Comparative Analysis of Clustering Techniques in Network Traffic Faults Classification

Karwan Qader, Mo Adda, Mouhammd Al-kasassbeh

School of Computing, Faculty of Technology, University of Portsmouth, Portsmouth, United Kingdom

School of Computing, Faculty of Technology, University of Portsmouth, Portsmouth, United Kingdom

Department of Computer Science, Information Technology College, Mutah University, Jordan

ABSTRACT: Ubiquitous high-speed communication networks play a crucial role in the modern life, demanding the highest level of reliability and availability. Due to the rapid growth of computer networks in terms of size, complexity and heterogeneity, the probability of network faults increases. Manual network administration is hopelessly outdated; complex automated fault diagnosis and management are essential to ensure the provision and maintenance of high quality service in computer networks. Guaranteed Service with higher levels of reliability and availability for real-time applications can be achieved with a systematic approach for real-time classification of network faults, which helps in well-informed (often-automated) decision making. In this paper we discuss three different data mining algorithms as part of the proposed solution for network fault classification: K-Means, Fuzzy C Means, and Expectation Maximization. The proposed approach can help capture abnormal behavior in communication networks, thus paving the way for real-time fault classification and management. We used datasets obtained from a network with heavy and light traffic scenarios in the router and server and built a prototype to demonstrate the network traffic fault classification under given scenarios. Our empirical results reveal that the FCM is more accurate while causing computational overhead. The other two algorithms attain almost the same performance.

INDEX TERMS: Network fault diagnosis, network fault classification, clustering algorithms, Fuzzy Clustering Means, Fuzzy C Means

I. INTRODUCTION

A network fault is an unplanned failure of either hardware or software components which may result in an unusual functioning [1]. Network fault detection and management for networks from which users demand high availability and reliability [2] are indispensable in high-speed Evaluation processing time.High-quality services can be rendered with an online and automated fault diagnosis system that can prevent serious damages [3]. Data communication networks became ubiquitous with ever-increasing size, complexity and heterogeneity, thus automated fault diagnosis and management is crucial for supporting real-time applications [4]. With increasing expansion ofnetwork scale and functionality, network management becomes very complicated in terms of fault diagnosis and management, particularly with traditional tools [5]. There is a continuous progress in the technology of computing nodes and interconnects which demand for high reliability and self-diagnosable abilities [6, 7]. Network faults might be caused by both hardware and software components. Web services commonly used to integrate heterogeneous systems can also cause faults in the overlay networks [8]. Therefore, there is a need for embedded intelligent fault diagnosis capability in computer networks.

As part of network management, fault management is meant to identify and often classify faults in the network. There are ever-growing numbers of networks in the real world that make use of critical components in their network infrastructure. Faults in such networks can cause reliability and availability problems besides increasing the network downtime. This can jeopardize the interests of users of commercial networks, as their real-time applications entail

demand for high quality of service (QoS)[2]. It is commonly knownthat it is very difficult to diagnose and manage network faults that mainly occur due to the heterogeneity and dynamic nature of current networks. Another factor contributing to network faults is the fact that network components are manufactured by different vendors in an open environment [9]. Existingfault management implementations rely on human experts (network managers) to derive knowledge and contribute towards making a set of rules used for fault diagnosis. Such rules and thresholds can cause network monitoring systems to raise alarm to alert network managers [2].

The literature of network management reviews many prior network fault management approaches such as probabilistic approaches [11], advanced database techniques [12], finite state machines (FSMs) [13] and expert systems [10]. An important review of algorithms for network fault identification can be found in [14]. These approaches need fault specifications to be detected and thus cause limitations in performance. Moreover network traffic, applications and configurations are dynamic in nature, thus it is not feasible to have pre-defined fault specifications for all. Further research on fault management is focused on learning machines to address issues in detecting anomalies. However, these techniques do not have an efficient way to collate information collected in time or space.It is necessary to tackle such problems to have automated fault detection mechanisms with adaptive learning for real-time monitoring of networks.

In this paper we built a fault classification model based on three data mining algorithms K-Means, Fuzzy C Means (FCM) and Expectation Maximization (EM). The datasets were collected from the network in the laboratory of the School of Computing at Portsmouth University. The network contains two subnets, two switches and one router. A total of 16 PCs were connected to the network. The first subnet has seven PCs and the second holds nine PCs. The main contributions of this paper is the building of a fault classification system based K-Means, FCM, and EM.

The remainder of the paper is structured as follows. Section II reviews literature on prior works related to network fault management as part of network management pertaining to communication networks. Section III provides an overview of the algorithms deployed in the proposed model. Section IV presents experimental results. Section V presents the results of evaluation. Finally, section VI concludes the paper and gives directions for future work.

II. RELATED WORKS

This section reviews literature on fault classification, especially network fault classification using Probabilistic Neural Network (PNN). Wu et al. [3] proposed an expert system for fault classification using PNN pertaining to internal combustion engines. It works in two phases; in the first phase, sound emission signals are recorded and in the second phase artificial NN is trained to diagnose engine fault conditions. Relating to power distribution networks, Xu et al. [15] investigated fault diagnosis approaches using artificial neural networks (ANNs) and logistic regression (LR).Menget al. [16] proposed a classifier based on radial basis function (RBFNN)that could classify transformer faults using quantum-inspired particle swarm optimization (QPSO) and fuzzy C-means (FCM) techniques, yielding an improvement in comparable accuracy. Similar to [3], Wu and Liu [4] used NN and wavelet packet transform (WPT), which is a well-known technique for signal processing. The NN used is a generalized regression neural network (GRNN), which was proved to provide 95% accuracy.

