DATA MINING OF NS-2 TRACE FILE

Ahmed Jawad Kadhim

IRAQ- Ministry of Education-
General Directorate of Education in Qadisiyah-
Ibn al Nafis School for Boys

ABSTRACT

Data mining is important process to extract the useful information and pattern from huge amount of data. NS-2 is an efficient tool to build the environment of network. The results from simulate these environment in NS-2 is trace file that contains several columns and lines represent the network events. This trace file can be used to analyse the network according to performance metrics but it has redundant columns and rows. So, this paper is to perform the data mining in order to find only the necessary information in analysis operation to reduce the execution time and the storage size of the trace file.

KEYWORDS

Data mining, NS-2, old trace file, new trace file.

1. INTRODUCTION

Data mining is more suitable process to find the novel, useful, and previously unknown information and knowledge from huge data [1]. This information and knowledge can be used for many applications such as fraud detection, science exploration, and customer to make certain decision to develop their businesses. Many people use term Knowledge Discovery from Data (KDD). But others see that KDD contains sequence of steps that are data cleaning, data integration, data selection, data, transformation, data mining, pattern evaluation, and knowledge presentation. They consider that data mining is an essential and required step in KDD [2]. There are many algorithms used for KDD such as classification, regression, clustering, neural networks, decision tree, association rule, genetic algorithm, and artificial intelligence [3].

2. THE NETWORK SIMULATOR (NS-2)

The network simulator (NS-2) is a discrete event simulator that has several capabilities to build the network environment with all network layers. It can simulate the wire, wireless network, sensor, satellite, and etc [4]. This simulator is building depending on two programing language that are C++ and OTCL [5]. It has many types of mobility models and traffic generators. The results of the simulation that performed based on NS-2 are NAM file (display file) and trace file (analysis file) that differ in storage size according to the network size [6]. There are two types of the trace file:

DOI : 10.5121/ijwmn.2014.6510

2.1 OLD TRACE FILE

This type of the trace file has about 21 columns that records each event occurs through simulation time such as event type, event time, and packet number, etc.

2.2 NEW TRACE FILE

This type takes more memory size than old trace file because it has about 51 columns [7].

3. SIMULATION AND RESULT

The simulation of the mobile ad-hoc network is performed using NS-2 with variant numbers of nodes in order to obtained many trace files that differ in size, number of columns, and number of rows. The simulation applied with old and new trace file. The resulted trace files have large storage size and take more time in the analysis operation. So that, the data mining is applied on these trace files to extract the useful and essentially information for the analysis process. The most useful columns in the analysis of the network events in old trace file only 6 columns that are: event type, event time, object type, packet id, packet type, and packet size and all other columns are often redundant. In new trace file, only 7 columns are important that are: the event type, event time, object type, packet id, packet type, packet size, and the energy. These columns are choosing because only these columns are important and useful in the analysis operation that based on NS-2 to compute the performance metrics that determine the behaviour of the network, routing protocols, traffic generation, and etc. Table1 illustrates the results of data mining for number of nodes with old and new trace file. Figure1, figure2, figure3, figure4 shows the effect of data mining on the size of the old trace file for network with 30, 45, 60, 75 nodes.

Table1: The results of data mining for old and new trace file

Number of Nodes	Trace File Type	Trace File Size		Number of Column		Number of Rows	
		Before Mining	After Mining	Before Mining	After Mining	Before Mining	After Mining
30	Old	1.8 MB	231 KB	21	6	24969	7205
45	Old	3 MB	460 KB	21	6	38843	13913
60	Old	4 MB	698 KB	21	6	49460	20770
75	Old	6.6 MB	1.1 MB	21	6	78611	32660
30	New	4.1 MB	270 KB	51	7	24969	7205

