A NEW FRAMEWORK FOR PREDICTING THE IMPACT OF TRAFFIC ON THE PERFORMANCE OF MOBILE AD-HOC NETWORK (MANET): USING REGRESSION AS DATA MINING APPROACH

Kamal Moh’d Alhendawi
Al-Quds Open University, Gaza, Palestine
e-mail: alhendawi.kamal@gmail.com or hindawi.kamal@yahoo.com

ABSTRACT

With the rapid technological advances in wireless communication and the increasing of usage of portable computing devices, it is expected that mobile ad hoc networks are increasingly developed towards enhancing the flexibility, scalability and efficiency of communication technology. The wireless ad-hoc network is a collection of mobile nodes in which these nodes have the ability to connect each other without backbone infrastructure (i.e. infrastructure less). Although many studies have been done on the performance assessment of MANET routing protocols, there is a need for investigating the impact of traffic load on the performance of MANET in order to justify the use of some routing protocol in MANET. This study is one of the fewest that aims at proposing a new framework towards predicting and validating the results of future scenario using data mining techniques. The regression analysis is used as data mining method in the prediction of the future scenarios. Practically, two experiments with eight scenarios are conducted. The findings indicate that the network size and traffic loads are proportionally related to the throughput. However, the findings show that the network size is inversely related to the delay in case of medium FTP traffic, and proportionally related in case of high FTP traffic. The results also indicate that data mining approach specially regression is an effective approach towards the prediction of the future network behavior.

Keywords: MANET; Mobility; Ad hoc network; Prediction model; Regression analysis; Traffic; Delay; Performance metrics.

INTRODUCTION

Basically, MANETs stands for Mobile ad hoc networks where MANET consists of a set of stations called nodes, and these nodes can communicate to each other through wireless links (Cheng et al., 2013, Chaki & Chaki, 2014). In ad-hoc network a route between two mobile hosts may include of hops through one or more nodes in the MANET. Furthermore, Cheng et al.(2013) indicated that MANETs can be realized by different wireless communication technologies such as Bluetooth, IEEE 802.11, and Ultra-Wide Band (UWB). Figure 1 illustrates the MANET topology, where circle denotes the transmission range, blue line denotes the wireless links and the mobile represents the mobile node.
According to Murty & Das (2011), the routing protocols of the ad-hoc network can be divided into two types: on demand protocols and proactive protocols. On demand routing protocol explore the route update when the route is required. However, the proactive routing protocols periodically update the routing information in order to disseminate topology updates on the network. Most of previous studies indicated the importance of network transmission metrics including traffic, delay and throughput in the evaluation of the network performance (Saad and Zukarnain, 2009; Rohankar et al., 2012; Cheng et al., 2013; Murty & Das, 2011). With regard to the performance metrics, the most commonly used measures are network Packet Delivery Ratio, delay, and throughput (Rohankar et al., 2012; Shah et al., 2008).

Since the topology of ad-hoc network has dynamic changes, investigating the impact of traffic on the performance of the network becomes increasingly required (Murty & Das, 2011; Cheng et al., 2013). Additionally, the previous studies discussed the impact of traffic on the network performance (Murty & Das, 2011; Saad and Zukarnain, 2009), they did not provide a flexible mathematical mechanism to minimize the huge number of the required experiments and scenarios. Furthermore, the published researches did not provide a comprehensive method to validate the accuracy of the obtained findings. Thus, the findings that are extracted after conducting simulation might be biased or give inaccurate implications.

For these reasons, this study is one of the fewest that focuses on simulating several scenarios in an effort to investigate the influence of traffic on ad-hoc network performance, and also, minimizes the scenarios through predicting the results of scenarios using data mining techniques, specially, the regression analysis. This study also provides a new theoretical framework for evaluating, predicting and validating the performance metrics (i.e. delay and throughput).

**Theoretical Foundations and Related Work**
As the study discusses the impact of traffic on performance metrics including delay time and throughput, it seems that it is appropriate to divide this section into two sub sections
in order to cover the concepts and terminologies of performance factors in addition to the essential information concerning the environment of wireless mobility models.