Hao and Cai-xin [18] proposed a novel classification algorithm based on artificial immune network classification (AINC) for faults in power transformers, while RBF NN was used in [16]. In [18], AINC was usedsimilarly, however their approach to solving the problem differed fundamentally as one uses NN and the other uses an immune network based on learning, memory and network suppression features. Thukaramet al. [19] focused on fault classification using a combination of Artificial Neural Network (ANN) and Support Vector Machine (SVM). For data analysis, Principal Component Analysis (PCA) was also deployedused. Rajakarunakaranet al. [20] employed ANN for fault diagnosis in rotary systems. Janik and Lobos [21] used the combination of SVM and RBF networks for classifying power quality disturbances. RBF was used in [16] also but it was applied for transformer faults. Auld et al. [22] proposed a classifier based on Bayesian neural networks for classifying Internet traffic, achieving an accuracy of 95-99%. Saxena and Saad [23] used a combination of ANN and genetic algorithms (GAs) to classify ideal parameters in rotating mechanical systems.

Radial Basis Function Neural Network (RBFNN)is also used by Samantaryet al. [24]. Other studies usedit differently [16, 21, 25].In [24], RBFNN was used along with HS-transform for classification of objects for protection of transmission lines. The proposed method makes use of HS-transform prior to applying RBFNN variants for fault detection and localization. Wang et al. [25] employed Pulse Coupled Neural Network (PCNN) and PNN for fault

diagnosis of mechanical equipment. PCNN is used for feature vector extraction, while PNN, BP and RBF are used for fast and efficient fault diagnosis. Demetgulet al. [26] employed ANN algorithms for fault diagnosis in pneumatic systems. Khomfoiet al. [24] proposed a system based on NN to diagnose Multilevel Inter Drive (MLID) system, which achieved 85-90% performance in fault diagnosis. Su and Chong [25] proposed an analytical redundancy method using NN for monitoring condition of induction machines. NN is trained with vibration signals. Faults are detected from the vibration spectra modelling error. The system proved to be robust in its functionality. Kanmaniet al. [30] applied PNN for fault prediction in object-oriented software. Burgess [31] proposed a mechanism for resource anomalies in event streams. Gao and Zhou [32] employed PNN for fault diagnosis of computer network. When comparing results of NN methods, PNN displayed the best performance[3, 5, 25].

In this paper,we focused on in K-Means, FCM and EM. K-Means is widely used in data mining algorithms, undermany variations [33, 36-39]. FCM is an improved version of K-Means with wider usage[33-35] is flexible and often deployed to supports hard and soft clusters [40].

III.ALGORITHMS USED FOR FAULT CLASSIFICATION

A. K-Means Algorithm

K-means algorithmhas been usedfor many years to performhard clustering, bywhich an object or a network fault belongs to only one cluster. It is based on similarity measure,depleted to ensure intra-cluster faults are highly similar while inter-cluster faults are highly dissimilar. Its general flow is as shown in eq. (1):

$$J_m = \sum_{i=1}^{k} \sum_{j=1}^{n} \left\| x_i^{(j)} - c_j \right\|^2 \qquad \text{eq. (1)}$$

The specific steps of the algorithm are: 1) identify initial group centroids; 2) use distance measure to identify objects that are closer to a group and assign it; 3) once all objects are assigned, recomputed the position of centroids; and 4) repeat the second and third steps until convergence.

B. Fuzzy C Means

Fuzzy C Means is an improved form of K-Means which is flexible in making clusters. It is generally considered as soft clustering, as an object or network fault can belong to multiple clusters with certain probability. It makes use of the following objective function:

$$J_m = \sum_{i=1}^{N} \sum_{j=1}^{C} u_{ij}^m \left\| x_i - c_j \right\|^2, 1 \le m < \infty \qquad \text{eq. (2)}$$

Where m represents any number, which is greater than 1. The degree of membership is representedby u_{ij}.Then the iterative optimization process makes use of the following equation:

$$u_{ij} = \frac{1}{\sum_{k-1}^{C} \frac{\left\| x_i - c_j \right\|}{\left\| x_i - c_k \right\|}^{\frac{2}{m-1}}}, \quad C_j = \frac{\sum_{i-1}^{N} u_{ij}^m \cdot X_i}{\sum_{i-1}^{N} u_{ij}^m} \qquad \text{eq. (3)}$$

To calculate and update centroids, eq. (4) and eq. (5) are used.

$$C_j = \frac{\sum_{i-1}^{N} u_{ij}^m \cdot X_i}{\sum_{i-1}^{N} u_{ij}^m} \qquad \text{eq. (4)}$$

$$u_{ij} = \frac{1}{\sum_{k-1}^{C} \frac{\left\| x_i - c_j \right\|}{\left\| x_i - c_k \right\|}^{\frac{2}{m-1}}} \qquad \text{eq. (5)}$$

C. Expectation Maximization

Expectation maximization is another data mining algorithm that can be used for clustering. It has many benefits, including its optimization for large number of variables simultaneously and it givesgood estimates for missing values. It supports both hard and soft clustering.

