Analysis and Predictive Maintenance of Network Management System

Shambhavi G P
Department of ISE
Sir MVIT

Shibani Naik
Department of ISE
Sir MVIT

Surabhi Suresh
Department of ISE
Sir MVIT

Mrs.Rashmi Amardeep
Faculty,Dept. of ISE
Sir MVIT

***Abstract*— This paper lays emphasis on the analysis of Network Management System and its predictive maintenance using an open source prediction tool called Rapid Miner. The network information provided by Network Management System includes device level data such as availability and port level information like errors and discards. This information is archived on periodic basis in a database and can be used for network planning and predictive maintenance.**

Given this information, an independent tool can be developed to analyze the archived data for early detection of potential faults as the basis for network administrators to proactively maintain their network thereby reducing costly downtime.

Keywords—Availability, Error rate, Error discarded packets, Utilization, Naïve Bayes Algorithm, Rapid Miner.

I. INTRODUCTION

A network management system (NMS) [1], [2], [3] consists of software and hardware tools that allow an Information Technology professional to describe components and functions for monitoring, planning and controlling individual components within a large network management framework. Data is usually recorded from the remote points of a network to carry out central reporting to the administrator of a system.

Network failures cause disruption not only by preventing access to the field devices by plant operators, but often also mean that there can be no communication among each other within field devices. Production could be brought to a standstill in the worst case.

A network management system can help the network administrator or the user in the following aspects:

- Discovery of network devices
- Monitoring of network devices
- Analysis of network performance
- Management of network devices
- Customizable alerts or intelligent notifications

Section II presents a typical Network Management System Infrastructure model. Section III introduces the proposed model which in detail talks about the two phases- Analysis and Prediction. The outcome of the previous section is presented and discussed in Section IV. Section V winds up with the conclusion.

II. ARCHITECTURE

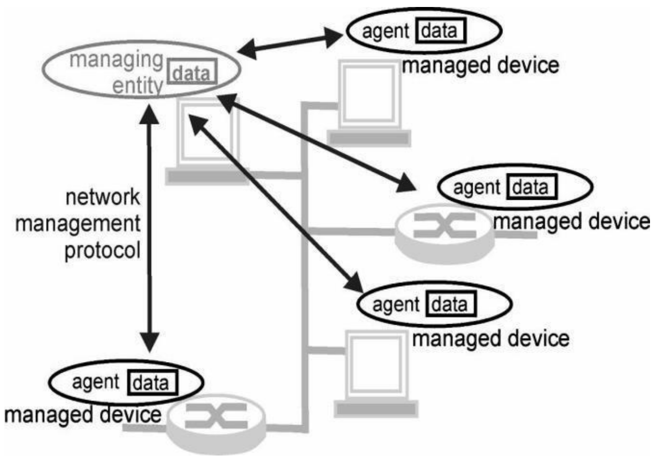

Figure 1. Components of a network management infrastructure

As shown in Fig.1, following are the three principle components of a network management architecture [4]: a managing entity, the managed devices, and a network management protocol.

- The managing entity is the central spot of activity for network management - it controls the aggregation, preprocessing, analysis, and display of information related to network management.
- A managed device is a component of network equipment along with its software that is situated in a managed network. A managed device might be a modem device, host, bridge, router, hub, printer, etc.
- The Network management protocol runs between the managed devices and the managing entity, allowing the status of managed devices to be queried by the managing entity.

III. PROPOSED MODEL

This paper highlights on two phases.

A. *Analysis Phase*

The analysis phase emphasizes on the following instances:

a) *Availability*

Network monitoring primarily looks for faults and tracks the state of the network. Availability determines the reachability factor i.e., to know if the devices are "up" or "down".

The simplest formula for availability is to calculate it as a ratio of the expected value of the uptime of a system to the aggregate of the expected values of up and down time.

$$Availability = \frac{E[Uptime]}{E[Uptime] + E[Downtime]}$$

Where,

E[Uptime] = the expected value of time for which the device is active.

