Modeling and Evaluation of Web Services in Mobile Networks using Stochastic Colored Petri Nets

Homayun Motameni1, Mohammad Vahidpour2, Babak Shirazi3, behnam barzegar4
1) Department of Computer Engineering, Islamic Azad University, Sari Branch, sari, Iran
2) Department of Computer Engineering, University of Science and Technology of Mazandaran, Babol, Iran
3) Department of Computer Engineering, University of Science and Technology of Mazandaran, Babol, Iran
4) Department of Computer Engineering, Islamic Azad University, nowshahr Branch, nowshahr, Iran
motameni@iausari.ac.ir; vahidpour@ustmb.ac.ir; babak@ustmb.ac.ir; barzegar@iauns.ac.ir

Received: 2013/06/14; Accepted: 2013/08/03

Abstract
With recent developments in communication technologies, it is possible to access information any time and in any place; however, mobile environments bring forward certain requirements and exclusive challenges including low connection speeds and bandwidth limitations. Thus, we need to consider mobile limitations when using web services. Providing services and limited resource allocation to users and at the same time maintaining performance and quality is a challenging and somewhat difficult task. In this paper, an architecture is introduced and described for showing the mobile agent request process in mobile environments. A mathematical model is introduced for performance analysis; however, the main challenge is the state space explosion in Markov Chain, which complicates the relations. Hence, a model has been suggested using Stochastic Colored Petri Nets that encompasses two sections. A case study performed on mobile networks and the results of implementation are compared with the proposed method. Therefore, by using the proposed method, it is easy to analyze the effect of network limitations and connection methods on performance.

Keywords: Performance Evaluation, Mobile Environment, Petri Nets, Markov Chain

1. Introduction
With the increasing development of processing technologies, it is possible to access and using information in any time and in any place. With developments in communication technologies, providing services has become possible on mobile devices. Smart phones and web services are very useful and have become an important part of our daily life. Web services are a set of standards and methods, which allow different software to share data and suitable connection. The purpose of web services in mobile services are making an environment that enables communication technology and cell phone industries to produce products and services, which satisfy customer's requirements, whereas common frameworks for web services do not have this ability [1]. Mobile environments present requirements and exclusive challenges. Mobile devices exhibit certain limitations including low processing speeds, low connection speeds, memory limitation, and battery life [2]. Programming methods for mobile
environments are different from common ones. Factors including geographical barriers reduce radio frequencies and increase noise. Thus, error rates will increase and bandwidth will decrease [3]; therefore, it is necessary to consider mobile environment limitations in using web services. Moreover, using incorrect connection models in web services of mobile environments will reduce performance of services and increase communication costs.

In this paper, we will investigate the architecture, process in mobile environment, and mathematical model for analysis of performance. Then performance of the Petri model will be presented. Hence, the proposed method could reduce the difficulty of Markov Chain and calculate the effect of network limitation, connection mode, user migration, and network failure. The purpose of this paper is to evaluate performance in mobile environments, implementation web service provider for evaluation and presentation of a model according to conditions and limitation of mobile environment.

This paper is organized as follows: in section 2, related works will be introduced. In section 3, the architecture of mobile environments will be investigated and we will describe the Markov Chain for mobile networks. In section 4, we will present the results of calculations, and finally, in section 5, calculations and future work will be presented.

2. Related Works

A Mobile network is a digital connection system, which functions by making geographical regions in a cell model and reusing frequency. In mobile communication networks, a cell is the smallest coverage limitation and the coverage of radio cells is cleared by a central transceiver station. All communication networks need special structures for servicing input calls because of changeability and movement of users in mobile networks [4]. There have been different studies on designing and modeling of mobile networks, which all aimed at analyzing the efficiency, strategies and the performance improvement methods of mobile networks based on the mobility of users, the increase in the number of users and the limited sources of the network [5], [6], [7]. According to the proposed quality model in [8] and [9], it can be concluded that the performance indicator affects the operation and the performance evaluation has a significant impact on the performance improvement of the network. The proper operation of the real time applications is very important in the mobile network. However, the challenge raised is difficulty of evaluating a system before implementing it. Xu et al. developed a system interface which could calculate the loss rate, delay rate and predict the network performance by gathering information from the network [10].