45	New	6.6 MB	537 KB	51	7	38843	13913
60	New	8.8 MB	814 KB	51	7	49460	20770
75	New	14.2 MB	1.28 MB	51	7	78611	32660

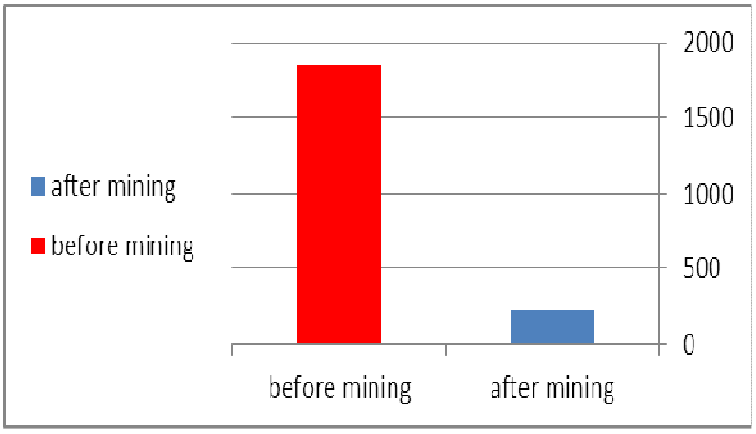

Fig1: The effect of data mining on the size of old trace file for network with 30 nodes

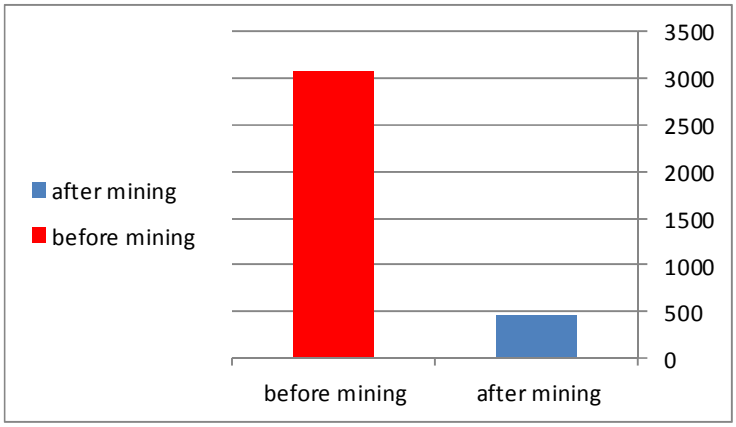

Fig2: The effect of data mining on the size of old trace file for network with 45 nodes

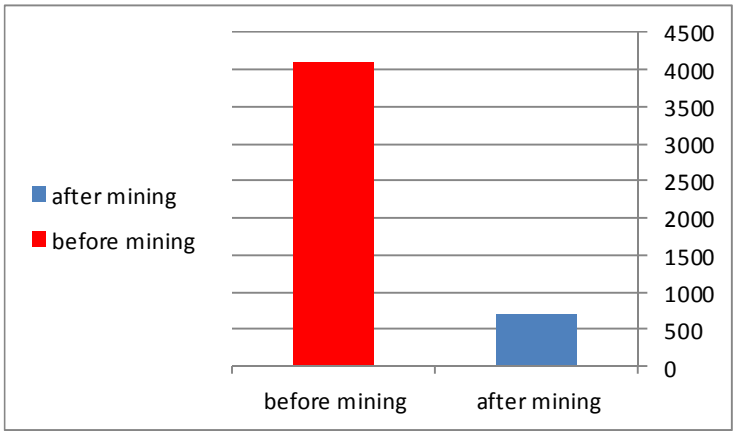

Fig3: The effect of data mining on the size of old trace file for network with 60 nodes

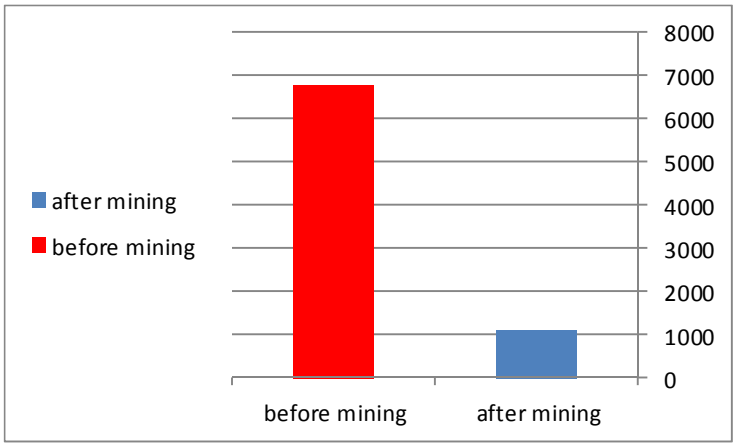

Fig4: The effect of data mining on the size of old trace file for network with 75 nodes

Figure5, figure6, figure7, figure8 illustrates the effect of data mining on the number of lines for the old trace file for network with 30, 45, 60, 75 node respectively.

