ANOMALY IDENTIFICATION AND FAILURE DIAGNOSIS

Swapnil B Kadam[1], Dr.S K Shirgave[2]
[1]Research Scholar, D Y Patil College of Engg. and Technology,Kasba Bavda,Kolhapur
[2]Guide,DKTE's college of Engineering & Technology,Ichalkaranji

Abstract

When we work in a large scale network, number of problems arises, the total time required to deal with these type of problems depends on how severe the problem is? As system takes more time to recover from failures, maintenance cost goes on increasing,it also causes loss of processing cycles. To deal with such type of loss, the information at various nodes in network is collected and verification of failure reasons is performed. In traditional system this process of dealing with failures was handled by humans, but such a manual processing was leading to various problems such consumption of time, scalability of network and many more. As scalability of network goes on increasing we should think on automation of anomaly identification to perform failure diagnosis.

Index Terms : Abnormal, anomaly, communication, detection, failure, timer, network, diagnosis, identification.

I. INTRODUCTION

As we go with more complex network, diagnosis of fault becomes a difficult task for network operators. Typically, one fault in the communication system produces large amount of alarm information,which is called alarm burst.Because of the huge information,manual cause identification becomes time consuming and error prone. Therefore, automated fault diagnosis in computer networks is a problem. The occurance of faults could be disastrous in terms of human mortality and environmental impact. Fault detection in process and manufacturing industries is also important to improve production efficiency, product quality and production cost.

In todays increased complexity of computer networks,one single fault in one network component might cause considerably high volume of alarms,which is called alarm burst.Alarm burst may be a result of

1) fault re-occurrence.
2) Multiple service invocations provided by a faulty component.
3) Multiple alarm generation by a device for a single fault.
4) Identification of and issuing a notification about the same fault by many devices at same time.
5) Error introduced in network devices causing them to fail,which causes generation of additional alarms.

Thus, it is a challenge for network operators to quickly and correctly identify the root cause, by analyzing those large amounts of alarms.

II. RELATED WORK

Under different working environment, such diagnosis algorithm behaves differently. Different models like interaction model, time and clock model, communication model and failure model explain the working environments. There are plenty of algorithms designed for the diagnosis of faults in such systems. The algorithm implemention is based on adaptive distributed system diagnosis. The word adaptive indicates that at different stages of algorithm dynamic decisions are taken according to the situation. The System level means set of indivisible units called processing elements. Each system node works independently and shares information with other system by passing message.Existing diagnosis algorithms like ADSD [7] (Adaptive distributed system level Diagnosis) and HiADSD [8], [9] (Hierarchical Adaptive distributed systemlevel

Diagnosis) compromise minimization of diagnosis latency to reduce network bandwidth consumption. Both algorithms have considered fully connected network topologies. Diagnosis algorithm performance is described in terms of correctness.Referred to as bounded correctness, consists of three properties,

1) Bounded Diagnostic Latency: all working nodes must learn about node failure or repair within a bounded time.
2) Bounded start-up: Nodes recovered must determine a valid state for every other node within a bounded time S of entering the working state.
3) Accuracy: ensures that any working node records no spurious events.

III.COMMUNICATION FAILURE DETECTION

Whatever existing system available to detect communication failure are manual or not providing good sensivity and specificity under various fault.To overcome existing system of communication failure detection mentioned is a way of providing automation of anomaly detection for various components in a distributed network. Component failure is an important problem in distributed system.In distributed network,each working node must maintain correct node level information about the status (working or failed) of each node or component in the system. Distributed network have systems in which hardware or software components of network carries communication and coordination of their actions by passing messages. In such systems,it is difficult to predict the behavior of system under various faulty situations. Under faulty environment, it becomes necessary to handle the faults more effectively and keep systems in a running condition. To handle faults in such a environment, it becomes esseential to detect and diagnose these nodes. Under different working situations,such diagnosis algorithm behaves in different ways.Different models like interaction model, time and clock model,communication model and failure model explain the environment of working situation. The goal of fault detection and diagnosis research is improving the security, efficiency,maintainability and reliability of network. A fault is called intermittent if its effects on the system are hidden for discontinuous periods of time. Even though a fault is tolerable, it must be diagnosed as early as possible,as it may lead to serious consequences in time.A fault diagnosis system is a monitoring system that is used to detect faults and diagnose their location and significance in a system. The system performs the following tasks:

1) Fault detection - to indicate weather a fault occurred in the system or not.
2) Fault isolation - to determine the location of the fault.
3) Fault identification - to estimate the size and nature of the fault.