**Performance Metrics**
Currently, evaluating the performance of mobile ad hoc network (MANET) is considered as an active area of research and it is still under discussion because many parameters factors could affect the performance of MANET. The primary focus of this study is to evaluate and validate the influence of traffic on the performance of MANET. To decide on the performance, it is essential to use standard metrics such as Average routing overhead, Packet delivery ratio, end-to-end delay, and throughput (Mutry & Das, 2011; Sanjay et al., 2009).

- **Packet Delivery Ratio (PDR)**
The number of data packets sent from the source to the number of received at the destination.

\[
PDR = \frac{\text{control packets sent} - \text{delivery packet sent}}{\text{control packets sent}}
\]

- **Average routing head (ARH)**
Ad hoc networks are intended to be scalable. As the network grows, a variety of routing protocols perform differently. The amount of routing traffic increases as the network grows. Therefore, it is considered as a measure of the scalability of the protocol, and thus the network. Average routing overhead is the total number of routing packets divided by total number of delivered data packets expressed in bits per second or packets per second.

\[
\text{ARH} = \frac{\text{Total no of routing packets}}{\text{Total no of delivered data packets}}
\]

- **Average End-to-End Delay (AEED):**
It consists of all the delays in the network such as buffer queues, transmission time and delays induced by routing activities and MAC control exchanges. Average End-to-End delay (seconds) is the average time it takes a data packet to reach the destination.

\[
\text{AEED} = \frac{\sum_{k=0}^{n} (\text{time packet received} - \text{time packet sent})}{\text{total no of packets received}}
\]

- **Throughput:**
The throughput is the rate of successfully transmitted data per second in the network during the simulation.
Mobility Models
The mobility model is designed to illustrate the movement pattern of mobile users, and how their location, velocity and acceleration vary over time. Since mobility patterns may play an important role in specifying the protocol performance, it is desirable for mobility models to emulate the movement pattern of targeted real life Applications in a reasonable way. The details of three popular mobility models are given below in order to demonstrate how the mobile nodes emulate the real world scenario (Rohankar et al., 2012):

**Random Waypoint**
Divecha et al. (2007) pointed out that the Random Waypoint model is the most popular model used in mobility research community. At every instant, a node randomly chooses a destination and moves towards it With a velocity selected randomly from a uniform distribution \([0, V_{\text{max}}]\), where \(V_{\text{max}}\) is the maximum allowable velocity for every mobile node. The node stops for a period of time identified by the ‘pause time’ parameter after reaching the destination. It also considered as the basis for establishing the other mobility models (Quispe & Galan, 2014; Pazand et al., 2007). Figure 2 shows the movement of mobile node.

![Random Waypoint Mobility Model](image)

**Figure 2. Random Waypoint Mobility Model**

**Random Walk Mobility Model**
The random walk mobility model is considered as the simplest mobility model in term of generating pattern. It is established to conduct simulations in case of the movement pattern of the mobile nodes is unpredictable. For the random walk mobility model, the mobile node randomly select the direction and travel speed whenever it moves from its current location to the new location, and the random speed between \([\text{SpeedMin}, \text{SpeedMax}]\) (Quispe & Galan, 2014). Figure 3 reveals the movement of mobile node in the random walk mobility model.
A new framework for predicting the impact of traffic on the performance of mobile ad-hoc network (MANET): using regression as data mining approach

**Random Direction Mobility Model**
In this mobility model, the mobile node utilizes the Random Waypoint Mobility Model in choosing the new destinations. There is a high probability for selecting a new destination located in the center of the simulation area, or requires travel through the middle of the simulation area. Figure 4 illustrates the trend of random direction mobility model (Pazand et al., 2007).

As an alternative classification, the mobility models can be differentiated based on different kinds of dependencies and restriction that are considered as shown in Table 1.