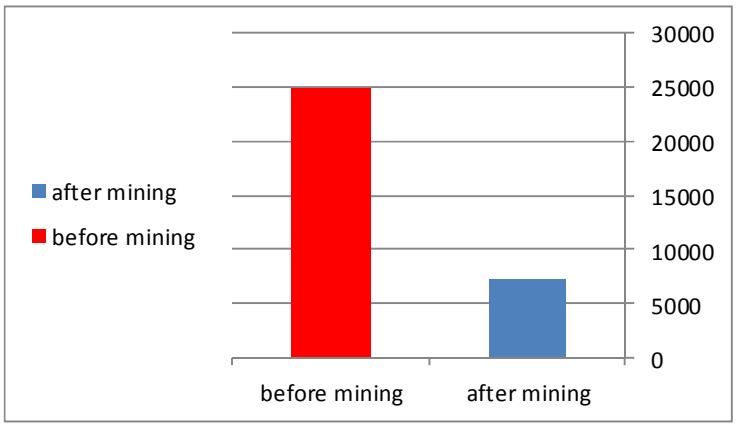

Fig5: The effect of data mining on the lines number of old trace file for network with 30 nodes

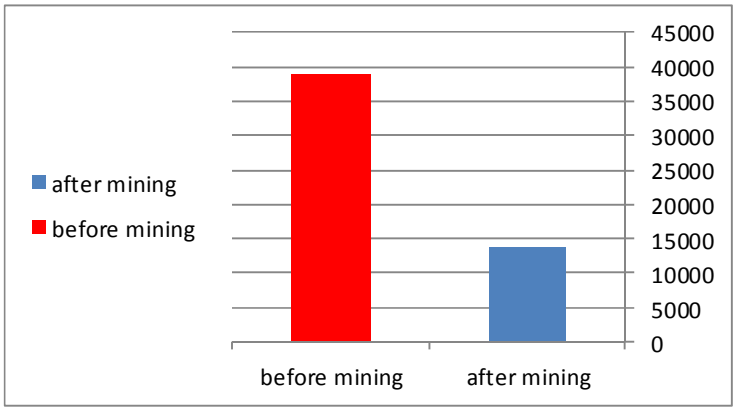

Fig6: The effect of data mining on the lines number of old trace file for network with 45 nodes

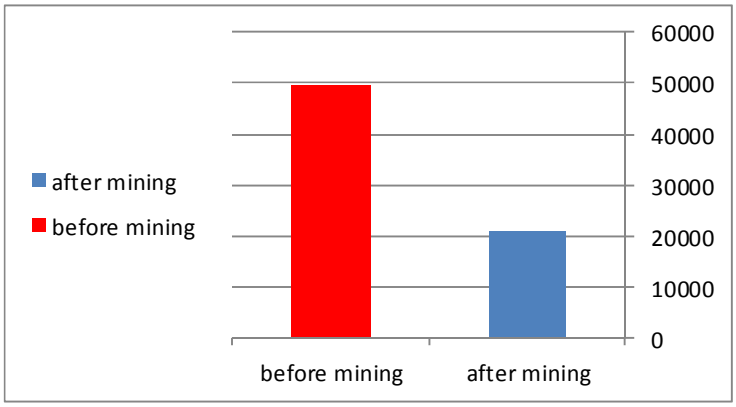

Fig7: The effect of data mining on the lines number of old trace file for network with 60 nodes