The first step in evaluating the performance is network analysis using Markov chain and queuing theory. Researches which used Markov chain and queuing theory to analyze the performance will be examined. Castel et al. use mathematical relations that are based on sudden Markov chain for calculating performance in mobile networks. For complicated networks, a multidimensional Markov Chain is created. They proposed a solution for simplifying and decreasing state space [11]. Saab et al. introduced a model for mobile networks in which users could connect to related services. Then, they introduced Markov chains for evaluating system performance and presented an equation based on the Jackson network relation for calculating possibility of sustainable state [12]. Chen et al. submitted a model for web providers and calculated response time using the Markov Chain. Their response time was dependent on the number of works in queue, input rate, and system interest. Based on results, more users and message volume
cause lower system performance. With decreasing response time, the possibility of error occurrence will increase and the number of users will decrease [13]. Luqun conclude that performance of the system is related to reporter buffer size, input rate, and waiting time in queue; also buffer increase does not affect response time of the system; however it will improve systems reliability and efficiency [14]. It is hard to analyze the performance of complicated systems using traditional methods. Therefore, a three-dimensional model is introduced for network evaluation in [8] and several indicators are defined in each of the dimensions. According to the suggested method for performance evaluation, first the network is analyzed, the requirements and objectives are identified, and then the simulation or evaluation takes place.

The challenge posed in mobile networks is too much delay, high error rate and low bandwidth. So, different strategies are defined to evaluate the performance and improve the operation in [15]. They examined each factor separately but the other challenge is the simultaneous presence of various factors in the network that complicates the problem. Hence, simulation is required to analyze various factors and Petri Nets is a perfect tool for identifying and correcting network errors [16]. In this research, we used the Stochastic Colored Petri Nets which is a combination of two kinds of Stochastic Petri Nets and is a perfect tool for describing and analyzing complex stochastic systems. These networks have some features such as using transitions with immediate or delayed firing by which we can analyze the reliability, performance and error of communication systems [17]. The studies which used Petri net to analyze the performance will be examined. Yu li et al. used Queuing network models and Petri Nets for evaluate the performance of GPRS in the mobile network and study network utilization for different states of the queue [18]. Mokdad et al. used Markov chain to verify that the main problem is the state space explosion, which should decrease. They also introduced a pattern for improving synchronous and asynchronous connections, and then investigated the performance of connections between nodes using stochastic Petri Nets [19]. Blade et al. implemented a web service for mobile environments and calculated the results of service performance, and then investigates different states of user movement in mobile environments and finally presented a Petri model for user behavior. They conclude that response time of mobile users is slightly longer than fixed users [20]. Yin et al. introduced a framework and model for mobile services using generalized stochastic high-level Petri Nets model. They modeled transactions between services and messages flow by introducing and presenting a framework and related regulations to model change of services and message flows. The purpose of paper is preventing deadlock of web flows and ensuring the accuracy of mobile services transmission using modeling [21]. Gharbi investigates the effect of network errors and blocked calls on the network performance using GSPNs models and computational methods and to solve the problem of state space explosion, he suggested an algorithm which automatically calculates Markov chain without taking the accessibility graph into account [22].

Since mobile agents are moving in the communication networks, the handover performance in next generation networks is an important challenge that in [23] the handover time is calculated based on the existing standards. The other challenge is the effect of the number of mobile agent on the network performance which we will explain in this article. Another challenge is the examination and evaluation of non-functional requirements in Petri Nets and Markov chains. We can calculate the security, reliability, availability and the performance using the relations in [24] and [25].
In short, the state space explosion in the Markov chain is a major challenge and many different studies has been done in this field to solve the problem but we need large storage space and longer time to perform and analyze. Hence, it is suggested to use Stochastic Colored Petri Nets. Reviewing the related articles revealed that several factors affect the network performance and the simultaneous presence of different factors complicates the problem more and more. In this study, we examine the effect of the simultaneous presence of different factors on the performance.

