Journal of Theoretical and Applied Information Technology
20th October 2013. Vol. 56 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

PROBABILISTIC QUEUING SCHEME FOR SERVICING E-MAILS USING MARKOV CHAINS

[1]OMAR SAID, [2]ALAA ELNASHAR

[1]College of Science, Menoufia University, Shbeen El_Kom, Egypt.
[2]College of Science, Minia University, Minia, Egypt.
[1,2]Information Technology Department,
College of Computers and Information Technology
Taif University, Taif, Saudi Arabia.
E-Mail: [1]o.saeed@tu.edu.sa, [2]a.ismail@tu.edu.sa

ABSTRACT

Recently, huge number of e-mails is sent and received. These e-mails are classified into spam and non spam ones which are processed with the same priority. To guarantee a higher priority service to non-spam e-mails than that is provided to spam e-mails, a two-priority queue scheme was proposed. There are some drawbacks in the two-priority queue scheme such as it fails to provide a Quality of Service (QoS) in case of network bottlenecks. In this paper, an algorithm for servicing non-spam e-mails with high probability is proposed. Markov chain is used to analyze the servicing probability of e-mails in case of two and three priority queue schemes. Results proved that the three-queue scheme provides a higher probability service to the most important non spam e-mails than that is provided to the same class in case of the two-queue scheme.

Keywords: *E-mail Systems; Internetworking; Markov Chain; Queuing Theory.*

1. INTRODUCTION

There are millions of e-mails which represent a load on the transmission channels [1] between senders and receivers. Spam constitutes dramatic percentage of the total e-mail traffic [2]. The spam control systems have accepted performance regarding e-mail blocking [3] ignoring the consumed QoS used by spam e-mails before blocking process. As the network QoS is an important issue, the handling process of e-mails should be started at the sender side and continued until receiver side. In single-queue scheme, e-mails are received and before serviced (a spam filtering [4], [5], mail classification [6], or spam detection [7]) by Mail Transfer Agent (MTA), they are stored in a buffer, which represented by a common queue with FIFS (First-Come First-Served) base. Spam control models try to develop classification techniques with acceptable accuracy [8], [9]. Learning or spam detection are required in most of the classification techniques [10], [11], [12]. Time performance (fast spam detection) is an important factor in high classification accuracy. In two-queue scheme, the e-mail prioritization idea is proposed to minimize the non-spam emails transmission delay and loss. The spam e-mails are inserted into a slow queue which results in a service delay. The non-spam ones should be inserted in a faster queue with high QoS. A well optimize classification may guarantee a high QoS which leads to high e-mails system efficiency [13]. This paper presents a model for three levels prioritized e-mails servicing which can classify not only spam e-mails but also non-spam ones.

The paper is organized as follows; Section 2 presents the related work. Section 3 introduces the new three-queue scheme. Section 4, the proposed queuing data model with extracted results and recommendations are showed. Finally, conclusion and future work are demonstrated in Sections 5 and 6 respectively.

2. RELATED WORK

Several studies have been presented to develop high reliability classification techniques that classify e-mail contents to differentiate between non-spam and spam e-mails [2], [5], [7]. There are very few studies that aim to improve classification timing performance [3]. Some techniques such as single-queue [8] and two-queue prioritized schemes have been proposed to reduce that delay [2], [13].

In this paper, the previous techniques are extended to develop a new three-queue prioritized scheme.

Journal of Theoretical and Applied Information Technology
20th October 2013. Vol. 56 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

2.1 Single queue service scheme

The single queue scheme is a specific module which used to reassemble e-mail packets and then inserts them in a single queue. Correspondingly, the spam detector services e-mails in the FIFO queue and determines to which folder will forward them (recipients' mailboxes or special junk e-mail). The e-mails, which are coming from assembler, have the same priority with constant delay as shown in Fig. 1. The delay value depends on the queue occupancy. So, there are two delay parameters; size of the e-mail queue and e-mail arrival rate. Hence; there is a relation between the number of coming e-mails and the service which should be available for the non-spam emails. This relation is formalized as follows; high spam arrival rate means non-spam e-mails queue delay. Queue overflow (mass-mailed worm outbreaks or Denial-of-Service (DoS) attacks) causes single-queue system fails. This scheme uses M/M/1/B queuing model [14].

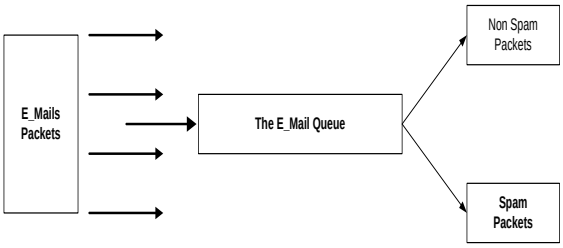

Figure1: Single-queue e-mail servicing scheme.

2.2 Two-queue service scheme

The two-queue scheme is considered as two-class classification scheme. E-mails are classified as either non-spam or spam. Based on layer-3 e-mail, by estimating e-mail classes, a prioritized e-mail servicing was proposed. The non-spam e-mails have higher priority of servicing than that is provided to spam e-mails. E-mails are forwarded either to a fast non-spam queue or to a slow spam queue. The proposed two-queue scheme is not limited to the layer-3 e-mail class estimation. If the e-mail class information can be known when e-mails are posted to queues, the two-queue system may work with other class estimation techniques. Fig. 2 shows the proposed two-queue scheme. In this system, there are four strategies to describe the e-mail services using discrete-time Markov chain analysis [15].