IV. NETWORK FAULTS AND THEIR CHARACTERISTICS

Simple network management protocol (SNMP) is used to collect and organize information about devices of an internet protocol (IP) network such as routers, switches, hubs, printers and servers. SNMP makes use of management information base (MIB) to obtain such information, which includes number of octets or packets sent, delivered, received and dropped. There are 178 MIB variables out of which 12 variables of interface (IF) category were consideredin our experiments. These variables exhibit more sensitivity with respect to the behaviour of traffic in the Internet. The datasets pondered in [41] contain the following 12 parameters to diagnose network faults easily:

Table 1 –Parameters of IF and IP of network

Network Parameter #	Parameter Name
P1	ifInNUcastPkts22
P2	ifInNUcastPkts33
P3	ifInOctets22
P4	ifInOctets33
P5	ifInUcastPkts22
P6	ifInUcastPkts33
P7	ifOutNUcastPkts22
P8	ifOutNUcastPkts33
P9	ifOutOctets22
P10	ifOutOctets33
P11	ifOutUcastPkts22
P12	ifOutUcastPkts33

Network faults are classified into different problems, such as IP conflicting, problems in connection, cable problems, system crashes, and software problems. These flaws could happen due to various reasons. For instance, transmission link failure, network element failure, software failure, protocol failure, malicious attack, unexpected increase in traffic, interface disconnection and miss-configuration [29]. The MIB variables shown in Table 1 are of interface category. The details of datasets are provided in Table 2.

Table 2 – Dataset details

Dataset Name	Number of Instances	Number of Attributes
Heavy Scenario in Router	960	12
Light Scenario in Router	960	12
Heavy Scenario in Server	960	6
Light Scenario in Server	960	6

In this study, four common faults have selected. Datasets were collected in two scenarios, known as heavy weight and light weight (referring to traffic volume), in both server and router [29].

A. Server crash and link failure

This is a kind of network fault (s) which casuses server crash and link failure. MB variables provide useful insights related to abrupt change when the client machines start flooding the server with excessive number of requests. This burst of FTP requests cases the server to get crashed. The MIB variables collected with the environment where 5PCs are used to flood the server simultaneously. For the sake of flooding DoSHTTP tool was used which exploits 5 PCs to sent packets server continously. The attack was performed for 7 minutes and later on 9 minutes. The server got crashed and went down. Two faults are injected in this experiment. They are known as server crash and link failuree.

B. Broadcast Storm

This is another kind of fault identified. This fault is caused in the network environemnt where one of the PCs broadcast messages to other PC to request services and information. The PCs that get broadcast messages start replying them.

This causes load on the netowrk unnecessarily. This network fault causes performance isssue and it also propagates to different parts of the netowrk.

C. Babbling node

It is yet another network fault. It is probably the result of a problem caused by one of the ptotocols of a network or even from the Ethernet card used in a PC. This fault is captured when a PC sendng packets to network continosly. An important observation is that the experimental setup captured faults wich high accuracy in the light scenario when comapred with that of heavy scenario.

V. EXPERIMENTAL RESULTS

Experiments are made to classify network faults in the datasets shown in Table 2under the three data mining algorithms: K-Means, FCM and EM. These clustering algorithms are performed on the given datasets with a prototype application leveraged by Weka support. The results of algorithms with respect to detecting and clustering faults are presented in this section.
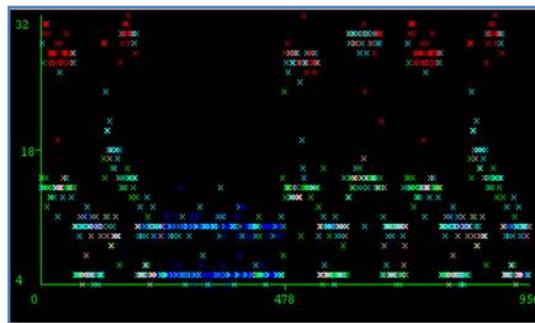
Figure 1 – Results of K-Means for heavy scenario in router dataset

As shown in Figure 1,the network faults are grouped into different clusters. There are five clusters normal traffic, link failure traffic, server crash, broadcast storm and babbling node, denoted by blue, red, green, sky blue and light red colours respectively.
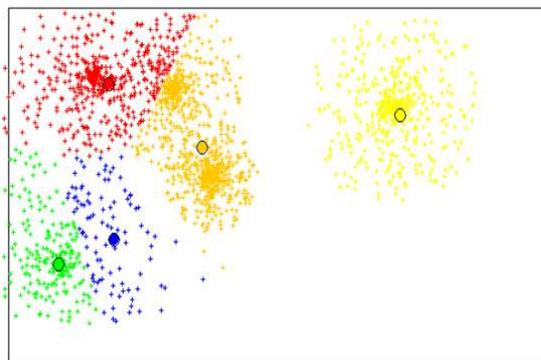

Figure 2 – Results of Fuzzy C Means for light scenario in router dataset

Figure 2 shows network fault clusters representing normal traffic, link failure traffic, server crash, broadcast storm and babbling node (blue, yellow, red, green, light brown, respectively) for FCM algorithm for light scenario in router dataset.
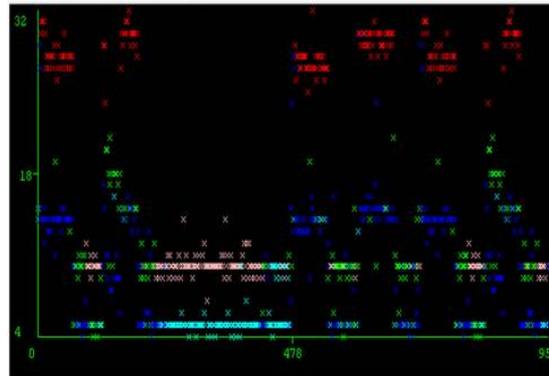
Figure 3 – Results of EM for heavy scenario in router dataset