E[Downtime] = the expected value of time when the device is not reachable.

b) *Error Discarded Packets*

Several factors may contribute to the dropping of packets that need to be forwarded to a neighboring device along a routing path or to the adjacent network. Such a loss in the transmission of data packets or loss of traffic includes the following clauses:

- Condition of system overload
- Physical cable faults leading to disruption
- Outages in the network

During analysis, the default acceptable value of error discarded packets are considered to be zero and if discards are detected, the devices from which they're being identified are flagged for further maintenance. Such analysis is made for every interface per device in terms of both transmitting and receiving discards.

c) *Error Rate*

The bit error rate (BER) is the number of bit errors per unit time. The bit error ratio is calculated as the number of incorrect bits received upon the total number of bits transferred for a particular interval of time. BER is often expressed as a percentage and considered to be a unitless entity.

In the simplest form, Error Rate is expressed in terms of total number of incorrect bits to the total number of bits transferred.

$$Error = \frac{Total\ no\ of\ incorrect\ bits}{Total\ no\ of\ bits\ received}$$

In a communication system, Bit Error Rate on the receiver side maybe subjected to the following disruptions due to-interference, transmission channel noise, distortion, wireless multipath fading, bit synchronization problems, attenuation, etc.

d) *Utilization*

Utilization rate [5] is generally referred to as the resource usage for a particular period of time which is represented as percentage of usable resource to the full capacity. It is specifically applied in the areas where finding a potential blocking area or a bottleneck is of main interest. The importance of utilization ratio surfaces because, the resource utilization ratio determines the exponential increase in response time. Also, underutilized devices or resources that cannot be fully exploited are identified by measuring the utilization.

Messages would be discarded if the interface queue is full. Utilization ratio helps in troubleshooting the problems on a timely basis so that network paralysis can be avoided by carrying out necessary route adjustment or load balancing.

Below is the formula to calculate utilization in a full duplex scenario:

$$Utilization = \frac{Max(\Delta ifInOctets, \Delta ifOutOctets) \cdot 8 \cdot 100}{IntervalSeconds \cdot ifSpeed}$$

Where,

ifInOctets = Number of Inflow data bytes.
ifOutOctets = Number of Outflow data bytes.

ΔifInOctets = Difference between two successive checking of set intervals i.e., inflow data in the time interval.

ΔifOutOctets = Difference between two successive checking of set intervals i.e., outflow data in the time interval.

B. *Prediction phase*

Prediction refers to supervised and unsupervised learning task where the data is used directly to predict the class value of a new instance. For the predictive analysis of Network management system an open source tool Rapid Miner [6],[7] has been used which is a software platform that provides an integrated environment for machine learning, business analytics, text mining, data mining and predictive analytics.

Following are the steps followed for prediction:

Step 1: Pre-processing- Data loading of the previous machine runs along with information about occurrence of failures.

Step 2: Using various attribute weighting algorithms to determine the influence factors and averaging their weights results.

Step 3: Training a model-consists of the optimization phase where validation of the model is done.

Step 4: Post processing- Loading fresh data and applying the machine failure model to predict potential failures in the machine.

The algorithm used is Naïve Bayes Classifier [8],[9],[10] which assumes that the value of a particular feature is independent of the value of any other feature, given the class variable. It is a technique for constructing classifiers: models that assign class labels to problem instances, where the class labels are drawn from some finite set.