A system fails when it cannot meet its promises. In particular, if a distributed system is designed to provide its users with a number of services, the system fails when one or more of those services cannot be completely provided. An error is a part of a systems state that may lead to a failure. Failures can be further classified as shown below.

1) Crash failure: A server works correctly until it halts.
2) Omission failure: When a server fails to respond to incoming requests, omission failure occurs.
3) Response failure: When a servers responses incorrectly, response failure occurs.
4) Arbitrary failure: Arbitrary failure occurs when a server may produce arbitrary responses at arbitrary times.
5) Timing failure: It can be a performance failure or clock failure. Performance failure occurs when either process exceeds the bounds on the interval between two steps or a messages transmission takes longer than the stated bound. Clock failure is the failure in which processes local clock exceeds the bounds on its rate of drift from real time.

Implemented diagnosis algorithm considers crash faults in nodes. It can be assumed that network delivers messages reliably. However, diagnosis algorithms can be transformed into test based algorithms and vice versa. Using this transformation, the algorithms could be easily converted to ones that use testing and the crash fault assumption could then be removed.The status of a node is modeled taking into consideration failed state and working state. Failed state nodes do not send messages nor do they perform any computation. Working state nodes respond properly to the diagnosis procedure.To reduce overhead, the heartbeat algorithm is implemented with multicast,for fully connected networks and with unicast for not completely connected networks.Both mechanisms are implemented over UDP/IP.

IV. SYSTEM MODULES

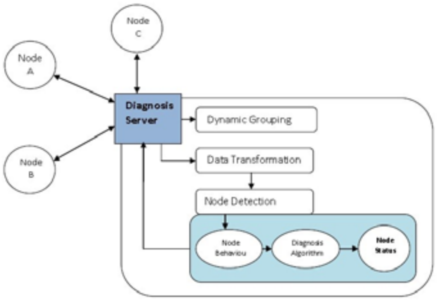


Fig. 1. Overview

Figure shows the framework for Anomaly Identification and Failure Diagnosis. The Node A, B and C are distributed node in the network. Diagnosis server periodically discovers the new node and form dynamic grouping based on the node status. Diagnosis server implements data transformation to perform node detection. Node detection provides the node status which is observed based on the node behaviour and diagnosis algorithm. Detection system has been consist of Dynamic Grouping :

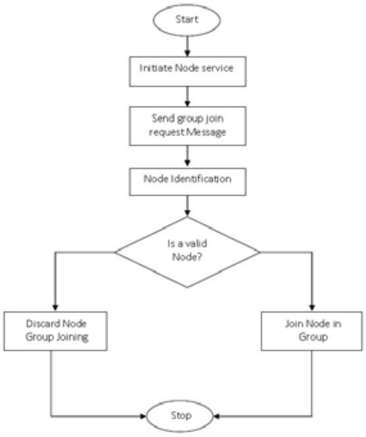


Fig. 2. dynamic grouping

Node grouping dynamically can be performed either unicast or multicast broadcasting in a network.It is possible to implement the algorithm using any type of communication algorithm to discover a faulty node. This module implementation for dynamic grouping considers both completely connected and not completely connected networks. In a completely connected network, there is a direct communication channel between all pair of nodes. This is a requirement to achieve bounded correctness with an arbitrary number of node failures. In notcompletely connected networks,intermediate nodes relay messages between some source-destination pairs. Hence, the number of node failures is reduced such that the network remains connected at all the time. To reduce over head, the grouping algorithm is implemented with multicast, for completely connected networks and with unicast for notcompletely connected networks.