<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random Models</td>
<td>There are neither dependencies nor any other restrictions modeled which are similar to RWP model.</td>
</tr>
<tr>
<td>Model with Temporal Dependency</td>
<td>The mobile nodes tend to travel in a correlated manner</td>
</tr>
<tr>
<td>Model with Spatial Dependency</td>
<td>The actual movement of a node is influenced by the nodes around it.</td>
</tr>
<tr>
<td>Models with Geographic Restriction</td>
<td>The area in which the node is allowed to move is restricted.</td>
</tr>
</tbody>
</table>
Regression Data mining Approach
Based on literature review (Ahmed, 2004; Giraud-Carrier and Povel, 2003; Turban et al., 2007; Ngai et al., 2009), there are several data mining model that could be provided by data mining techniques where classification, clustering, regressions and forecasting are examples on the data mining models. In statistical modelling, regression analysis is a statistical process for approximating the relationships between variables including independent and dependent variables where the independent variables have the role of predictors (Field, 2009). Accordingly, regression analysis is broadly used for prediction. Regression analysis is also employed in recognizing how independent variables are related to the dependent variable, and to explore function which models the data with the least error (Field, 2009).

Figure 5. Illustration of regression on a data set

Figure 5 shows the regression on a data set depending on the function: \( Y = f(X, \beta) \); where the regression model relates \( Y \) to a function of \( X \) and \( \beta \).

RELATED WORK
Hinds et al. (2013) conducted a comparison between the routing protocols in MANET in order to determine the optimal protocol that can operate on MANET. Shah et al. (2008) evaluate the performance of ad-hoc routing protocols using simulation. Network load, mobility, and network size are used as parameters for the simulation.

Cheng et al. (2013) focused in his study on the basic MANET topology and algorithms with special attention to the routing protocols used in the Video streaming Network (VANET) including single and multipath routing protocols.

Saad and Zukarnain (2009) investigated the impact of mobility model on the performance of AODV for overhead throughput and Packet delivery ratio parameters for number of nodes. Moreover, Mutry and Das (2011) discuss the performance of MANET with special focus on the efficiency of on demand routing protocols DSR and AODV based on packet delivery ratio, mobility, packet delivery latency, variation with total number of errors, end-to-end delay where the node density is varying.
Rohankar et al. (2012) analyzed the performance of the routing protocols for random mobility in the mobile ad-hoc network. Performance is analyzed in terms of end-to-end delay, throughput and Packet delivery ratio for varying node densities.

Based on literature review, the simulation studies mostly used the random waypoint as a model of mobility (Quispe & Galan, 2014; Rohankar et al., 2012). Further, the most wireless communication technology is Bluetooth or IEEE 802.11. To provide accurate results, the simulation setting should consider encouraging results and recommendation of the previous related work in selecting the mobility model and network topology. More details regarding the simulation settings are provided in section 5.

**PROPOSED FRAMEWORK AND METHODOLOGY**

To evaluate the impact of traffic on the performance of ad-hoc network, two experiments with six scenarios are performed where each experiment considers the variety of the traffic within a fixed number of mobile stations (i.e. Traffic FTP). The first experiment assumes medium FTP traffic while FTP traffic in the second experiment is supposed to be high. Figure 6 reveals the proposed framework which demonstrates the methodological steps followed by the research toward evaluating the effects of traffic on performance.
Figure 6. The proposed Framework for predicting the performance of Mobile Ad-hoc Net
The first step is to decide regarding the design of experiments and scenarios. In the
second step the simulation parameters should be identified (see next section). In the third
step, the simulation is conducted, and therefore, the simulation results are extracted. In
the fourth and fifth steps, the data mining method namely, regression analysis is used in
order to estimate the results of the future scenarios. In the sixth step, the results of
regression analysis are extracted. Finally, a set of simulation results are compared with
the prediction results to measure the effectiveness of using data mining in the prediction
(regression analysis).

Simulation assumptions
This study is mainly concerned with three parameters such as traffic sent, delay and
throughput. Moreover, two experiments are actually used with different number of nodes
(10, 20 and 30 respectively). Table 2 shows the parameters values that should be input as
a simulation’s settings or configurations.