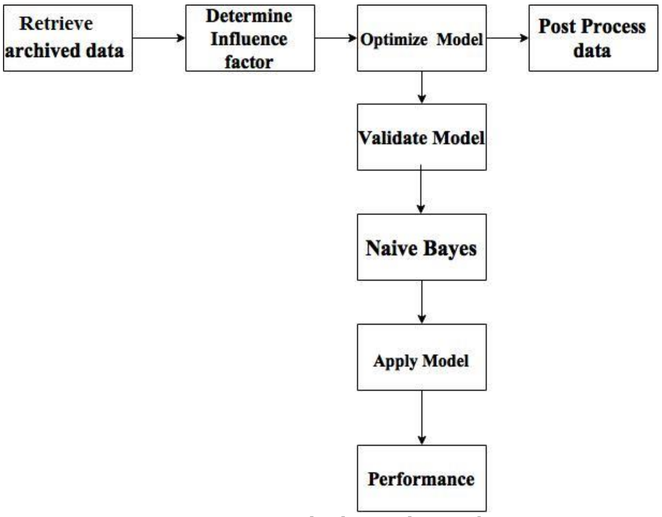

Figure 2. Steps involved in Prediction phase

IV. RESULTS AND DISCUSSION

A. *Availability*

Table 1. Reachability table per device

DEVICE_ID	REACHABLE STATE	STATUS_CHANGE_ TIMESTAMP
16	NotReachable	12/2/15 11:25 AM
16	Reachable	12/2/15 11:32 AM
16	Reachable	12/2/15 11:32 AM
16	Reachable	12/2/15 11:32 AM

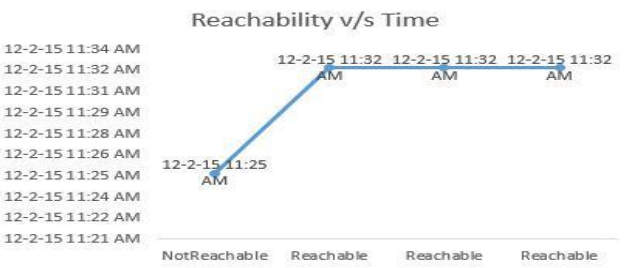

Figure 3. Availability graph for device level data

The graph shown in Fig.3 depicts the reachability factor for a particular device at different intervals of time as recorded in the Table 1. This serves the purpose of determining at what time the device is available.

B. *Error discarded packets*

Table 2. Interface Error discarded packets

DEVICE ID	INTERFACE	INTERFACE DISCARDS	TIMESTAMP
65	1004	42	12-14-15 10:49
65	1001	46	12-14-15 10:49
65	1004	16	12-14-15 10:50
65	1001	19	12-14-15 10:50

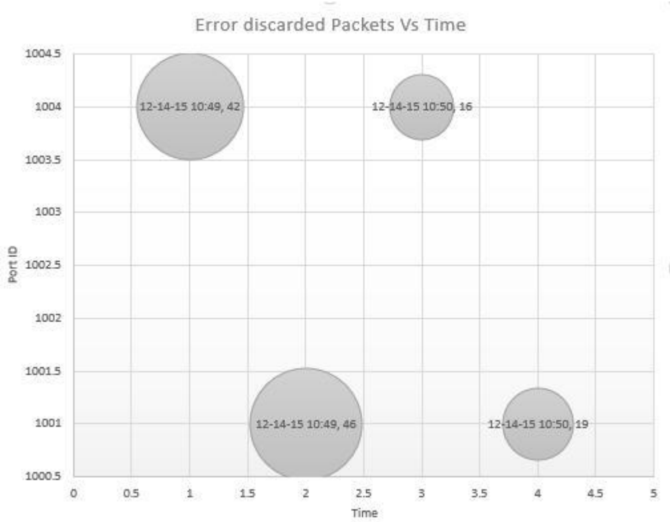

Figure 4. Graph of error discarded packets against timestamp

The balloon graph in Fig.4 illustrates the number of error discarded packets for different ports per device as tabulated in Table 2. Lesser the number of discards, better is the efficiency since retransmission is avoided.