Figure 3 shows network fault clusters representing normal traffic, link failure traffic, server crash, broadcast storm and babbling node (blue, red, green, sky blue and light red, respectively) for EM algorithm.
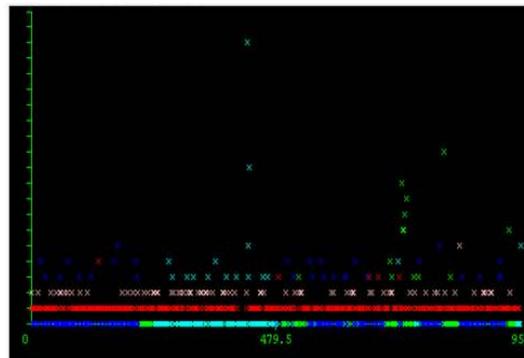

Figure 4 – Results of K-Means for heavy scenario in server dataset

Figure 4 shows network fault clusters representing normal traffic, link failure traffic, server crash, broadcast storm and babbling node (blue, red, green, sky blue and light red, respectively) for K-Means algorithm for heavy scenario in server dataset.
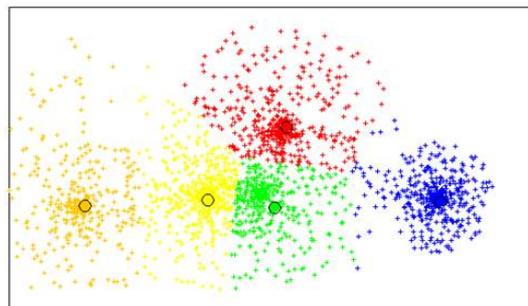

Figure 5 – Results of Fuzzy C Means for heavy scenario in server dataset

Figure 5 shows network fault clusters representing normal traffic, link failure traffic, server crash, broadcast storm and babbling node (blue, yellow, red, green and light brown, respectively) for FCM algorithm for heavy scenario in server dataset.
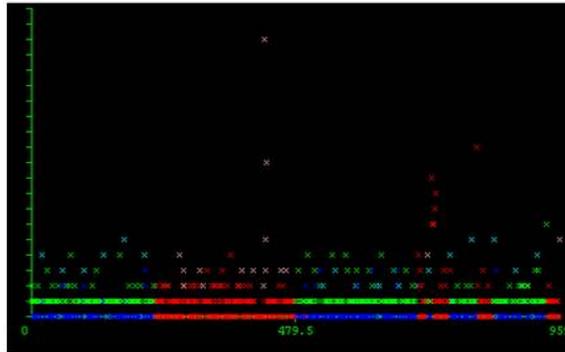

Figure 6 – Results of EM for heavy scenario in server dataset

Figure 6 shows network fault clusters representing normal traffic, link failure traffic, server crash, broadcast storm and babbling node (blue, red, green, sky blue and light red, respectively) for EM algorithm for heavy scenario in server dataset.
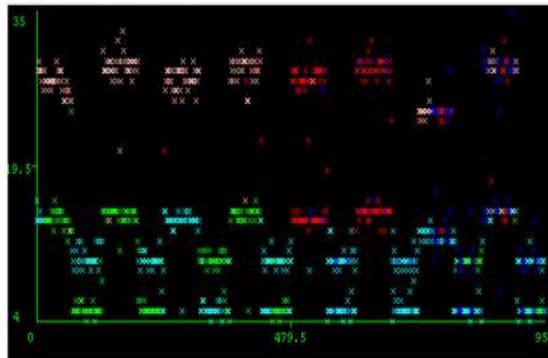

Figure 7 – Results of K-Means for light scenario in router dataset

Figure 7 shows network fault clusters representing normal traffic, link failure traffic, server crash, broadcast storm and babbling node (blue, red, green, sky blue and light red, respectively) for K-Means algorithm for light scenario in router dataset.
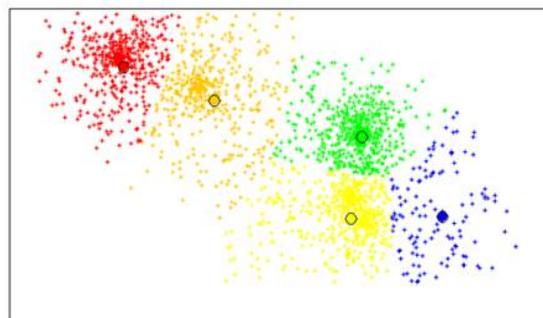

Figure 8 – Results of Fuzzy C Means for light scenario in router dataset

Figure 8 shows network fault clusters representing normal traffic, link failure traffic, server crash, broadcast storm and babbling node (blue, yellow, red, green and light brown, respectively) for Fuzzy C Means algorithm for light scenario in router dataset.
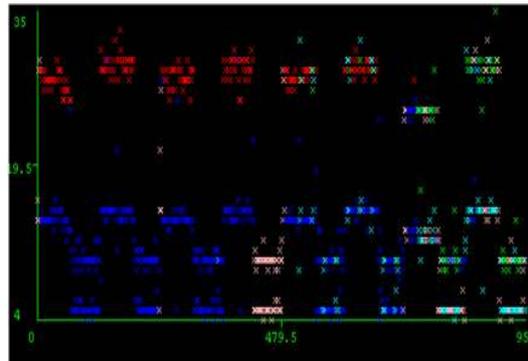