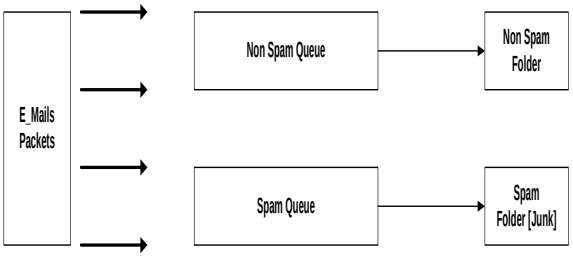

Figure 2: Two-queue e-mail servicing scheme

This scheme improved the loss and delay of non-spam queue when the services are prioritized. Adapting categorization probabilities to the queues' tenures gave better overall cost to process e-mails [11]. In addition, this scheme protects e-mail servers from being overloaded under spam attacks.

3. THREE-QUEUE SERVICE SCHEME STRUCTURE

We present a new scheme based on a three-priority queuing system. In this scheme, we try to prioritize non-spam e-mails on servers to decrease both the probability of delay and loss. The non-spam e-mails are filtered into two classes, namely class1 and class2. The most important non spam e-mails are assigned to class1 based on predefined rules. Non-spam e-mails having less importance are assigned to class2. Class1 e-mails are enqueued into a fast queue that has higher service priority. Class2 e-mails are enqueued into another queue with middle priority. Spam e-mails are separated and enqueued into a low priority queue. The proposed scheme structure is shown in Fig. 3.

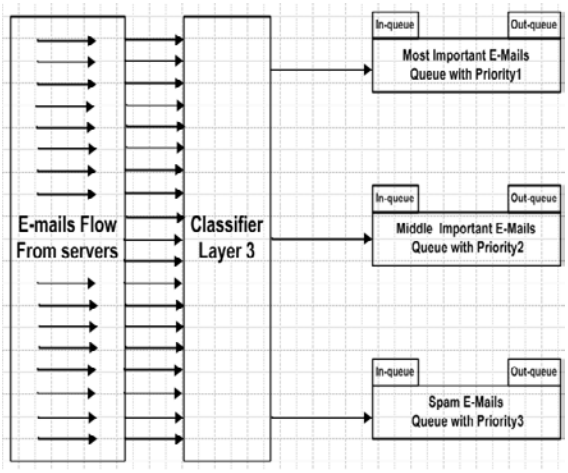

Figure 3: Three-queue e-mail servicing scheme structure.

For integration, next-generation of routers should be designed to make e-mails pre-classification on layer 3.

Journal of Theoretical and Applied Information Technology
20th October 2013. Vol. 56 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

3.1 Scheme mechanism

In this scheme, a linear data model is used. The linear formula used is $P_{ij} = \lambda_i + \mu_j$, where P_{ij} represents the transition probability from state i to state j, λ and μ represent the birth and death rates respectively.

The scheme represents the three e-mail classes by the corresponding queues Q1, Q2, and Q3 respectively. At any time, any e-mail may be enqueued into one of these three queues by the classifier as shown in Fig. 3. The three-queue scheme mechanism is summarized in the following steps:

Step1: The classifier distributes the incoming e-mails into the system's queues according predefined priorities within a time interval determined by the server. The initial probabilities of the classifier selection of system queues are pr_1, pr_2 and pr_3, where $\sum_{i=1}^{3} pr_i$.

Step2: During servicing, within a time interval, if the serviced e-mail cannot be completed due to low bandwidth or queue congestion, the proposed system will give the QoS priority to the next e-mail which should be serviced in the same queue.

Step3: The incomplete processed e-mail should be transferred to the next queue with the lower priority.

Step4: While Q3 is empty, the e-mail processing will continue in Q1 with a predefined time interval provided that QoS, which is required for Q1, is available.

Step5: Once Q1 becomes empty, the system starts to service e-mails in Q2.

Step6: If an e-mail has incomplete service in Q2, it should be moved to the next queue Q3.

Step7: While the system servicing e-mails in Q3, if an e-mail is received into Q1, the system jumps immediately to Q1 and service this new coming e-mail.

3.2 Scheme state transitions

To clarify the Markov chain model for the proposed queuing system, we define Q1 as state1, Q2 as state2, and Q3 as state3. The time which is used by the system to service one e-mail is denoted by τ , ($\tau = 1, 2, 3 \ldots$). The transitions over these states are for either servicing or waiting. The transition from one state to any other state is shown in Fig. 4.

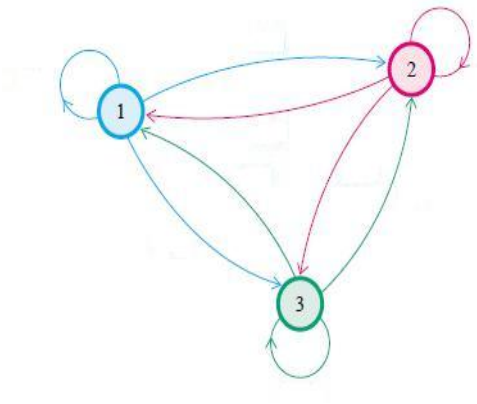

Figure 4: Three states transition diagram

Let $\{Y^{(\tau)}, \tau \geq 1\}$ is a Markov chain where $Y^{(\tau)}$ denotes the selected queue at the τ^{th} time interval. The state space of the random variable *Y* is $\{Q_1, Q_2, Q_3\}$. The initial selection probabilities of queues are: $Pr_1 = P[Y^{(0)} = Q_1]$,

$Pr_2 = P[Y^{(0)} = Q_2]$ and $Pr_3 = P[Y^{(0)} = Q_3]$ where $\sum_{i=1}^{3} pr_i = 1$.

Let $S_{ij}(i, j = 1, 2, 3)$ be the transaction probabilities of the system over the three states, the state transition matrix for $Y^{(\tau)}$ is expressed as:

$Y^{(\tau)}$

$Y^{(\tau-1)}$	$\mathbf{Q_1}$	$\mathbf{Q_2}$	$\mathbf{Q_3}$
$\mathbf{Q_1}$	S_{11}	S_{12}	S_{13}
$\mathbf{Q_2}$	S_{21}	S_{22}	S_{23}
$\mathbf{Q_3}$	S_{31}	S_{32}	S_{33}

To handle all transitions, we applied the row dependent model $P_{ij} = \lambda_i + i(\mu_j)$.