3. Architecture Modeling of Mobile Network

For achieving suitable performance, we need to analyze mobile environment complexity. Hence, a kind of architecture should be stated that reflects the communication cells behavior, Web services and users' requests.

A communication mobile service is composed of a set of cells, each cell covering a geographical limitation to satisfy user requests. We suppose that channel capacity is limited. There is a station and users can enter a cell, using network and exit from cell at any time. There is a channel of network for every user who wants to use the network. In this situation, we consider an agent that moves from one cell to another. In this way, the number of users will decrease and if the channel capacity is not complete, the number of users will be added to other cell, which is shown in Figure 1.

![Figure 1: An architecture of Markov Chain in communication cell](image)

The mobile agent can move freely between communication cells and they will be registered in that network after entrance to the cell. Every request will be sent to the web provider after passing through the network queue and a response will be returned to users via the communication network. The request process of mobile agent is shown according to time in Figure 2, which is moving between cells.
Figure 2: Mobile agent request process in communication network

$t_m =$ time duration of mobile agent migration
$t_r =$ response time of successful register in a cell
$t_{qc} =$ delay in cell queue
$t_c =$ delay of Communication link
$t_{qs} =$ delay in service queue
$t_p =$ delay of processing for handling of protocol, encoding and decoding
$t_s =$ response time of service

The total response time which is described in equation 1 is equal to the total of delays, waiting time in queue and services response

$$T = \sum (t_m + t_r + 2(t_{qc} + t_c) + t_{qs} + t_p + t_s)$$

(1)

In Equation 1, delay in queue and network is multiplied by number 2 in Equation 1, because of packages sweep. Usually communication cells cover all part of the city and we can discard the weak or lack of coverage among two cells; thus the roaming time of mobile agent among two adjacent cells can be considered as zero ($t_m = 0$). In any cell, response time of mobile agents is a little further than static agents [20]. Hwan et al. specified four states for the movement mobile agents, and response time is calculated for each one. Usually, the movements of agents between cells are randomly; According to results, the respond time in accidental movement can be discarded [26]. In the following sections, we will investigate response time in mobile environments.

3.1 Mobile Network Modeling using Markov Chain

Network performance is the ability of providing services to mobile users during connection. In order to achieve these benefits, design and analysis of network architecture is necessary. For analyzing performance, a mathematical model is presented, which is based on architecture described in section 4. In this model, each user can have move freely in the communication network. The network is composed of some communication cells which are shown with variable $C$, and each cell has limited capacity designated by variable $Z$. A channel will be dedicated to every user who requests network usage, in case of remaining capacity. After usage termination, the
dedicated capacity will be freed. We consider \( N \) number of web service providers. We suppose that in cell of \( i \), according to Poisson’s law, users are entered with the input rate of \( \lambda_{ci} \) where \( 1 \leq i \leq C \). Response time in any cell of \( i \) has nominal distribution with the rate of \( \mu_{ci} \).

The described model can be displayed using Continuous Time Markov Chain, which is shown as \( X(t) \). The most important benefit of Continuous Time Markov Chain is its simplicity, because it shows the complex condition of systems. Moreover these models can be used for measuring system performance. For these two reasons, Continuous Time Markov Chain could be a suitable option. For describing this chain, \( X \) variable will described as \( (x_{c1}, x_{c2}, ..., x_{cp}, x_{s1}, x_{s2}, ..., x_{sN}) \) in which \( x_{ci} \) indicates the number of requests in cell \( i \) that \( (\forall 0 \leq c_i \leq C) \) and \( x_{si} \) indicates the number of requests to web service providers that \( (\forall 0 \leq s_i \leq N) \). \( \lambda_{ci} \) is equal to user migration rate from one cell to another, and \( \lambda_{si} \) is equal to transfer rate to other web services. For different events, we have the following transfers:

\[
\begin{align*}
\text{(1)} & \quad x(1) \rightarrow (..., x_{ci} + 1, ...) \text{ if } (x_{ci} \leq Z_{ci}) \text{ with rate } \lambda_i \\
\text{(2)} & \quad x(2) \rightarrow (..., x_{ci} - 1, x_{cj} + 1, ...) \text{ if } (x_{cj} \leq Z_{cj} \text{ and } x_{ci} > 0) \text{ with rate } x_{ci} \cdot \lambda_{cj} \\
\text{(3)} & \quad x(3) \rightarrow (..., x_{ci} - 1, x_{sj} + 1, ...) \text{ if } (x_{ci} > 0) \text{ with rate } x_{ci} \cdot \mu_{ci} \\
\text{(4)} & \quad x(4) \rightarrow (..., x_{si} - 1, ...) \text{ if } (x_{si} > 0) \text{ with rate } x_{si} \cdot \mu_{si} \\
\text{(5)} & \quad x(5) \rightarrow (..., x_{si} - 1, x_{sj} + 1, ...) \text{ if } (x_{si} > 0) \text{ with rate } x_{si} \cdot \lambda_{sj}
\end{align*}
\]

In the stated equation, different lines are shown:

1) In cell \( i \) \( (0 \leq i \leq C) \), the service request enters with rate of \( \lambda_i \) and a channel will be dedicated to user, if cell capacity is not completed;
2) A user on cell \( i \) will change his position and move to cell \( j \). The transfer will be performed when the capacity of the target cell is not complete;
3) Request for using web services from the queue of cell \( i \) will be exited and will be sent toward the web service;
4) The web service that is placed on node \( i \) will respond to one of the requests with the rate of \( \mu_{si} \), and the result will be returned to the user;
5) The sent request need other web services for completion and the request will be sent to them. \( \lambda_m \) indicates the total users migration rate plus the rate of withdrawal or cancellation of the services by users.

Based on equation 2, the diagram of transfer state of Markov chain is indicated in Figure 3. Dotted lines indicate variable number of states.
More users and hosts will increase the size of the Markov chain. The size of the Markov chain for $M$ users and $N$ web services is equal to $(M + 1)^{N+1}$; thus, analysis and calculation of performance in large systems is very complicated, and we could use modeling in this way.

3.2 Modeling using Stochastic Colored Petri Net

Prediction of performance in service oriented applications is a complex task; thus, Petri Nets are used for calculating performance; the mobile Agents processes are modeled, models are transited to equivalent Petri Nets and the desired net will be evaluated per different states.

Colored Petri Nets have properties such as color and time. Stochastic Petri Nets have properties such as immediate firing, deterministic delay, fault tolerance, and the possibility for analysis of network performance. Stochastic Colored Petri Nets is a combination of two kinds of Petri Nets which uses both types of them and it is a suitable tool for description and analysis of complex stochastic systems. The firing delay in transitions is defined by the random variable with the exponential distribution $\lambda$ and the firing delay is randomly assigned to each transition. The formal definition of Stochastic Colored Petri Net is $(P, T, I, O, H, W, M_0)$, in which $P$ is the set of places, $T$ is the set of transitions, $I$ is the set of inputs, $O$ is the set of outputs, $H$ is the inhibition functions, $W$ is the delay to each transition and $M_0$ is the initial marking function which is defined on $P$. In this study we used the TimeNet software version 4.1 that is a useful tool for analysis of SCPN Petri Nets which allow error control and modeling complex component [17]. In modeling, the confidence level is 5%, and confidence interval is considered 95%.

3.3 Modeling Mobile Networks using Petri Net

The Performance model depicted in Figure 4 is composed of three sections including mobile users, mobile network, and web service. The queue of mobile network is M/M/m/K, and web service queue is M/M/1. The input state in network and web service is FIFO. Users will enter the network by distribution rate of Poisson $\lambda_{ci}$ and the time of service in the mobile network is exponential distribution with a rate of $\mu_{ci}$. If a user does not change its position in node $i$, the request will be delivered to the related service with the probability of $P_{ij}$, and the user can change its position to another cell with the

![Diagram of transfer state of Markov chain](image-url)
probability of $P_h$, and due to failure indication in network, some requests to the related service will not be delivered with the probability of $P_e$. Service rate of any web services in node $j$ ($1 \leq j \leq N$) is exponential distribution with a rate of $\mu_{sj}$. Results will be completed and delivered to users with the probability of $P_{j0}$. However, if necessary, for sending requests to other web services, it is shown by probability of $P_{sjk}$. Some users have services request again or request more information with probability of $P_r$.