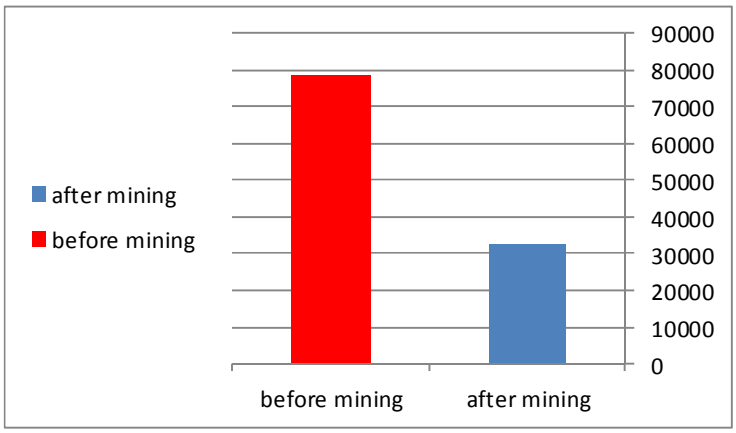

Fig8: The effect of data mining on the lines number of old trace file for network with 75 nodes

Figure9, figure10, figure11, figure12 represents the effect of data mining on the size of the new trace file for network with 30, 45, 60, 75 nodes.

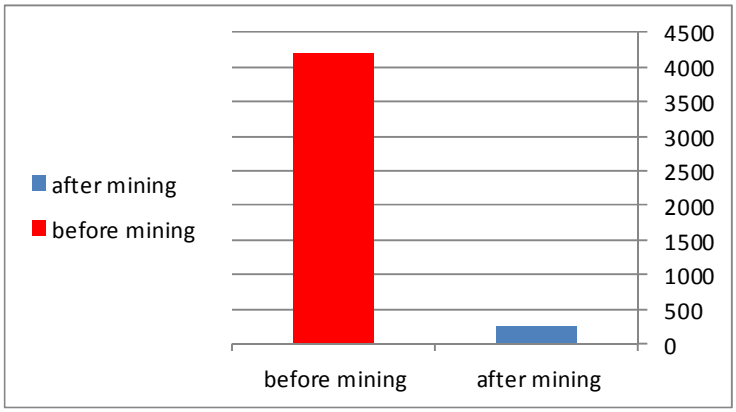

Fig9: The effect of data mining on the size of new trace file for network with 30 nodes

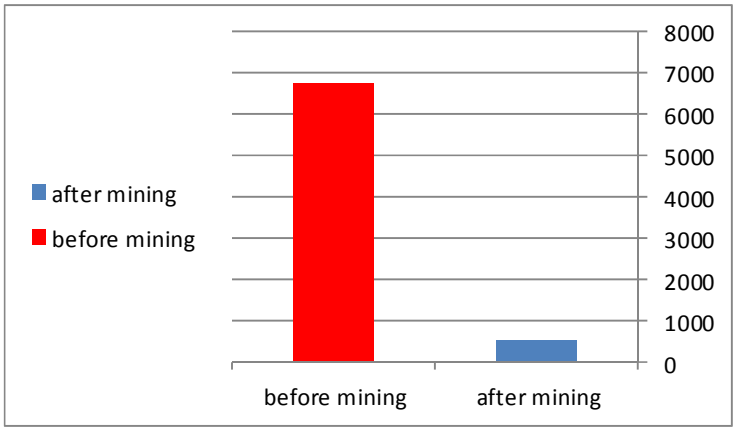

Fig10: The effect of data mining on the size of new trace file for network with 45 nodes

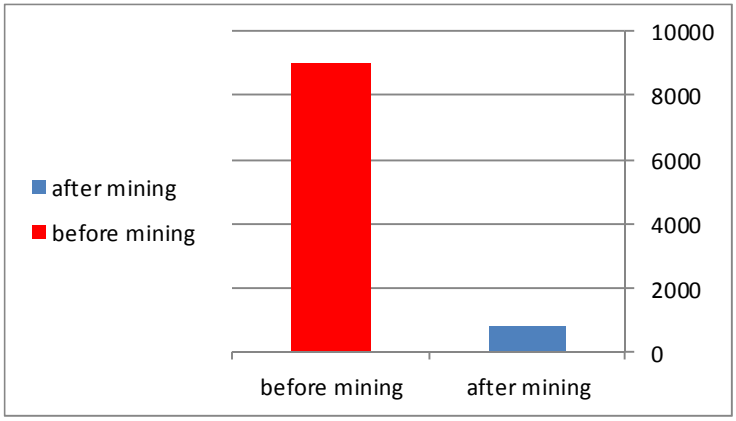

Fig11: The effect of data mining on the size of new trace file for network with 60 nodes