4. THREE-QUEUE SCHEME ANALYSIS

There are two parameters λ and μ used to manage the data analysis related to queue transition probabilities. Fig. 5 and Fig. 6 show the data model for both two-queue and three-queue schemes.

$Y^{(\tau)}$

$Y^{(\tau-1)}$	$\mathbf{Q_1}$	$\mathbf{Q_2}$

Journal of Theoretical and Applied Information Technology
20th October 2013. Vol. 56 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Q1	λ	$1-\lambda$
Q2	$\lambda+\mu i$	$1-(\lambda+\mu i)$

Figure 5: Two-queue scheme probabilities matrix

$Y^{(\tau)}$

$Y^{(\tau-1)}$	**Q1**	**Q2**	**Q3**
Q1	λ	$\lambda+\mu i$	$1-(2\lambda+\mu i)$
Q2	$\lambda+\mu i$	$\lambda+2\mu i$	$1-(2\lambda+3\mu i)$
Q3	$\lambda+2\mu$	$\lambda+3\mu i$	$1-(2\lambda+5\mu i)$

Figure 6: Three-queue scheme probabilities matrix

Where *i*, denotes the number of queues, (*i* = 3), in case of three-queue scheme.

Now, we search for the values of both λ and μ which give the best performance of the proposed system. In case of λ = 0.1 and μ = 0.002, we obtain the following transition probabilities matrices for both two-queue and three-queue schemes respectively.

$Y^{(\tau)}$

$Y^{(\tau-1)}$	**Q1**	**Q2**
Q1	0.1	0.9
Q2	0.104	0.896

$Y^{(\tau)}$

$Y^{(\tau-1)}$	**Q1**	**Q2**	**Q3**
Q1	0.1	0.104	0.796
Q2	0.102	0.108	0.79
Q3	0.104	0.112	0.784

Fig.7 shows that at the value of $\lambda = 0.1$ and the value of $\mu = 0.002$, the system shifts over to the spam queue in both schemes (Tow and Three queues). The probability of spam queue service decreases as the value of μ increases for the same value of λ (λ = 0.1 and μ = 0.004, 0.006 and 0.01).

Fig. 8 (a, b) shows that the probability of spam queue service is decreased more rapidly in case of the three-queue scheme. It is notable that for this value of λ, (λ = 0.1), the probability of non-spam queue(s) is still less than spam queue which is not a good sign and indicates that the λ value should be changed, see Fig. 7, Fig. 8, and Fig. 9.

$\lambda = 0.1, \mu = 0.002$

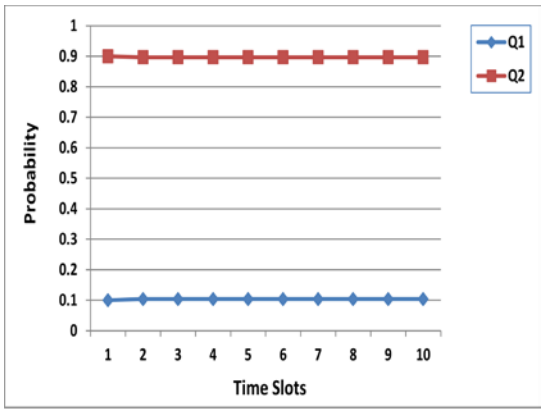

Figure 7: (a) Two-queue scheme service probability.

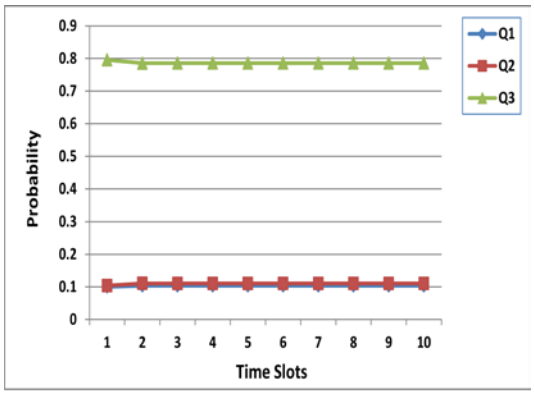

Figure 7: (b) Three-queue scheme service probability.

$\lambda = 0.1, \mu = 0.004$

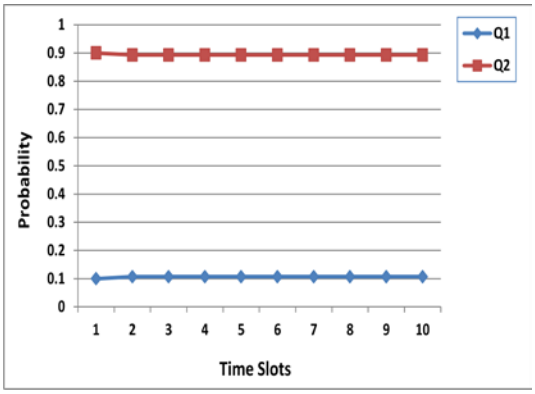

Figure 8: (a) Two-queue scheme service probability.

Journal of Theoretical and Applied Information Technology
20th October 2013. Vol. 56 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

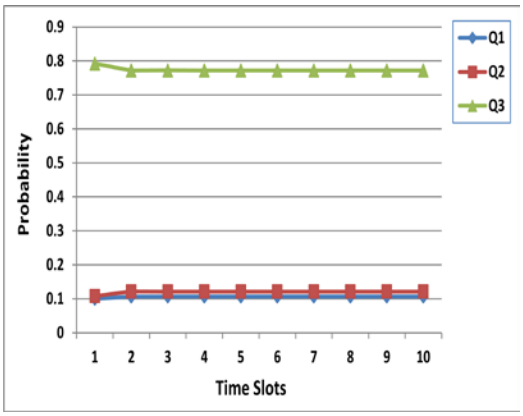

Figure 8: (b) Three-queue scheme service probability.