![Figure 4: Mobile environments performance model](image)

In the following section, based on the hypothesis and described requirements, the proposed model is analyzed using Petri Net.

![Figure 5: Proposed model for evaluation performance of mobile environments](image)

Figure 5 depicts the proposed model for the mobile environment, which is mapped using Petri Net. In this model, users send their request to the mobile network, and then requests are routed and delivered to the desired web service. Finally, responses will be returned to the users; however, some users send request again or request more information. Used tokens are described in Figure 5.
In Figure 6, there is a model for entering mobile users, in which the request will enter by exponential distribution $\lambda_{ci}$, and then requests will be sent by the mobile network to related services. Variable $C$ shows package counter and variable $T$ is shows input time. In mobile networks, users can migrate from neighbor cell to another cell with the rate of $\lambda_{Migration}$. Finally, server response will be delivered to the user and will be displayed on the mobile device. In situation, it is possible that after displaying service on mobile device, Services were reconnected to the server again which we show by Poisson distribution $\lambda_{Extra}$. Packet arrival rate is equal to the arrival rate of the new users' to the network and also the users' possible migration rate and the probability of the additional requests which is stated in equation 3.

$$\lambda_{in\ (Network)} = \lambda_{ci} + \lambda_{Extra}P_r + \lambda_{Migration}P_h \tag{3}$$

Figure 7 is modeled based on introduced architecture in section 3. The model is composed of three sections: receiving package line (receiving package from user and sending it to the router), sending package line (receiving package from the router and
sending it to the user), and network controlling (network capacity, failure controlling, and user migration). In receiving line, the package will enter the sub-system of the mobile network. Network capacity is limited; thus there is a channel for every package. Token \( m \) shows the number of channels in the network. The packets will stand in the queue of network. The structure of network queue is FIFO. Queue Length will increase after rise of net traffic and packets leave the net with exponential distribution rate of \( \mu_n \). The request for server may not be delivered because of failure. In this case, the token enters failure location with the rate of \( \lambda_{\text{Failure}} \) and frees the closed channel. If a user changes its position, then the data package will exit the communication cell, and the captured channel will become free. Synchronous data communication is similar to TCP and asynchronous method is similar to UDP. In a synchronous connection, capture capacity of the network will be maintained until the package reaches the consumer. However, in an asynchronous connection, capture capacity will become free upon package exit from the network. Variable \( S \) is used for recognizing the kind of connection (number one for synchronous communication or number two for synchronous communication), and finally the package will be sent to the router. The route of sending package is similar to the route of receiving package, but the difference is that in the network, type of connection is identified and if the connection is asynchronous, a channel will be dedicated to packet again. If user migration is taking place or there is a failure in the network, the package will exit the network; otherwise, the package will be placed in the network queue, and finally, the captured capacity in the network will be free before package exit and delivery to the user. The relation between the input and output rate to the network and the router is shown in relation 4.

\[
\lambda_{\text{in (Routers)}} = \lambda_{\text{in (Network)}} - (\lambda_{\text{Failure}}P_e + \lambda_{\text{Migration}}P_n) \\
\lambda_{\text{out (Network)}} = \lambda_{\text{out (Routers)}} - (\lambda_{\text{Failure}}P_e + \lambda_{\text{Migration}}P_n)
\]

![Figure 8: Routers sub-system modeling](image)