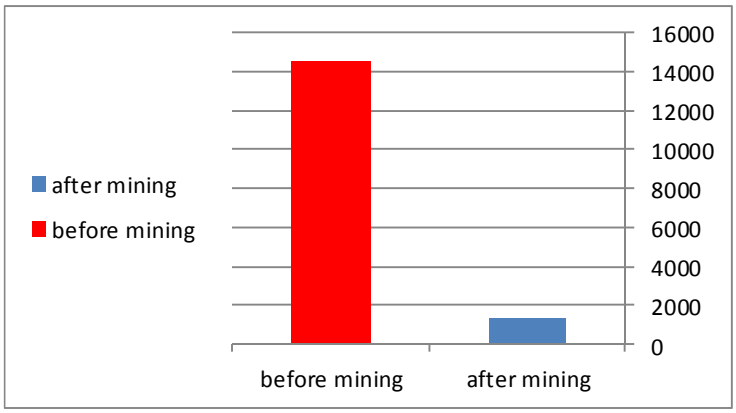

Fig12: The effect of data mining on the size of new trace file for network with 75 nodes.

Figure13, figure14, figure15, figure16 illustrates the effect of data mining on the number of lines for the new trace file to network with 30, 45, 60, 75 node respectively.

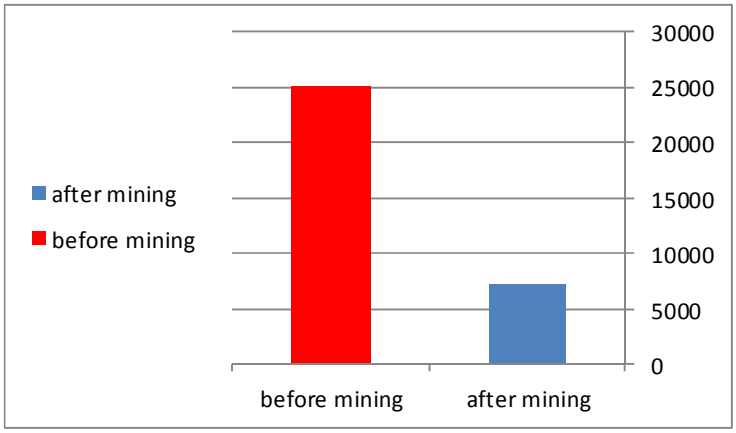

Fig13: The effect of data mining on the lines number of new trace file for network with 30 nodes

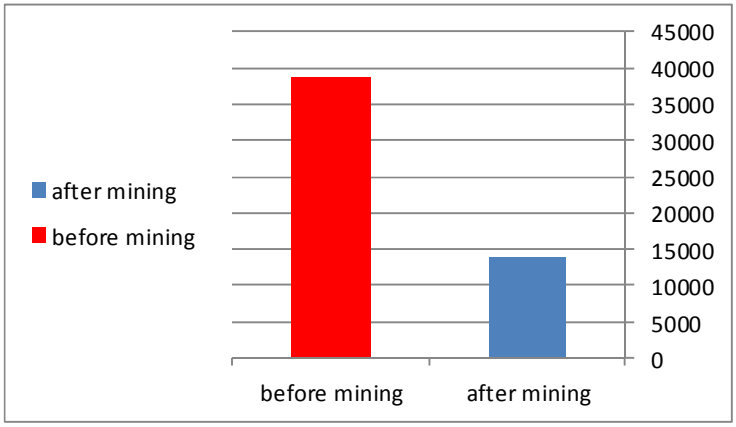

Fig14: The effect of data mining on the lines number of new trace file for network with 45 nodes