Figure 9 – Results of EM for light scenario in router dataset

Figure 9 shows network fault clusters representing normal traffic, link failure traffic, server crash, broadcast storm and babbling node (blue, red, green, sky blue and light red, respectively) FCM algorithm for EM algorithm for light scenario in router dataset.
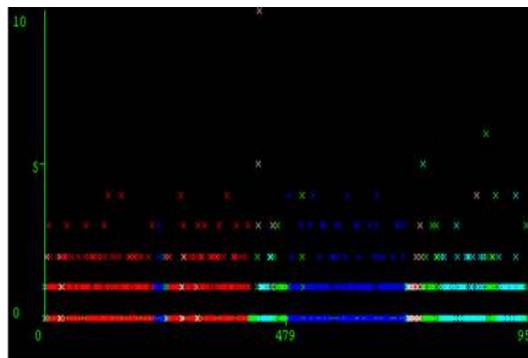

Figure 10 – Results of K-Means for light scenario in server dataset

Figure 10 shows network fault clusters representing normal traffic, link failure traffic, server crash, broadcast storm and babbling node (blue, red, green, sky blue and light red, respectively) for K-Means algorithm for light scenario in server dataset.
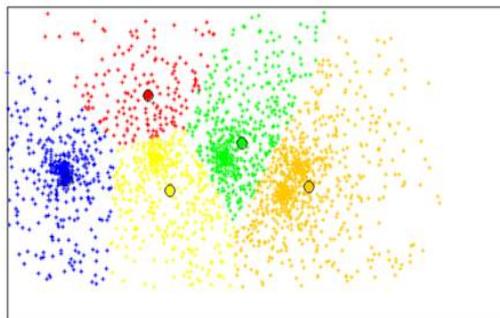
Figure 11 – Results of Fuzzy C Means for light scenario in server dataset

Figure 11 shows network fault clusters representing normal traffic, link failure traffic, server crash, broadcast storm and babbling node (blue, yellow, red, green and light brown, respectively) for FCM algorithm for heavy scenario in server dataset.

Figure 12 – Results of EM for light scenario in server dataset

Figure 12 shows network fault clusters representing normal traffic, link failure traffic, server crash, broadcast storm and babbling node (blue, red, green, sky blue and light red, respectively) for EM algorithm for light scenario in server dataset.
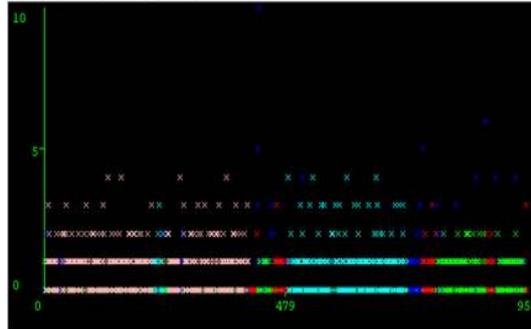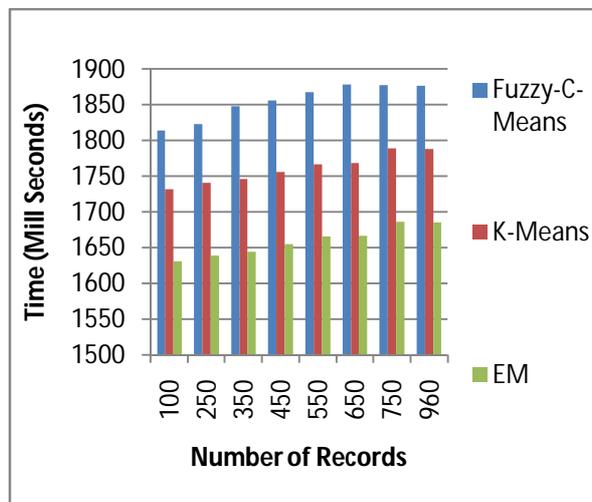

Figure 13 – Evaluation processing time for (FCM, K-Means andEM).

As shown in Figure 13, it is evident that there is a gradual increase in time taken to process data for fault classification. The three algorithms performed differently with respect to time. However, accuracy is greater with FCM,although it took more time than the other algorithms.

VI.EVALUATION OF RESULTS

Precision (number of correctly predicted faults in same cluster/ total number of predicted faults in thesame cluster)and recall (number of correctly predicted faults in same cluster/ total number of faults in the same cluster)are the two metrics used to measure accuracy of the three algorithms used in this paper.

Table 3- Confusionmatrix: precision and recall

	Actual Positive	**Actual Negative**
Predicted Positive	TP	FP
Predicted Negative	FN	TN
Recall = TP/TP + FN Precision = TP/TP+FP True Positive Rate (TPR) = TP/TP+FN False Positive Rate (FPR) = FP/FP+TN		

As shown in Table 3, confusion matrix is presented with precision and recall. The predicted positives are actual positives then it is considered as True Positive (TP). Second case is that predicted positives but actually negatives. Such results are known as False Positives (FP). The third case is predicted negative but actually positive. It is known as False Negative (FN). The fourth case is that predicted negative but actually negative. It is known as true negative.