$\lambda = 0.1, \mu = 0.006$, these values give results like in Fig. 7 and Fig. 8 (a, b) with minor difference.

$$\lambda = 0.1, \mu = 0.01$$

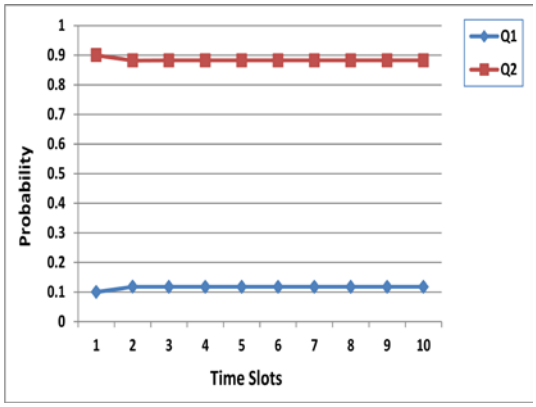

Figure 9: (a) Two-queue scheme service probability.

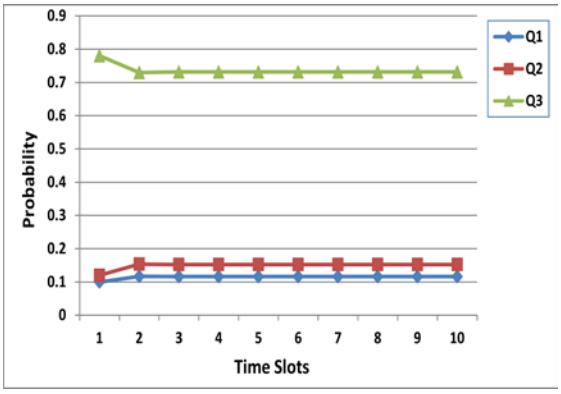

Figure 9: (b) Three-queue scheme service probability.

Increasing the value of λ to 0.2 and varying the value of μ = (0.004, 0.006 and 0.01) make the system so as to be enhanced; as the gap between the probability of servicing spam and non-spam queues is reduced. But the results are still unsatisfactory since the probability of servicing spam is still over that of non-spam; see Fig. 10, Fig. 11, and Fig. 12.

$$\lambda = 0.2, \mu = 0.002$$

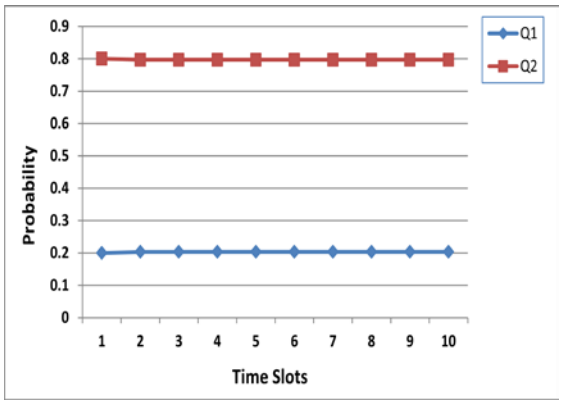

Figure10: (a) Two-queue scheme service probability.

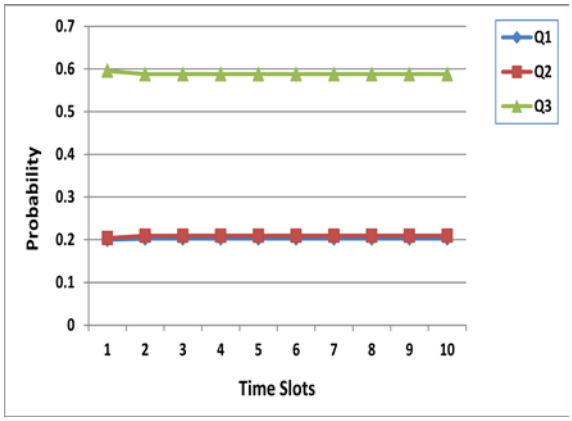

Figure10: (b) Three-queue scheme service probability.

$\lambda = 0.2, \mu = 0.004$, these values give the results like in Fig. 10 (a, b) with minor difference.

$$\lambda = 0.2, \mu = 0.006$$

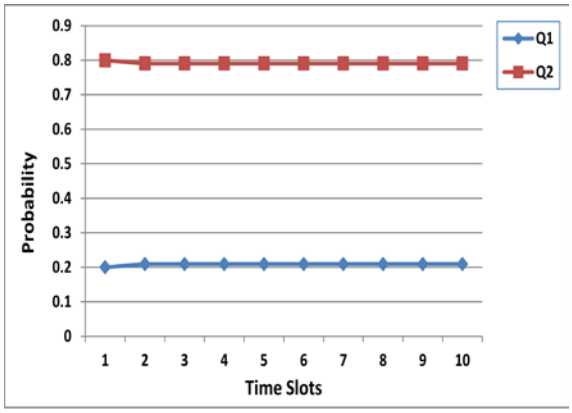

Figure 11: (a) Two-queue scheme service probability.