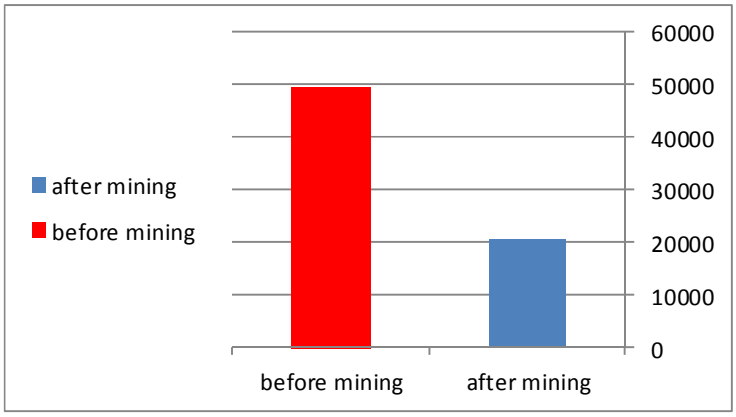

Fig15: The effect of data mining on the lines number of new trace file for network with 60 nodes

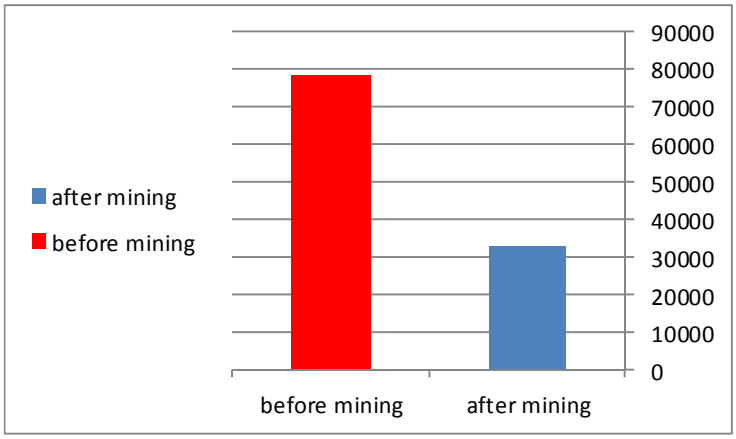

Fig16: The effect of data mining on the lines number of new trace file for network with 75 nodes

4. CONCLUSION

After applied the data mining on the trace files, the execution time decrease about 90% that the time for file before mining. Therefore, the computation speed increases about 90%. The size of the trace file decrease approximately 80%. These results for both old and new trace file. The data mining is a good approach to reduce time and size of NS-2 trace file.

5. REFERENCES

[1] *Bharati M. Ramageri*, " DATA MINING TECHNIQUES AND APPLICATIONS", Indian Journal of Computer Science and Engineering Vol. 1 No. 4 301-305,

[2] Jiawei Han and Micheline Kamber, (2006) "Data Mining Concepts and Techniques", second edition, ISBN 13: 978-1-55860-901-3, ISBN 10: 1-55860-901-6.

[3] David Hand, Heikki Mannila and Padhraic Smyth, (2001) " Principles of Data Mining", ISBN: 026208290x.

[4] Saad Talib Al-jebori, ahmed jawad kadhim, (2012) " Design and Simulation of Optimal MANET Routing Protocol Selection System ", MSc. Thesis, science college, department of computer science, Babylon university.

[5] Jyotsna Rathee, (2009) "Simulation, Analysis and Comparison of DSDV Protocol in MANETS", MSc. thesis, Computer Science and Engineering Department, Thapar University.

[6] Teerawat Issariyakul and Ekram Hossain, (2009) "Introduction to Network Simulator NS2", springer, ISBN: 978-0-387-71759-3.

[7] Bjrn Wiberg, (2002) "Porting AODV-UU Implementation to ns-2 and Enabling Trace-based Simulation", MSc. thesis, Information Technology, Department of Computer Systems, Uppsala University, Sweden.

AUTHORS' BIOGRAPHIES

AHMED JAWAD KADHIM AL SHAIBANY received his B.S. Degree in Computer Science from the Sciences College, Al-qadisyiah University, Iraq, 2007. M.S. degree in Computer Science from
Sciences College, Babylon University, Iraq, 2012**.**