Table 4– Evaluation results for heavy scenario in router dataset

	Precision	**Recall**
K-Means	40	35.51
Fuzzy C Means	85	25.24
EM	70	30.25

The precision and recall values are presented in Table 4. The precision of FCM is highest while the precision of K-Means is least. In recall of FCM is least while the recall of K-Means is highest. This is the case with the heavy scenario in router dataset.

Table 5 – Evaluation results for light scenario in router dataset

	Precision	**Recall**
K-Means	77	28.54
Fuzzy C Means	90	20.56
EM	85	12.5

The precision and recall values are presented in Table 5. The precision of FCM is highest while the precision of K-Means is least. The recall of EM is least while the recall of K-Means is highest. This is the case with the light scenario in router dataset.

Table 6 – Evaluation results for heavy scenario in server dataset

	Precision	**Recall**
K-Means	80	25.6
Fuzzy C Means	98	12.43
EM	90	7.5

The precision and recall values are presented in Table 6. The precision of FCM is highest while the precision of K-Means is least. The recall of EM is least while the recall of K-Means is highest. This is the case with the heavy scenario in server dataset.

Table 7 – Evaluation results for light scenario in server dataset

	Precision	**Recall**
K-Means	80	20.56
Fuzzy C Means	98	12.43
EM	90	7.5

The precision and recall values are presented in Table 7. The precision of FCM is highest while the precision of K-Means is least. The recall of EM is least while the recall of K-Means is highest. This is the case with the light scenario in server dataset.
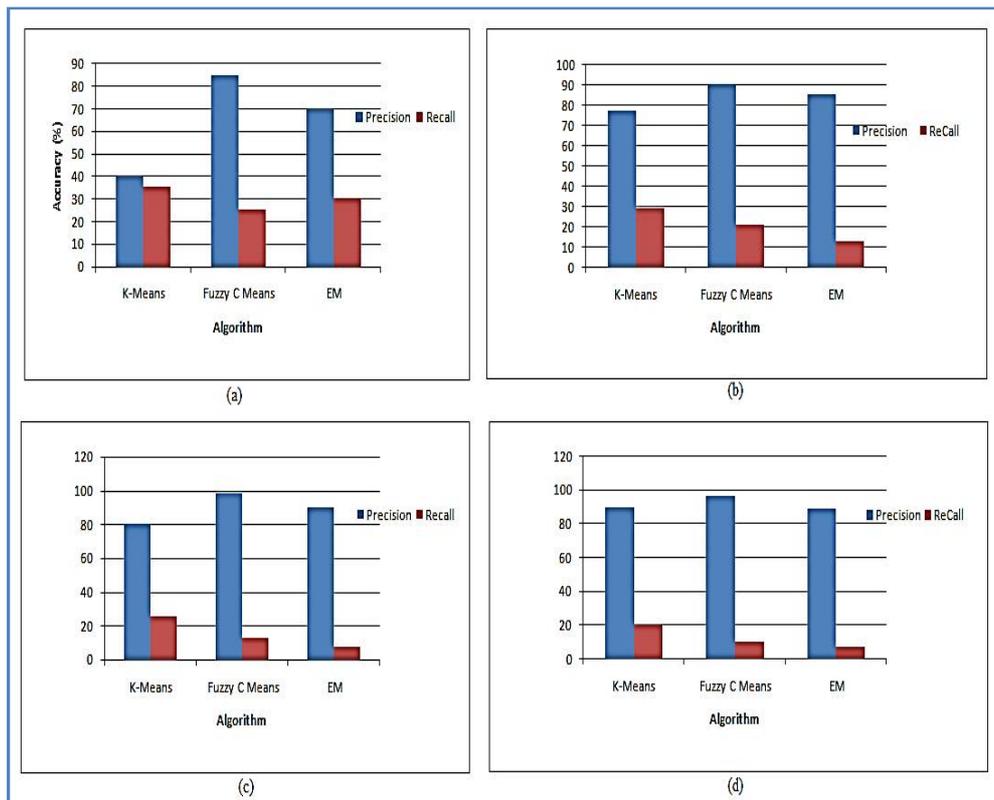

Figure 14 (a) Results of evaluation with precision and recall for heavy scenario in router dataset; (b)light scenario in router dataset, (c) heavy scenario in server dataset, (d) light scenario in server dataset

Figure 14 shows a comparison of the three algorithms' performance differences. FCM demonstrated superior performance in terms of accuracy, but it incurred an execution time penalty to trade-off the accuracy. EM showed better performance than K-Means. Precision and recall are the two measures used to evaluate the quality of the clustering made by different algorithms. Precision denotes the fraction of pairs that are correctly grouped into same cluster. Whereas the recall is the fraction of actual pairs that were identified. True Positive (TP) does mean two objects that are similar are kept in a same cluster. With respect to True Negative (TN) two objects that are not similar are kept in two different clusters. There are two error things. A False Positive (FP) condition assigns two objects that are not similar to the same cluster. A False Negative (FN) condition assigns two similar objects to different clusters. The following are the precision and recall formulae as follows.

$$P=TP/(TP+FP)$$
$$R=TP/(TP+FN)$$

The precession and recall measures are used to evaluate the clustering of different algorithms like K-Means, Fuzzy C Means and EM. The results are presented that reflect the accuracy of clustering of different clustering algorithms. The evaluation is made with different datasets that contain network faults. The datasets on which evaluation is made are heavy scenario in router dataset, light scenario in router dataset, heavy scenario in server dataset and light scenario in server dataset.