Journal of Theoretical and Applied Information Technology
20th October 2013. Vol. 56 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

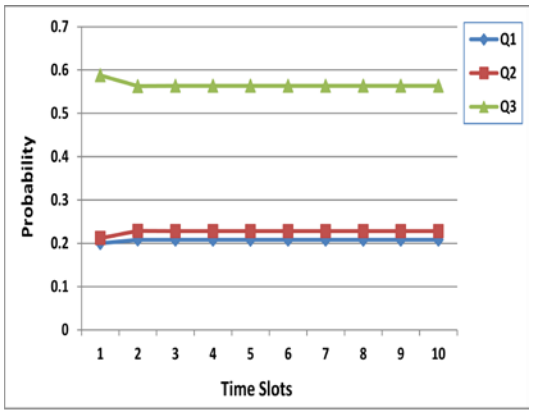

Figure11: (b) Three-queue scheme service probability

$$\lambda = 0.2, \mu = 0.01$$

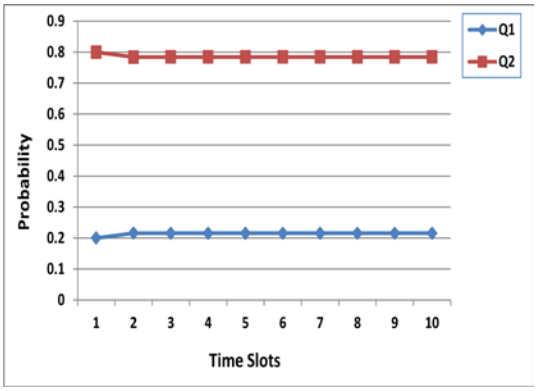

Figure 12: (a) Two-queue scheme service probability.

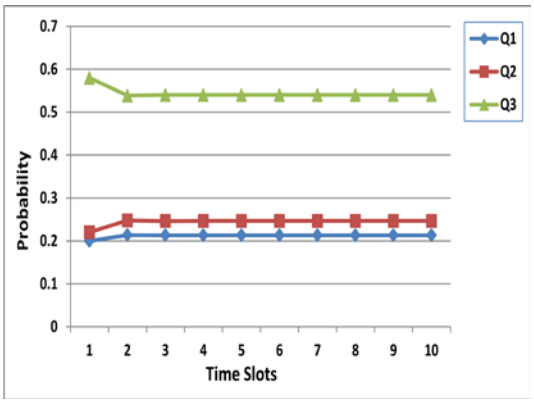

Figure 12: (b) Three-queue scheme service probability.

This gives an indication that the value of λ should be increased for arbitrary values of μ = (0.004, 0.006 and 0.01) to achieve a higher probability for servicing non-spam e-mails. At $\lambda = 0.3$, the system takes the same behaviour as in case of $\lambda = 0.1$ and 0.2 with more diversion between the probabilities of servicing spam and non-spam queues as shown in Fig. 13, Fig. 14, and Fig. 15.

$$\lambda = 0.3, \mu = 0.002$$

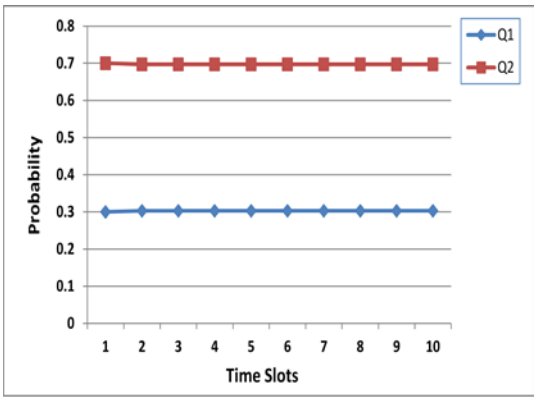

Figure13: (a) Two-queue scheme service probability.

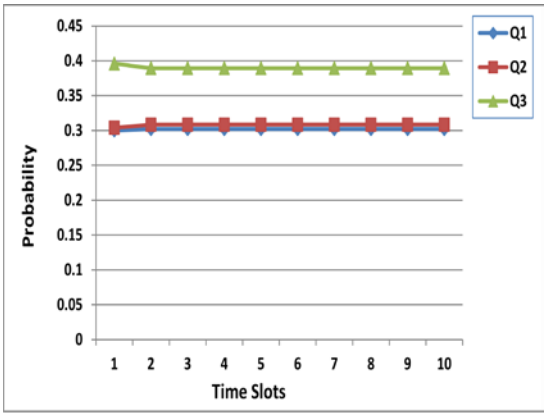

Figure 13: (b) Three-queue scheme service probability.

$\lambda = 0.3, \mu = 0.004$, these values give the results like in Fig. 13 (a, b) with minor difference.

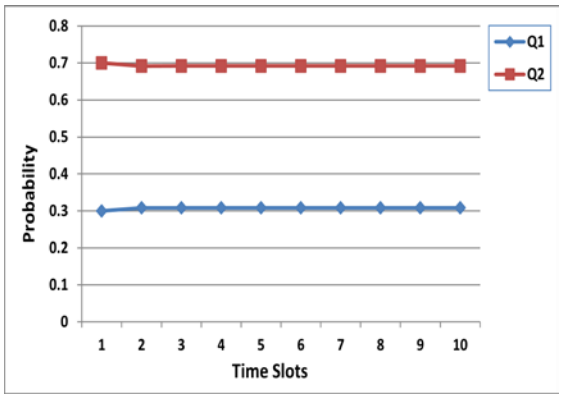

Figure 14: (a) Two-queue scheme service probability.

