)(1-P_q)b_2}{1-b_1 b_2 (1-P_q)^2}\right]...(6.1)$$

7. ISO-FAILURE: The failure probability of first browser B_1 is b_1 and second browser B_2 is b_2.

Let us define $b_1=b_2=b$ then this condition constitute the iso-failure browser probability and iso-failure curves.

Now the (6.1) expression will be

$$P_1 = (1-b)(1-pc)\left[\frac{p+(1-p)(1-pq)b}{1-b^2(1-pq)^2}\right] \quad ...(7.1)$$

8. ISO-FAILURES CURVES: According to fig. 1 the browser share and initial share are linearly related when the condition of Iso-failure is imposed on the expressions. The quitting probability p_c if high the browser share reduces but the linear relationship between p and P_1 remains maintained.

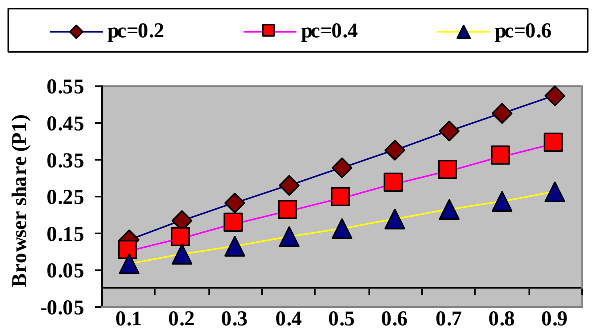

Fig. 1. Initial share (p)
(b=0.3, pq=0.5)

In fig. 2 and fig. 3 the similar pattern is observed when the Iso-failure level is kept high the browser share reduces.

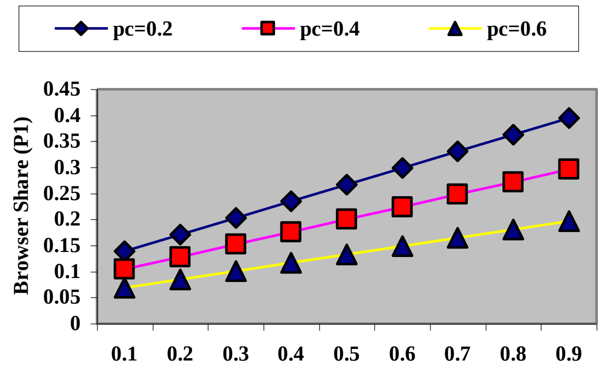

Fig. 2. Initial share (p)
b=0.5, pq=0.5)

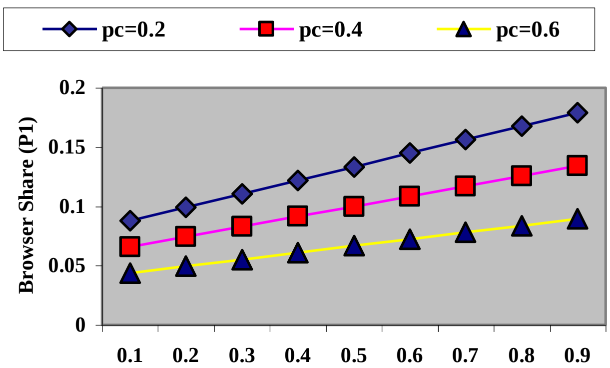

Fig. 3. Initial share (p)
(b=0.8, pq=0.5)

9. FITTING THE STRAIGHT LINE: We are to approximate the relationship between parameter P and p through a straight line $\hat{P}_1 = a + b.p$ where a and b are constants to be obtained by the method of least square. For the i^{th} observation p_i we write the relationship as $\hat{P}_{1i} = a + b.p_i$ (i=1, 2, 3,…, n). The normal equations are

$$\left.\begin{aligned} \sum_{i=1}^{n} P_{1i} &= n.a + b\sum_{i=1}^{n} p_i \\ \sum P_{1i}.p_i &= a\sum_{i=1}^{n} p_i + b\sum_{i=1}^{n} p_i^2 \end{aligned}\right\} \quad ...(9.1)$$

By solving the normal equations (9.1), the least square estimates of a and b are as $\hat{a}, \hat{b}$:

$$\hat{b} = \left\{ \frac{n\sum_{i=1}^{n} P_{1i} p_i - (\sum_{i=1}^{n} P_i)(\sum p_i)}{n\sum_{i=1}^{n} p_i^2 - (\sum_{i=1}^{n} p_i)^2} \right\} \quad \text{.....} \quad (9.2)$$

$$\hat{a} = \left\{ \frac{1}{n} \sum_{i=1}^{n} P_{1i} - \hat{b} \sum_{i=1}^{n} p_i \right\} \quad \text{...}(9.3)$$