VII. CONCLUSIONS AND FUTURE WORK

Network fault detection and identification is crucial for network administrators to make well-informed decisions. The availability and reliability of networks is essential, and fault classification can simplify the decision making process, bypassing the tedious and time-consuming task of manual correlation of faults.

In this paper we studieddata mining algorithms for network fault classification using three popular data mining algorithms as part of our solution for network fault classification: K-Means, Fuzzy C Meansand Expectation Maximization. We used datasets captured from a real-world network containingdifferent faults, including normal traffic, link failure, server crash, broadcast storm and babbling node.

The datasets were takenfrom two different scenarios, with heavy and light traffic loads in server and router; four datasets were thus used for experiments. The results revealed that there are subtle performance differences among the three algorithms. K-Means and EM are relatively faster than FCM, while evaluation with precision and recall in FCM is more accurate, although itrequires more time to process data. There is a trade-off between performance and time taken. We conclude that FCM provides higher accuracy than K-Means and EM. In future, we will extend our research to employ and combine PNN to propose a hybrid method for network faults classification.

REFERENCES

1. Hekmat, S., "Communication Networks", PragSoft Corporation, p.32-44, 2005.
2. Angjeli, A., Cheng, E., and Liptak, L., "Linearly Many Faults in Dual-Cube-Like Networks", *Elsevier*, 472, p.56-60, 2013.
3. Wu, J. D., Chiang, P. H., Chang, Y. W., and Shiao, Y. J., "An Expert System for Fault Diagnosis in Internal Combustion Engines Using Probability Neural Network", *Elsevier*, p2704–2713,2008.
4. Bistouni, F., andJahanshahi, M., "Pars Network: A Multistage Inter Connection Network with Fault-Tolerance Capability",*Elsevier*. 75, p.213-313, 2015.
5. Cisco., "Understanding Fault Management", *CISO*, p.12-17,2003.
6. Jayashree, K., and Anand, S., "Web Service Diagnoser Model for Managing Faults in Web Services", *Elsevier*. 36, p.25-34, 2013.
7. Gao, Y., and Zhou, X., "The Design of Network Fault Diagnosis System Based on PNN", *IEEE*, p.90-107, 2010.
8. Hood, C. S., andChuanyiJit., "Proactive Network Fault Detection", *IEEE*, p.755–768, 1997.
9. Yemini, Y., "A Critical Survey of Network Management Protocol Standards", Telecommunications Network Management Century, pp.1-10,2010.
10. Jakobson, G., and Weissman, M. D., "Alarm Correlation" , *IEEE Network,* pp. 52-59, 1993.
11. Deng, R. H., **Lazar,** A. A.**, and**Wang, W., "A Probabilistic Approach to Fault Diagnosis in Linear Lightwave Networks", *IEEE JSAC,* vol. 11, no. 9, pp. 1438-1448, 1993.
12. Wolfson, 0., Sengupta, *S.,* and Yemini, *Y.,* "Managing Communication Networks by Monitoring Databases", *IEEE Transactions on Software Engineering,* vol. 17, no. 9, pp.1-8, 1991.
13. Rouvellou, I., "Graph Identification Techniques Applied to Network Management Problems" , Columbia University, pp.1-14, 1993.
14. Wang, A. W., and Deng, R., "Models and Algorithms for Network Fault Detection and Identification: **A** Review", *KC,* Singapore, pp.1-14, 1992.
15. Xu, L., and Chow, M. Y., "A Classification Approach for Power Distribution Systems Fault Cause Identification", *IEEE*. 21 (1), p53-60, 2006.
16. Meng, K., Dong, Z. Y., Wang, D. H., and Wong, K. P., "A Self-Adaptive RBF Neural Network Classifier for Transformer Fault Analysis", *IEEE*. 25 (3), p1350-1360,2010.
17. Wu, J. D., and Liu, C. H., "An Expert System for Fault Diagnosis in Internal Combustion Engines Using Wavelet Packet Transform and Neural Network", *Elsevier*, p4278–4286,2009.
18. Hao, X., and Xin, S. C., "Artificial Immune Network Classification Algorithm for Fault Diagnosis of Power Transformer", *IEEE*. 22 (2), p930-935,2007.
19. Thukarm, D., Khincha, H. P., and Vijaynarasimha, H. P., "Artificial Neural Network and Support Vector Machine Approach for Locating Faults in Radial Distribution Systems", *IEEE*. 20 (2), p710-721, 2005.
20. Rajakarunakaran, S. A., Venkumar, P., Devaraj, D., and Suryaprakasa, R. K., "Artificial Neural Network Approach for Fault Detection in Rotary System", *Elsevier*, p740–748, 2008.
21. Janik, P., and Lobos, T., "Automated Classification of Power-Quality Disturbances Using SVM and RBF Networks", *IEEE*. 21 (3), p1663-1669,2006.
22. Auld, T., Moore, A. W., and Gull, S. F., "Bayesian Neural Networks for Internet Traffic Classification", *IEEE*. 18 (1), p223-239, 2007.
23. Saxena, A. A., and Saad, A., "Evolving an Artificial Neural Network Classifier for Condition Monitoring of Rotating Mechanical Systems", *Elsevier*, p441–454, 2007.
24. Samantaray, S. R. ,Dashb, P. K., and Panda, G. , "Fault Classification and Location Using HS-Transform and Radial Basis Function Neural Network", *Elsevier*. n.d (n.d), p897–905, 2006.
25. Wanga, C., Zhou, J., Qin, H., Li, C., and Zhang, Y., "Fault Diagnosis Based on Pulse Coupled Neural Network and Probability Neural Network", *Elsevier*, p14307–14313, 2011.