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simulator</td>
<td>OPNET 14.5</td>
</tr>
<tr>
<td>Routing Protocols</td>
<td>AODV</td>
</tr>
<tr>
<td>802.11</td>
<td>data rate 11 Mbps</td>
</tr>
<tr>
<td>Node</td>
<td>10, 20, 30</td>
</tr>
<tr>
<td>Scenario Size</td>
<td>3.5*3.5 km</td>
</tr>
<tr>
<td>Application Traffic</td>
<td>FTP</td>
</tr>
<tr>
<td>Simulation Time</td>
<td>60 min</td>
</tr>
<tr>
<td>Channel Type</td>
<td>Wireless channel</td>
</tr>
<tr>
<td>Network Interface Type</td>
<td>Phy/WirelessPhy</td>
</tr>
<tr>
<td>Performance Parameter</td>
<td>Throughput, Delay, Network Load</td>
</tr>
<tr>
<td>Model</td>
<td>Random waypoint</td>
</tr>
</tbody>
</table>

EXPERIMENTAL RESULTS

The simulations consist of two experiments with 6 scenarios and they concerned with
investigating the impact of traffic load on the performance of mobile ad-hoc network.
The two experiments use the same protocol namely, AODV. The simulations focus on
two performance metrics including delay and throughput.

Experiment #1
In experiment 1, the simulation assumes that all nodes have a medium FTP traffic. It
consists of three scenarios in accordance with number of mobile nodes (10, 20, and 30).
Concerning the delay time, it is found that the delay time decreased as number of nodes
increased. Figure 7 shows the impact of medium FTP traffic on the delay.
A new framework for predicting the impact of traffic on the performance of mobile ad-hoc network (manet): using regression as data mining approach

Figure 7. The effect of medium FTP load on delay time

Regarding the throughput, the findings show that throughput deceased as number of nodes increased. Figure 8 reveals the impact of medium FTP traffic on the throughput.

Figure 8. The effect of medium FTP load on throughput

Experiment #2

In experiment 1, the simulation assumes that all nodes have a High FTP traffic. It consists of three scenarios in accordance with number of mobile nodes (10, 20, and 30). The findings indicated that the delay time decreased as number of nodes increased. Figure 9 shows the impact of high FTP traffic on the delay time.

Figure 9. The effect of medium FTP load on delay time
Based on figure 10, the findings demonstrated that the throughput decreased as number of nodes increased.

![Impact of Traffic Load on Throughput](image)

**Figure 10. The effect of high FTP load on throughput**

**PREDICTING THROUGHPUT OF MANET (USING DATA MINING)**

As the simulation requires too many scenarios in order to consider the results, it is important to find an alternative data mining approach for predicting the future scenarios results. One of the methods that could be used is the *regression analysis*. In this study, the author keeps focus on throughput prediction.

To do so, the mean of scenarios for 10, 20, and 30 nodes are calculated. Secondly, the regression function is utilized toward predicting forecasting the relationship between the FTP traffic and the performance metrics (i.e. throughput) in order to predict the value of throughput. Therefore, it is important to build the mathematical equation of regression function and also it is essential to find the regression coefficients. As this study concerned with the effect of independent variable (x: Network size) on the dependent (y: Throughput), it seems that the linear regression is most appropriate tool in prediction which is expressed in the following equation:

\[
Y = a + bX,
\]

where \(X\) is the explanatory variable and, \(Y\) is the dependent variable.

So, the obtained value is actually an estimate for the average of the throughput, and therefore, the accuracy of estimation is reasonable. Figure 11 shows the mechanism of prediction for the throughput using regression analysis as a data mining technique.
A new framework for predicting the impact of traffic on the performance of mobile ad-hoc network (manet): using regression as data mining approach

In the following two sub section the researcher followed the proposed methodology in Figure 11 in order to extract and validate the prediction of the throughput values.