Where n is the number of observations in sample of size n, the resultant straight line is

$$\hat{P}_1 = \left\{ \hat{a} + \hat{b}.p \right\} \quad \text{....}(9.4)$$

The coefficient of determination (COD) as a measure of good curve fitting is

$$\text{COD}=\frac{\sum\left(\hat{P}_{1i} - \overline{P}_1\right)^2}{\sum\left(P_{1i} - \overline{P}_1\right)^2} \quad \text{...}(9.5)$$

Where $\overline{P} = \frac{1}{n}\sum P_{1i}$ is mean of original data of variable P_1 obtained through Markov chain model. The term $\hat{P}_{1i} = \hat{a} + \hat{b}.p_i$ is the estimated by values of P_{1i} given observation p_i. The COD lies between 0 to 1. If the line is good fit then it is near to 1. We generate pair of values (p, P_1) in tables (9.1, 9.2, and 9.3) by providing few fixed input parameters.

Table 9.1 (P_1 by expression (6.1), $\hat{P}_1$ by (9.4) with known p_c, b, p_q, (9.4.1))

Fixed parameter	P	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	**COD**
p_c=0.4 ,b=0.3 p_q=0.5	P_1	0.1009	0.1374	0.1740	0.2105	0.2470	0.2835	0.3201	0.3566	0.3931	**1.000**
	$\hat{P}_1$	0.1009	0.1374	0.1740	0.2105	0.2470	0.2835	0.3201	0.3566	0.3931	
$\hat{a} = 0.06445$; $\hat{b} = 0.365217$; $P_1 = \hat{a} + \hat{b}.p$; $\hat{P}_1 = 0.06445 + 0.365217(p)$ ….(9.4.1)											

Table 9.2 (P_1 by expression (6.1), $\hat{P}_1$ by (9.4) with known p_c, b, p_q, (9.4.2))

Fixed parameter	P	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	**COD**
p_c=0.4 b=0.5 p_q=0.5	P_1	0.1041	0.1281	0.1520	0.1761	0.2110	0.2240	0.2481	0.2721	0.2961	**1.000**
	$\hat{P}_1$	0.1041	0.1281	0.1521	0.1761	0.2110	0.2240	0.2481	0.2720	0.2961	
$\hat{a} = 0.081271$; $\hat{b} = 0.241254$; $P_1 = \hat{a} + \hat{b}.p$; $\hat{P}_1 = 0.081271 + 0.241254(p)$ …(9.4.2)											

Table 9.3 (P_1 by expression (6.1), $\hat{P}_1$ by (9.4) with known p_c, b, p_q, (9.4.3))

Fixed parameter	P	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	**COD**
p_c=0.4 b=0.7 p_q=0.5	P_1	0.0851	0.0984	0.1117	0.1251	0.1384	0.1517	0.1651	0.1784	0.191	**1.000**
	$\hat{P}_1$	0.0851	0.0984	0.1117	0.1251	0.1384	0.1517	0.1651	0.1784	0.191	
$\hat{a} = 0.071791$; $\hat{b} = 0.133333$; $P_1 = \hat{a} + \hat{b}.p$; $\hat{P}_1 = 0.071791 + 0.133333(p)$ …(9.4.3)											

10. CONFIDENCE INTERVAL: The 100(1- α) percent confidence interval for a and b are

$$\hat{a} \pm \left\{ t_{(n-2)} \frac{\alpha}{2} \right\} .s. \left[\sqrt{\frac{1}{n} + \frac{\bar{p}}{\sum_{i=1}^{n} (p_i - \bar{p})^2}} \right] \quad ...(10.1)$$

$$\hat{b} \pm \left\{ t_{(n-2)}, \frac{\alpha}{2} \right\} .s. \left[\sqrt{\sum_{i=1}^{n} (p_i - \bar{p})^2} \right] \quad ...(10.2)$$

Where s= $\sqrt{\frac{\sum (P_{1i} - \hat{P}_{1i})^2}{n-2}}$

Table: 10.1 Calculation of Confidence interval for a and b

Fixed parameter P_c =0.4, b=0.3, p_q=0.5	$\hat{a}$ =0.06445	$\hat{b}$ =0.365217	(a=0.06445, a=0.06445) (b= 0.365217 , b=0.365217)
P_c =0.4,b=0.5, p_q=0.5	$\hat{a}$ =0.081271	$\hat{b}$ =0.241254	(a=0.081271, a=0.081271) (b=0.241254 , b=0.241254)
P_c =0.4,b=0.7, p_q=0.5	$\hat{a}$ =0.071791	$\hat{b}$ =0.133331	(a=0.07179 , a=0.07179), (b=0.13333 , b=0.013333)
Average Estimate	$\bar{a} = 0.072504$	$\bar{b} = 0.2465983$	$\hat{P}_1 = \bar{a} + \bar{b}(p)$ $\hat{P}_1 = 0.072504 + 0.2465983(p)$