Journal of Theoretical and Applied Information Technology
20th October 2013. Vol. 56 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

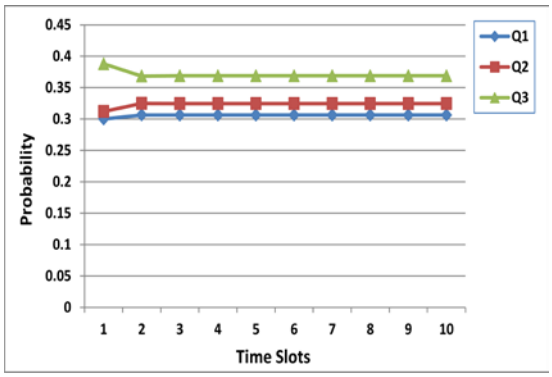

Figure 14: (b) Three-queue scheme service probability.

$$\lambda = 0.3, \mu = 0.01$$

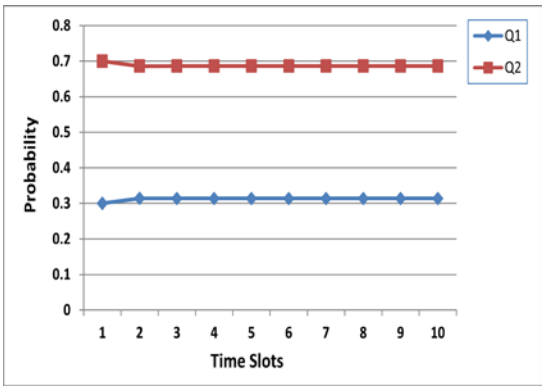

Figure15: (a) Two-queue scheme service probability.

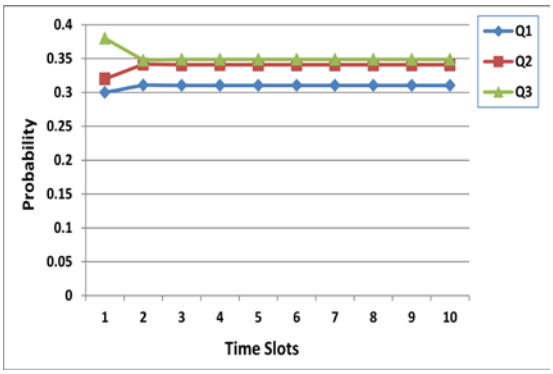

Figure15: (b) Three-queue scheme service probability.

From the previous results, it is notable that increasing the value λ enhances the overall system behavior, since the probability of serving non-spam queue also increases. Furthermore, the gap between the servicing probabilities of spam and non-spam is more reduced in case of three-queue scheme. Upon this fact, the λ value should be increased to get best result for the proposed scheme. Fig. 16, Fig. 17, and Fig. 18show that increasing the value of λ to 0.4 for the values of μ = (0.002, 0.004, 0.006, and 0.1) , the probability of servicing non-spam queues is changed to be greater than the probability of serving spam queues in case of three-queue scheme. This implies an advantage of the proposed scheme over the two-queue scheme which is clear in the optimal case, at $\lambda = 0.4$ and $\mu = 0.002$. In case of the two-queue scheme, the probability of servicing q1, which represents non-spam emails queue, is less than that of q2, which represents the spam queue. This reflects the scheme failure at these values of λ and μ as shown in Fig. 16 (a). Regarding the three-queue scheme; the probability of servicing q2 is over that of q1 with minor difference which means that the middle importance e-mails will be serviced firstly as shown in Fig. 16 (b). Fixing the value of $\lambda = 0.04$ and varying the values of μ = (0.004, 0.006, and .01), the non-spam emails servicing probabilities are still greater than that of the spam emails but the difference between the probabilities of servicing both q1 and q2 increases as the value of μ increases. This indicates that the chance of serving most important emails may be decreased, as shown in Fig 17 and Fig. 18.

$$\lambda = 0.4, \mu = 0.002$$

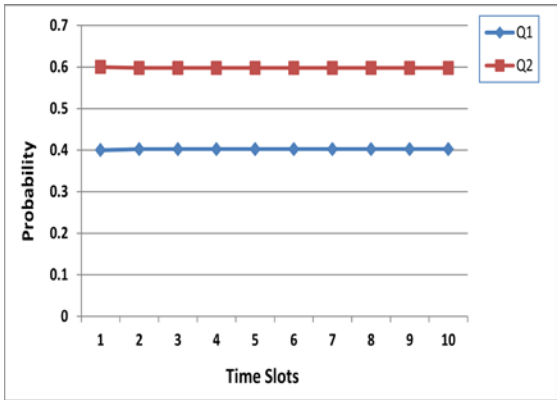

Figure16: (a) Two-queue scheme service probability.

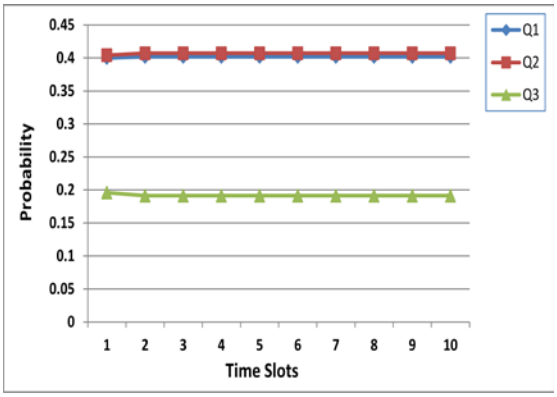

Figure 16: (b) Three-queue scheme service probability.

Journal of Theoretical and Applied Information Technology
20th October 2013. Vol. 56 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

$\lambda = 0.4, \mu = 0.004$, these values give the results like in Fig. 16 (a, b) with minor difference.

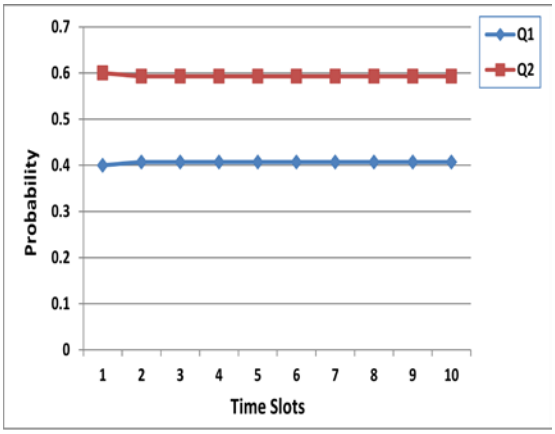

Figure17: (a) Two-queue scheme service probability.