26. Demetgul, M., Tansel, I. N., Taskin, S., "Fault Diagnosis of Pneumatic Systems with Artificial Neural Network Algorithms", *Elsevier*, p10512–10519,2009.
27. Khomfoi, S., and Tolbert, L. M., "Fault Diagnostic System for a Multilevel Inverter Using a Neural Network", *IEEE*. 20 (3), p1062-1069, 2007.
28. Su, H., and Chong, K. T., "Induction Machine Condition Monitoring Using Neural Network Modeling" , *IEEE*. 54 (1), p241-248, 2007.
29. Kasassbeh, M. A., and Adda, M., "Network Fault Detection With Wiener Filter-Based Agent", *Elsevier*. 32, p.32-44, 2009.
30. Kanmani, S., Uthariaraj, V. R., Sankaranarayanan , V., and Thambidurai, P., "Object-Oriented Software Fault Prediction Using Neural Networks", *Elsevier*, p483–492, 2007.
31. Burgess, M., "Probabilistic Anomaly Detection in Distributed Computer Networks", *Elsevier*. p1–26, 2006.
32. CISCO., "Understanding Fault Management", *CISO*, p.12-17, 2003.
33. Velmurugan, T., "Performance Based Analysis Between K-Means and Fuzzy C-Means Clustering Algorithms for Connection Oriented Telecommunication Data", *Elsevier*. 19, p.213-313, 2014.
34. John, V., and Liu, Z., "Pedestrian Detection in Thermal Images Using Adaptive Fuzzy C-Means Clustering and Convolutional Neural Networks", *MVA Organization*, p.32-44, 2015.
35. Huang , C. W., Lin, K. P., Wu, M. C., Hung, K. C., Liu, G. S., and Jen, C. H., "Intuitionistic Fuzzy C-Means Clustering Algorithm with Neighborhood Attraction in Segmenting Medical Image", *Springer*, p.56-60, 2014.
36. Pan, G., Li, K., Ouyang, A., Zhou, X., and Xu, Y., " A Hybrid Clustering Algorithm Combining Cloud Model Iwo And K-Means", *International Journal of Pattern Recognition*. 28 (6), p.619–627, 2014.
37. Astakhova, N. N., Demidova, L . A., Nikulchev, E. V., "Forecasting Method for Grouped Time Series with the Use of K-Means Algorithm", *Applied Mathematical Sciences*. 9 (97), p.213-313, 2015.
38. Maria, F., Lorenzo, B., and Rizzi, A., "Two Density-Based K-Means Initialization Algorithms for Non-Metric Data Clustering", *Springer-London*, p.2806–2821, 2015.
39. Wang, J., Song, J., Xu, X. S., Shen, H. T., Li, S., "Optimized Cartesian K-Means", *IEEE*, p.325–349, 2014.
40. Kolstein, M., Lorenzo, G. D., and Chmeissani, M., "Evaluation of List-Mode Ordered Subset Expectation Maximization Image Reconstruction for Pixelated Solid-State Compton Gamma Camera with Large Number of Channels", *International Workshop on Radiation Imaging Detectors*, p.23-33, 2014.
41. Qader, K., and Adda, M., "Fault Classification System for Computer Networks Using Fuzzy Probabilistic Neural Network Classifier", In International Conference on Engineering Applications of Neural Networks, p. 217-226, 2014.

BIOGRAPHY

Karwan Qader is currently a final year PhD researcher at School of Computing, Faculty of Technology, the University of Portsmouth-UK. His research work focuses on network fault management, specifically computer network faults classification in real time systems based on proposed novel Fuzzy Probabilistic Neural Network Classifier (FPNNC), which is a hybrid technique based on Fuzzy Cluster Mean (FCM) with Probabilistic Neural Network (PNN). He is interested in doing research in networking and performance evaluation, parallel and distributed systems, cloud computing, architecture and operating systems and artificial intelligence. He has a number of publishes in journals and conferences.

Mo Adda Principal Lecturer at the University of Portsmouth since 2002. He obtained a PhD in distributed systems and parallel processing from the University of Surrey. As a Senior Lecturer, He taught programming, computer architecture and networking for 10 years at the University of Richmond. From 1999-2002, He worked as a senior software engineer developing software and managing projects on simulation and modelling. He have been researching parallel and distributed systems since 1987. His research interests include multithreaded architectures, mobile networks and business process modelling, parallel and distributed processing, wireless networks and sensor networks, network security, embedded systems, simulation and modelling, mobile intelligent agent technology.

Mouhammd Al-kasassbeh Associate Professor at Mutah university, Information Technology. He obtained his PhD in network fault detections at Portsmouth University. His research interests are network traffic analysis, network fault detection, time series analysis, machine learning, intelligent mobile agent systems and intrusion detections systems. He has numbers of publishes in Journals, conferences and books.