The router sub-system in Figure 8 is composed of routers, the route of connecting network to the server (Line1) and server route to the network (Line2). Token \( r \) shows the numbers of routers. Router delays are added to tokens upon exit. The packets with the exponential distribution rate \( \mu_r \) go out of the router's queue. Hence, the total time spent on the router is \( \frac{1}{\mu_r} \)
The web service Sub-system in Figure 9 is composed of a number of servers. Token \( s \) shows the number of tokens. The entered package will be placed in the server queue. The structure of the server queue is FIFO. The package will exit the queue with the nominal distribution \( \mu_{s_j} \), and response will be sent to mobile users with the probability \( P_{f0} \), and with the Probability \( P_r \) need to complete the application by self-server. Nonetheless, if sending request to other services is required, the request will be sent to another server with the probability \( P_{sjk} \) and other servers respond by the exponential distribution rate \( \mu_{sk} \). Finally, the packages will be sent to the routers after completing the request. If the response time of the two servers in a ring way are \( t_1 \) and \( t_2 \) in a way that the probability of \( t_2 \) is \( \beta = 1 - \alpha \), so the total response time is \( (t_1 + at_2)/\beta \) [27]. Hence, the response time of the web service Sub-system which consists of two web providers in a ring with the probability of \( Ps_{jk} \) and the reapply route with the probability of \( P_r \) is the equation 5.

\[
T_{Servers} = \frac{t_{sj} + t_{sk} P_{sjk} + T_{Retry} P_r}{P_{f0}} \tag{5}
\]

\( T_{Retry} \) is the total time that a packet spends in the network and the router.

\[
T_{Retry} = T_{Network} + T_{Routers} = \frac{2}{\mu_{n}} + \frac{2}{\mu_{r}} \tag{6}
\]

The final response time is equal to the time spent in the network, routers, and servers which is shown in the equation 7.

\[
T_{Total} = T_{Servers} + T_{Network} + T_{Routers} = \frac{\mu_{sk} + \mu_{s} P_{sjk}}{\mu_{s_j} \mu_{sk} P_{f0}} + \frac{2(\mu_{n} + \mu_{r})(P_{r} + P_{o})}{\mu_{n} \mu_{r} P_{o}} \tag{7}
\]

4. Calculations

For evaluating performance, web services were designed in Java with SOAP protocol, and were implemented on a Windows 7 server. The client program on the mobile device was written using NetBeans and J2ME language. Different mobile users
connected to the server via the mobile network, and response times were calculated for simultaneous users.

Table 1: Input variable to model

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Mobiles</td>
<td>1-100</td>
</tr>
<tr>
<td>Number of Servers</td>
<td>1</td>
</tr>
<tr>
<td>Number of Router</td>
<td>1</td>
</tr>
<tr>
<td>Capacity of Network</td>
<td>Extreme</td>
</tr>
<tr>
<td>Capacity of Queue (Network &amp; Server)</td>
<td>Extreme</td>
</tr>
<tr>
<td>Delay of Network</td>
<td>256 ms</td>
</tr>
<tr>
<td>Server Service Time</td>
<td>14 ms</td>
</tr>
</tbody>
</table>

Figure 10: Comparison SOAP web service response time between using implementation and Petri Net

Described variables in Table 1 will be transferred to the simulation and obtained results are shown in Figure 10. The horizontal axis shows the number of users, and the vertical axis shows response time. Figure 10 explains that a Petri Net result is close to implementation. In order to decrease response time, it is advised that using more services will decrease load traffic on other servers.

Figure 11: Comparison response time for different mobile network capacity
In Figure 11, the time of simulation is shown for capacity of 1, 5, 10, and extreme. The more limited is the capacity of the network, the more are the packets waiting to be sent. The response time will increase by the increase of the network queue length. In simulation the response rate tends to a fixed rate by the increase of the network capacity which is shown in the equation 8. Hence, service providers should provision sufficient network capacity for consumers based on the number of users in one region.

\[ \lim_{\text{channel} \to \infty} T_{\text{network}} = k \quad (k \text{ is the fixed Rate}) \quad (8) \]

5. Conclusion

Today, providing services in mobile environments is a challenge that many researches are done around it. Any disturbance in the system has serious economic consequences, since for analysis of mobile network performance, Markov chain state space for \( M \) users and \( N \) services is equal to \((M + 1)^{N+1}\), which makes relations more complicated, so it is proposed that Petri Nets use for analyzing performance.

In this paper, we present a new method for modeling and analyzing performance in mobile environment using Stochastic Colored Petri Net, which is designed based on continuous time of Markov chain to easily and quickly calculate the result of performance.

6. Reference