**Experiment # 1**

Based on Table 3, the mean values of the delay along with number of nodes for the first experiment is as follows:

<table>
<thead>
<tr>
<th>Nodes</th>
<th>Throughput(bps)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>0.000667572</td>
</tr>
<tr>
<td>20</td>
<td>0.000541831</td>
</tr>
<tr>
<td>30</td>
<td>0.000364902</td>
</tr>
</tbody>
</table>

After finding the actual throughput value, these values should be represented through linear regression in order to find the regression coefficients and predict the value of the throughput. Figure 12 illustrates the linear regression of the throughput values.
The result of regression analysis provides a regression equation \( y = -2E-05x + 0.0008 \) and this equation is used to predict the values of the throughput for the nodes from 31 to 40.

Table 4. The predicted values of throughput (Exp.1)

<table>
<thead>
<tr>
<th>Nodes</th>
<th>Predicted Throughput</th>
</tr>
</thead>
<tbody>
<tr>
<td>31</td>
<td>0.00018</td>
</tr>
<tr>
<td>32</td>
<td>0.00016</td>
</tr>
<tr>
<td>33</td>
<td>0.00014</td>
</tr>
<tr>
<td>34</td>
<td>0.00012</td>
</tr>
<tr>
<td>35</td>
<td>1E-04</td>
</tr>
<tr>
<td>36</td>
<td>0.00008</td>
</tr>
<tr>
<td>37</td>
<td>6E-05</td>
</tr>
<tr>
<td>38</td>
<td>0.00004</td>
</tr>
<tr>
<td>39</td>
<td>2E-05</td>
</tr>
<tr>
<td>40</td>
<td>0</td>
</tr>
</tbody>
</table>

As important point Table 4 shows that the prediction stop at 40 because after this number the throughput becomes zero as the network traffic load will be very high. Figure 13 represent the predicted values throughput in comparison with the linear regression.
A new framework for predicting the impact of traffic on the performance of mobile ad-hoc network (manet): using regression as data mining approach

Figure 13. The throughput predicted values VS the actual values

Nothing can flow an exact linear pattern even in the time series or not. The blue line, which reveals the distribution of the predicted values, shows that the predicted values are closed to the regression line and it takes the oval shape. Also, the residual within the acceptable range and has a reasonable level of linearity (Field, 2009). Thus, the linear regression is successfully predicting the linear relationship.

Experiment # 2
Table 5 demonstrates the mean values of the throughput along with number of nodes as follows:

<table>
<thead>
<tr>
<th>Nodes</th>
<th>Throughput(bps)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>0.0013218</td>
</tr>
<tr>
<td>20</td>
<td>0.000781323</td>
</tr>
<tr>
<td>30</td>
<td>0.000581088</td>
</tr>
</tbody>
</table>

After finding the actual throughput value, these values should be represented through linear regression in order to find the regression coefficients and predict the value of the throughput. Figure 14 illustrates the linear regression of the throughput values.
The result of regression analysis provides a regression equation \( y = -4E-05x + 0.0016 \) and this equation is used to predict the values of the throughput for the nodes from 31 to 40.

Table 6. The predicted values of the throughput

<table>
<thead>
<tr>
<th>Nodes</th>
<th>Predicted Throughput</th>
</tr>
</thead>
<tbody>
<tr>
<td>31</td>
<td>0.00036</td>
</tr>
<tr>
<td>32</td>
<td>0.00032</td>
</tr>
<tr>
<td>33</td>
<td>0.00028</td>
</tr>
<tr>
<td>34</td>
<td>0.00024</td>
</tr>
<tr>
<td>35</td>
<td>0.0002</td>
</tr>
<tr>
<td>36</td>
<td>0.00016</td>
</tr>
<tr>
<td>37</td>
<td>0.00012</td>
</tr>
<tr>
<td>38</td>
<td>0.00008</td>
</tr>
<tr>
<td>39</td>
<td>4E-05</td>
</tr>
<tr>
<td>40</td>
<td>0</td>
</tr>
</tbody>
</table>

As important point the prediction stop at 40 because after this number the throughput becomes zero as the network traffic load will be very high. Figure 15 represent the predicted values throughput in comparison with the linear regression.
A new framework for predicting the impact of traffic on the performance of mobile ad-hoc network (MANET): using regression as data mining approach

Figure 15. The throughput predicted values VS actual values

The blue line shows that the predicted values have an acceptable range of residual as the predicted values are closed to the regression line.