11. DISCUSSIONS:

While considering fig. 1 we observe that there is a linear trend exists between final browser share and initial share using (6.1). This trend increases with the increase of initial share. If p_c probability is high then browser share level is low. It seems final browser share is inversely proportional to the quitting probability p_c which is usual also. The same pattern appears in fig. 2 and fig. 3. The linear pattern between p and $\bar{p}_1$ is replaced by a direct equation of a straight line in the form $\hat{P}_1 = \hat{a} + \hat{b}.p$. The least square estimates of $\hat{a}$ are 0.06445, 0.081271, 0.07179 and $\hat{b}$ are 0.365217, 0.241254, 0.13333 respectively. The three possible equations of linear relationship between p and $\hat{P}_1$ are $\hat{P}_1$ =(0.06445+0.365217.p), $\hat{P}_1$ =(0.081271+0.241254.p), $\hat{P}_1$ =(0.071791+0.13333.p)

The coefficients of determination (COD) in each case are exactly 1 therefore the estimated values of a and b are very close to the real values. The confidence intervals are

For $\hat{a}$: (0.06445, 0.06445), (0.08271, 0.081271), (0.071791, 0.071791)

For $\hat{b}$: (0.365217, 0.365217), (0.241254, 0.241254), (0.13333, 0.13333)

The average equation of linear relationship is

$$\hat{P}_1 = \bar{a} + \bar{b}(p); \; \hat{P}_1 = 0.72504 + 0.2465983(p).$$

12. CONCLUSION:

The least square based line fitting between P_{1i} *and* $\hat{p}_1$ is accurate because of high values of COD which is nearly equal to unity. The confidence interval for $\hat{a}$ and $\hat{b}$ are overlapping with the true values showing the robustness of the estimates. It seems the fittings of straight lines are good approximations of the complicated relationships between final browser share probability and initial browser share probability. The resultant relationship is

$$\hat{P}_1 = \bar{a} + \bar{b}(p); \; \hat{P}_1 = 0.072504 + 0.2465983(p)$$

which could be used us a rule for quick calculations

REFERENCES

[1] **Medhi, J. (1991):** Stochastic models in queuing theory, Academic Press Professional, Inc., San Diego, CA.

[2]. **Medhi, J. (1992):** Stochastic Processes, Ed.4, Wiley Eastern Limited (Fourth reprint), New Delhi.

[3]. **Chen, D.X. and Mark, J.W. (1993):** A fast packet switch shared concentration and output queuing,

IEEE Transactions on Networking, vol. 1, no. 1, pp. 142-151.

[4]. **Hambali, H. and Ramani, A. K., (2002):** A performance study of at multicast switch with different traffics, Malaysian Journal of Computer Science. Vol. 15, Issue No. 02, Pp. 34-42.

[5]. **Naldi, M. (2002):** Internet access traffic sharing in a multi-user environment, Computer Networks. Vol. 38, pp. 809-824.

[6]. **Newby, M. and Dagg, R. (2002):** Optical inspection and maintenance for stochastically deteriorating systems: average cost criteria, Jour. Ind. Statistical Associations. Vol. 40, Issue No. 02, pp. 169-198.

[7]. **Francini, A.** and **Chiussi, F.M. (2002):** Providing QoS guarantees to unicast and multicast flows in multistage packet switches, IEEE Selected Areas in Communications, vol. 20, no. 8, pp. 1589-1601.

[8]. **Dorea, C.C.Y., Cruz and Rojas, J. A. (2004):** Approximation results for non-homogeneous Markov chains and some applications, Sankhya. Vol. 66, Issue No. 02, pp. 243-252.

[9]. **Paxson, Vern, (2004):** Experiences with internet traffic measurement and analysis, ICSI Center for Internet Research International Computer Science Institute and Lawrence Berkeley **National Laboratory.**

[10]. **Yeian, C. and Lygeres, J. (2005):** Stabilization of class of stochastic differential equations with Markovian switching, System and Control Letters. Issue 09, pp. 819-833.