C. *Error Rate*

Table 3. Receiver side error rate

DEVICE ID	INTERFACE INDEX	INTERFACE RX ERROR	TIMESTAMP
20	1001	10.15	12/3/15 15:54
20	1001	12.68	12/8/15 15:34
20	1001	14.34	12/8/15 15:59
20	1001	15.85	12/8/15 16:04
20	1001	11.6	12/8/15 16:09

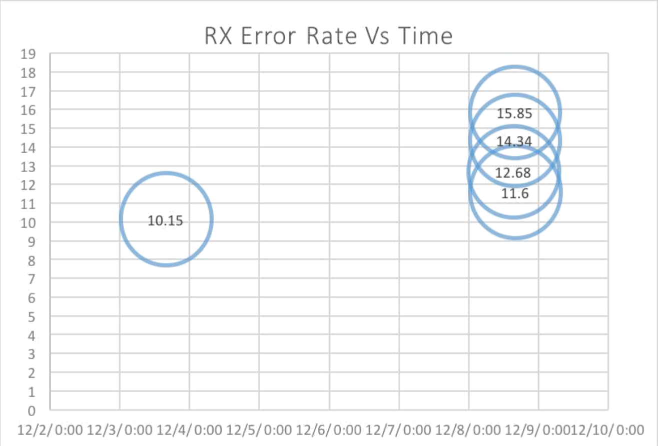

Figure 5. Graph for Receiver side error rate against time

Fig.5 demonstrates the error rates for a particular port of a device within a span of few minutes as recorded in Table 3.

D. Prediction result

Table 4. Accuracy table obtained considering only the reachability factor

	Ports	Accuracy
Total	17888	
Genuine Accepted Rate(GAR)	8937	49.96%
False Rejection Rate(FRR)	8951	49.58%

For a given set of data, the availability was analyzed and predicted using Naïve Bayes classifier, taking into account only the reachability factor. The accuracy for Genuine Accepted Rate was found to be 49.96% as shown in Table 4. And 49.58% reckons for cases that were falsely predicted by the model which is quite reasonable but could be improved further.

Table 5. Improved accuracy table with the addition of error and utilization

	Ports	Accuracy
Total	2317	
Genuine Accepted Rate(GAR)	2290	98.83%
False Rejection Rate(FRR)	27	1.89%

Table 5 shows the improved accuracy for Genuine Accepted Rate, which is as high as 98.83% for predicting the same availability with the before said algorithm after combining multiple factors such as error and utilization.

V. CONCLUSION

In this paper, the devices on a network are analyzed based on parameters such as availability, error rate, discarded packets and utilization rate. Further, the availability of the device has been predicted using the Naive Bayes algorithm, through a data mining tool called Rapid Miner. The accuracy of the prediction has been estimated to be approximately 98% upon combining multiple factors. On the whole, this paper presents a solution for the network administrator to analyze the network and predict forthcoming failures and is an open idea for future enhancements on the same.

REFERENCES

[1] S. Chantler, “Network management system (NMS), Manweb,” in *Fourth International Conference on Power System Control and Management*, 1996, pp. 210–215.

[2] J. Burchell, “The strategy for a network management system,” in *Fourth International Conference on Power System Control and Management*, 1996, pp. 216–220.

[3] S. F. Bush and A. B. Kulkarni, *Active networks and active network management: a proactive management framework*. Springer, 2001.

[4] M. Rose, "Simple Book, The: An Introduction to Internet Management, Revised Second Edition," Prentice Hall, 1996.

[5] Weiming Shen, Zongkai Lin, Jean-Paul A. Barthes, Tangqui Li(Eds.), “Computer Supported cooperative work in Design I”,2004.

[6] V. Kotu and B. Deshpande, “Getting Started with RapidMiner,” in *Predictive Analytics and Data Mining*, Elsevier, 2015, pp. 371–406.

[7] T. Verma, R. Renu, and D. Gaur, “Tokenization and Filtering Process in RapidMiner,” *Int. J. Appl. Inf. Syst.*, vol. 7, Apr. 2014.

[8] D. Barber, “Naive Bayes,” in *Bayesian Reasoning and Machine Learning*, Cambridge University Press, 2011, pp. 243–255.

[9] A. C. de Pina and G. Zaverucha, “Combining attributes to improve the performance of Naive Bayes for Regression,” in *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*, pp. 3210–3215.

[10] J. Lin and J. Yu, “Weighted Naive Bayes classification algorithm based on particle swarm optimization,” in *2011 IEEE 3rd International Conference on Communication Software and Networks*, pp. 444–447.