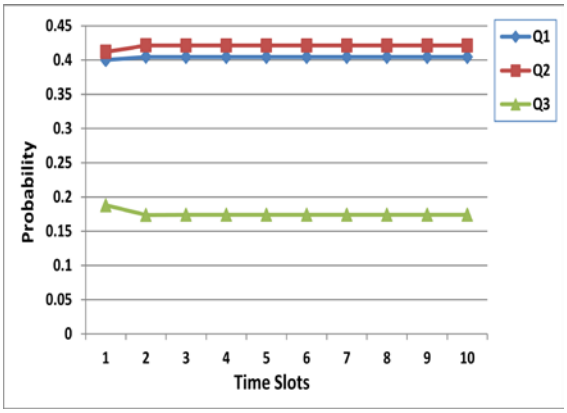

Figure17: (b) Three-queue scheme service probability.

$\lambda = 0.4, \mu = 0.01$

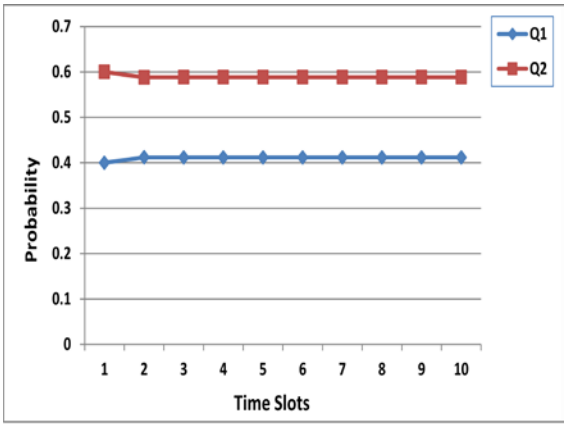

Figure18: (a) Two-queue scheme service probability.

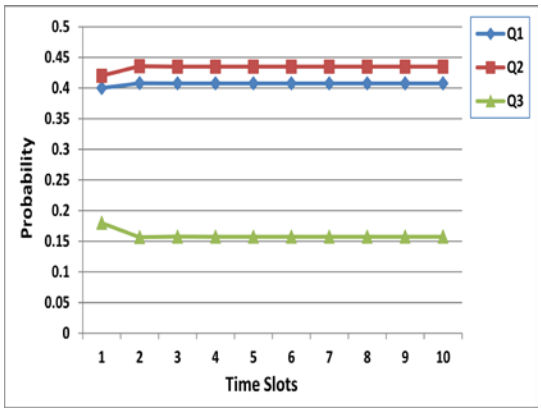

Figure18: (b) Three-queue scheme service probability.

Increasing and then fixing the value of λ to 0.05 and varying the values of μ (0.004, 0.006, and .01), the spam emails servicing probabilities in both the two and the three-queue schemes became negative ; this is can't occur practically, see Fig. 19 and Fig. 20.

$\lambda = 0.5, \mu = 0.002$

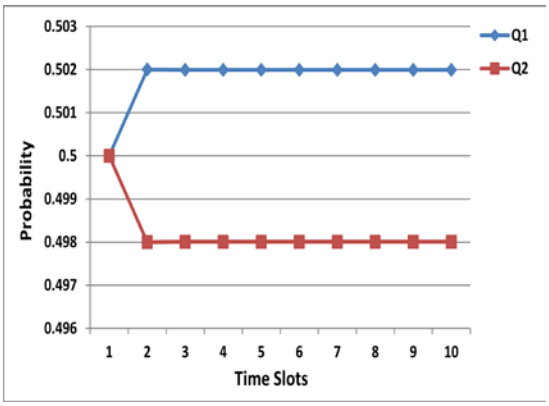

Figure19: (a) Two-queue scheme service probability.

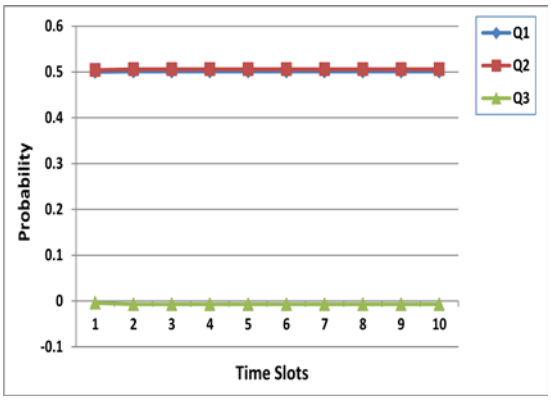

Figure19: (b) Three-queue scheme service probability.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

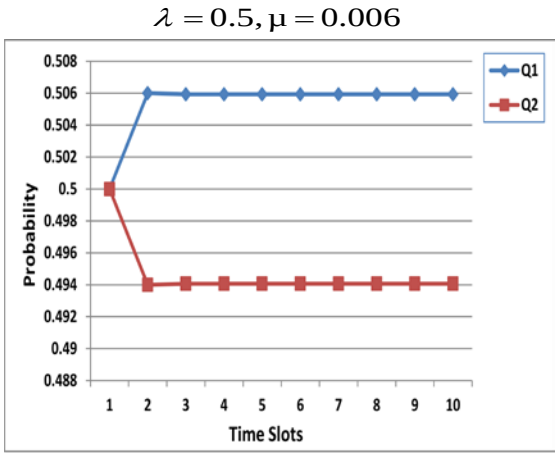

Figure 20: (a) Two-queue scheme service probability.