**Comparing the two experiments**
In an effort to compare the results in the two experiments, Figure 16 illustrates the difference in the throughput based on the variety of the FTP traffic.

Figure 16. The comparison of the throughput for experiments 1&2

Based on Figure 16, it is clear that the throughput has a proportional relationship with the network size. Also, it is indicated that the throughput reduction in case of experiment 1 has slow trend than that of experiment 2. This is because of the high traffic in experiment 1. Generally, the trend of the predicted values in the two regression models is approximately similar.

**DISCUSSION**

Based on the simulations results, it can be inferred that the network size has reverse relationships with throughput and delay in case of medium traffic, and this makes sense as the network has many mobile stations and there is no congestion. Even this result has
a logic, sense and systematic justification, it doesn’t support the findings obtained by Pal and dhir (2013).

However, in case of high FTP traffic, the network size has a proportional relationship with the delay, and this seems to be correct as the network might have high congestion or collisions. This result is in agreement with many related studies such as (Murty & Das, 2011; Cheng et al., 2013; Chaki & Chaki, 2014).

Concerning experiment 1, the results indicates that ($R^2 = 99.06\%$) which means the independent variable (i.e. no of nodes) highly contribute to the MANET throughput as number of nodes explains 93.4% of total variances in the throughput. Additionally, based on Figure 13, the residuals are within the acceptable range. Consequently, the prediction model represented by the equation ($Y= -2E-05x + 0.0008$) has a reasonable contribution in the prediction of the results of the future scenarios in case of having medium FTP load.

With regard to experiment 2, the results shows the ($R^2 = 0.9343$) which means the independent variable (i.e. no of nodes) highly contribute to the MANET throughput as number of nodes explains 93.4% of total variances in the throughput. Moreover, based on Figure 15, the residuals are within the acceptable range. Therefore, the prediction model ($Y = -4E-05x + 0.0016$) has a reasonable contribution in the prediction of the future performance (throughput) in case of having high FTP load.

In addition, the similarity of the trends of the regression models in the two experiments provides evidence on the consistency of the prediction models. The results also emphasize the relationship between the network size and throughput, as the previous studies were also indicated the proportional relationship.

As evidence on the suitability of linear regression in predicting the relationship between network load and throughput, the predicted values are closed to the regression line and it takes the oval shape. The residuals within the acceptable range and have reasonable level of linearity (Field, 2009). Thus, the linear regression is successfully predicting the linear relationship.

Finally, it can be inferred that this study contributes to the related studies as it discuss three important issues (i.e. the effect of network size, traffic and forecasting future performance) which are still under discussion. Above all, the findings provide an evidence on the validity of the proposed conceptual framework and prediction method.

**CONCLUSION**

This study aims at investigating the impact of network size (i.e. Number of nodes) and the FTP traffic load on the performance of the Mobile Ad-hoc Network (MANET). The performance is evaluated in terms of the delay time and the throughput. It is worth
mentioning that the current study also develops a new conceptual model for the prediction of the results of the scenario simulations. The developed prediction models are obtained using the data mining techniques, specially the regression analysis. The encouraging results indicate that network size and FTP traffic are proportionally correlated with throughput. Furthermore, the results demonstrated that the delay time is inversely related to network size in case of medium traffic. In case of large traffic, it is found that the network size is proportionally related to the delay time. As a future research direction, it is suggested to test the validity of the proposed conceptual framework, and also, it is important to verify the relationship between network size, delay and throughput for other mobility models or routing protocols. Finally, based on the encouraging results, it is expected that this study contribute to the related empirical studies.

REFERENCES