[11]. **Shukla, D., Gadewar, S. and Pathak, R.K. (2007 a):** A stochastic model for space division switches in computer networks, International Journal of Applied Mathematics and Computation, Elsevier Journals, Vol. 184, Issue No. 02, pp235-269.

[12]. **Shukla, D. and Thakur, Sanjay, (2007 b**) Crime based user analysis in internet traffic sharing under cyber crime, Proceedings of National Conference on Network Security and Management (NCSM-07), pp. 155-165, 2007.

[13]. **Agarwal, Rinkle and Kaur, Lakhwinder (2008):** On reliability analysis of fault-tolerant multistage interconnection networks, International Journal of Computer Science and Security (IJCSS) Vol. 02, Issue No. 04, pp. 1-8.

[14]. **Shukla, D., Tiwari, Virendra, Thakur, S. and Deshmukh, A. (2009 a):**Share loss analysis of internet traffic distribution in computer networks, International Journal of Computer Science and Security (IJCSS), Malaysia, Vol. 03, issue No. 05, pp. 414-426.

[15]. **Shukla, D., Tiwari, Virendra, Thakur, S. and Tiwari, M. (2009 b) :**A comparison of methods for internet traffic sharing in computer network, International Journal of Advanced Networking and Applications (IJANA).Vol. 01, Issue No.03, pp.164-169.

[16]. **Shukla, D., Tiwari, V. and Kareem, Abdul, (2009 c)** All comparison analysis in internet traffic sharing using markov chain model in computer networks, Georgian Electronic Scientific Journal: Computer Science and Telecommunications. Vol. 06, Issue No. 23, pp. 108-115.

[17]. **Shukla, D., Tiwari, Virendra, and Thakur, S. (2010 a):** Effects of disconnectivity analysis for congestion control in internet traffic sharing, National Conference on Research and Development Trends in ICT (RDTICT-2010), Lucknow University, Lucknow.

[18]. **Shukla, D., Gangele, Sharad and Verma, Kapil, (2010 b):** Internet traffic sharing under multi-market situations, Published in Proceedings of 2nd National conference on Software Engineering and Information Security, Acropolis Institute of Technology and Research, Indore, MP, (Dec. 23-24,2010), pp 49-55.

[19]. **Shukla, D., and Thakur, S**. **(2010 c)**: Stochastic Analysis of Marketing Strategies in internet Traffic, INTERSTAT (June 2010).

[20]. **Shukla, D., Tiwari, V., and Thakur, S., (2010 d):** Cyber Crime Analysis for Multi-dimensional Effect in Computer Network, Journal of Global Research in Computer Science(JGRCS), Vol. 01, Issue 04, pp.31-36.

[21]. **Shukla, D., Gangele, Sharad, Verma, Kapil and Singh, Pankaja (2011 a):** Elasticity of Internet Traffic Distribution Computer Network in two Market Environment, Journal of Global research in Computer Science (JGRCS) Vol.2, No. 6, pp.6-12.

[22]. **Shukla, D., Gangele, Sharad, Verma, Kapil and Singh, Pankaja (2011 b):** Elasticities and Index Analysis of Usual Internet Browser share Problem, International Journal of Advanced Research in Computer Science (IJARCS),Vol. 02, No. 04, pp.473-478.

[23]. **Shukla, D., Gangele, Sharad, Verma, Kapil and Thakur, Sanjay, (2011 c):** A Study on Index Based Analysis of Users of Internet Traffic Sharing in Computer Networking, World Applied Programming (WAP), Vol. 01, No. 04, pp. 278-287.

[24]. **Shukla, D., Gangele, Sharad, Singhai, Rahul and Verma, Kapil, (2011 d):** Elasticity Analysis of Web Browsing Behavior of Users, International Journal of Advanced Networking and Applications (IJANA), Vol. 03, No. 03, pp.1162-1168.

[25]. **Shukla, D., Verma, Kapil and Gangele, Sharad, (2011 e):** Re-Attempt Connectivity to Internet Analysis of User by Markov Chain Model, International Journal of Research in Computer Application and Management (IJRCM) Vol. 01, Issue No. 09, pp. 94-99.

[26]. **Shukla, D., Gangele, Sharad, Verma, Kapil and Trivedi, Manish, (2011 f):** Elasticity variation under Rest State Environment In case of Internet Traffic Sharing in Computer Network, International Journal of Computer Technology and Application (IJCTA) Vol. 02, Issue No. 06, pp. 2052-2060.

[27]. **Shukla, D., Singhai, Rahul [2011]:** Analysis of User Web Browsing Using Markov chain Model, International Journal of Advanced Networking and Application (IJANA), Vol. 02, Issue No. 05, pp. 824-830.