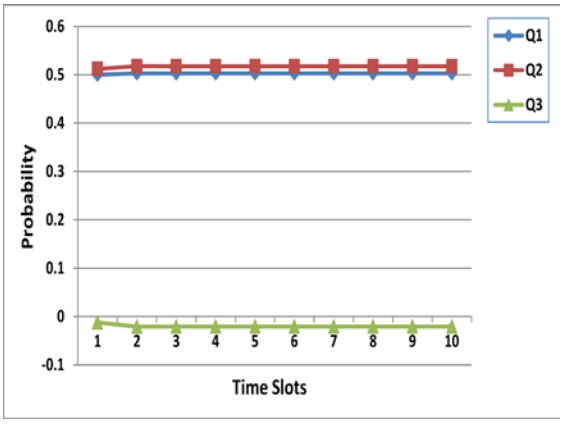

Figure 20: (b) Three-queue scheme service probability.

Our practical results showed that the optimal state is obtained when the value of $\lambda = 0.04$ and the value of $\mu = 0.002$, which is very close to the real world e-mail systems; where the number of middle importance e-mails is greater than the most importance ones and less than that of spam ones. Hence, the middle importance e-mails can be considered as the most common ones, which are sent and received periodically. So, this type of e-mails should be serviced with a high probability to decrease the network starvation especially in the bottleneck points which may face the e-mails within their trip. For the most important e-mails, the probability of serving should be less than that provided to the middle important ones. Also, the spam e-mails represent a network load and there is no problem if they lost. The most important e-mails should be assigned the next priority since they don't represent a load on the network. In addition, their needs to the best QoS may lead to some delay waiting for the availability of the required QoS. Furthermore, the waiting time may give a considerable chance to move from q2 to q1 in servicing process which is settled in the three-queue scheme.

5. CONCLUSION

In this paper, three-level prioritized e-mail system using Markov chain is introduced. In this system, the e-mails are distributed among three queues. The first priority queue services the e-mails which contain highest importance data. The second one is allocated for the middle importance e-mails. The last queue contains spam e-mails which should be serviced with the lowest priority. The application, which uses the proposed system, should specify a quota form each type of e-mails for each user. Also, in this paper, a data model for two and three queue schemes is introduced. The results showed that the optimal state of the proposed system is obtained at the value of $\lambda = 0.04$ and the value of $\mu = 0.002$ which provides high chance for the most important e-mails (depending on the sender perspective) to be firstly serviced and benefit from the network QoS.

6. FUTURE WORK

In the future work, two main topics should be covered. The first one is to study the system generalization which provides n-levels of prioritization for e-mails. The second one is related to adapting of the proposed system to handle several network parameters and loads such as delay and loss.

REFRENCES:

[1] James Turnbull, Peter Lieverdink, Dennis Matotek Pro Linux System Administration, "Mail Services" chapter 10, Apress, 2009, pp 443-516, ISBN: 1430219122, 978130219125

[2] Twining RD, Williamson MM, Mowbray M, Rahmouni M. Email prioritization: reducing delays on legitimate mail caused by junk mail. HP Digital Media Systems Laboratory, Bristol, UK, Technical Report HPL-2004–5(R.1), May 2004.

[3] Goodman J, Heckerman D, Rounthwaite R. Stopping spam. Scientific American 2005; 292(4): 42–49.

[4] Bratko A, Filipič B. Spam filtering using compression models. Technical Report IJS-DP-9227, Department of Intelligent Systems, Jožef Stefan Institute, Ljubljana, Slovenia, 2005.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[5] Segal R, Crawford J, Kephart J, Leiba B. Spamguru: an enterprise anti-spam filtering system. In Proceedings of the First Conference on Email and Anti-Spam (CEAS), Mountain View, CA, July 2004

[6] Borg, Anton "E-mail Classification Using Social Network Information", International Conference on Availability, Reliability and Security (ARES), 2012, Prague, Page(s): 168 – 173, 20-24, Aug. 2012.

[7] Tseng, Chi-Yao Sung, Pin-Chieh; Chen, Ming-Syan Syan, Cosdes: A Collaborative Spam Detection System with a Novel E-Mail Abstraction Scheme" IEEE Transactions on Knowledge and Data Engineering,, Volume: 23, Issue: 5, May 2011, Page(s): 669 – 682.

[8] Marsono MN, El-Kharashi MW, Gebali F, Ganti S, A distributed e-mail classification for spam control. In Proccedings of the 2006 Canadian Conference on Electrical and Computer Engineering (CCECE), Ottawa, Canada, May 2006, Page(s):438–441.

[9] SpamAssassin. Available: http://spamassassin.apache.org/. [July 2007].

[10] Yoshida K, Adachi F, Washio T, Motoda H, Homma T, Nakashima A, Fujikawa H, Yamazaki K. Density-based spam detector. In Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), Seattle, WA, Page(s): August 2004, 486–493.

[11] Panigrahi, Prabin Kumar "A Comparative Study of Supervised Machine Learning Techniques for Spam E-mail Filtering", International Conference on Computational Intelligence and Communication Networks (CICN), 2012, Uttar Pradesh, India, 3-5 Nov. 2012, Page(s): 506 – 512.

[12] Madi, Nadim K M, Implementation of secure email server in cloud environment, International Conference on Computer and Communication Engineering (ICCCE), Kuala Lumpur, Malaysia, 3-5, July 2012, Page(s): 28 – 32.

[13] M. Marsono, M. Watheq El-Kharashi, Fayez Gebali, Prioritized e-mail servicing to reduce non-spam delay and loss: a performance analysis, International Journal of Network Management, Volume, 18 Issue 4, August 2008, , Page(s): 323-342.

[14] Gebali F. Computer Communications Networks: Analysis and Design (3rd edn). North Star Digital Design: Victoria,BC, Canada; 2004.

[15] SHUKLA, Diwakar; OJHA, Shweta; JAIN, Saurabh, Data Model Approach And Markov Chain Based Analysis Of Multi-Level Queue Scheduling, Journal of Applied Computer Science & Mathematics;, Issue 8, 2010, Page(s): 50-56.