- Data Transformation:

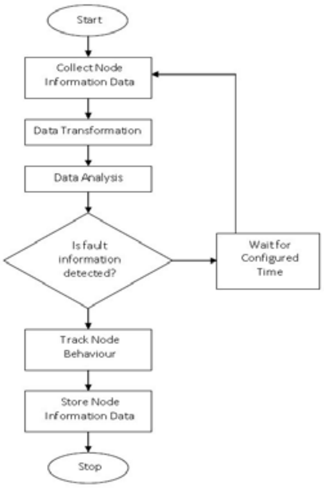

Start
Collect Node Information Data
Data Transformation
Data Analysis
Is fault information detected?
Wait for Configured Time
Track Node Behaviour
Store Node Information Data
Stop

Fig. 3. Data Transformation

This module implements the function to collecting related information across the system and assembling them into a uniform format, this format is called as feature matrix. Here, a feature of a node is defined as any individually measurable characteristic or variable of the node being observed. A system fails when it cannot meet its promises. In particular, if a distributed system is designed to provide its users with a number of services, the system fails when one or more of those services cannot be completely provided. An error is a part of a systems state that may lead to a failure. The cause of an error is called a fault. Implementing the diagnosis algorithm considers crash faults in nodes. It can be assumed that network delivers messages reliably. However, this algorithm can be transformed into test based algorithms and vice versa. Using this transformation, the algorithms could be easily converted to ones that use explicit testing and the crash fault assumption could then be removed.

- Feature Extraction

  A feature extraction is applied on the feature matrix to obtain a matrix which has much lower dimensionality while keeping the most relevant information in the data. This not only gives acceleration to data analysis by reducing data dimensionality but also improves the quality of data analysis by removing inherent data dependency.

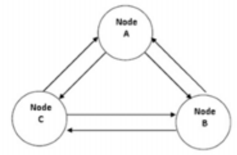


Fig. 4. feature extraction

- Node Detection

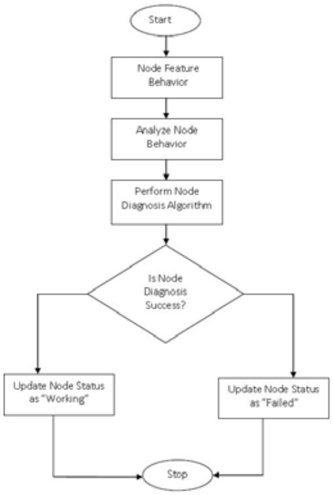
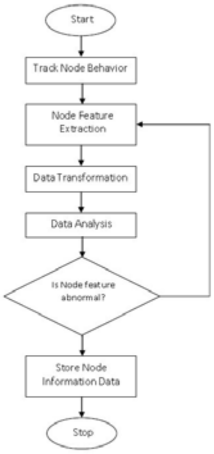

Start
Node Feature Behavior
Analyze Node Behavior
Perform Node Diagnosis Algorithm
Is Node Diagnosis Success?
Update Node Status as "Working"
Update Node Status as "Failed"
Stop

Fig. 5. Node status

Node detection module is used to determine the nodes that are behaving differently from the majority of node and this behavior is termed as anomalous (i.e. abnormal behavior). By analyzing this matrix generated by feature extraction, an outlier detection algorithm such as cell based algorithm is used to quickly identify the outliers.The status of a node is modeled by a state machine with two states, failed and working. Failed nodes do not send messages nor do they perform any computation. Working nodes execute faithfully the diagnosis procedure.



Whenever node A respond properly it means the status of this node is working and initially status of all other nodes is unknown. Detection algorithm sets send message timer for sending messages and receive message timer for receiving messages. It sends messages periodically. On receiving message from node B or C, node A sets status of corresponding nodes as working and again resets timer

Fig. 6. Node Exchanging heartbeats

for Receive. When timer expire for Send, it sends message to all network nodes and sets send timer with predefined interval period.When receive timer expires, host node sets status of corresponding node as failed.

Let us assume that X node enters into working state it executes as, Above flow chart shows the exchange of

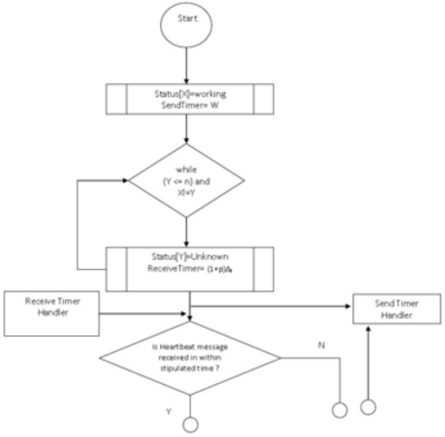


Fig. 7. Node Exchanging heartbeats

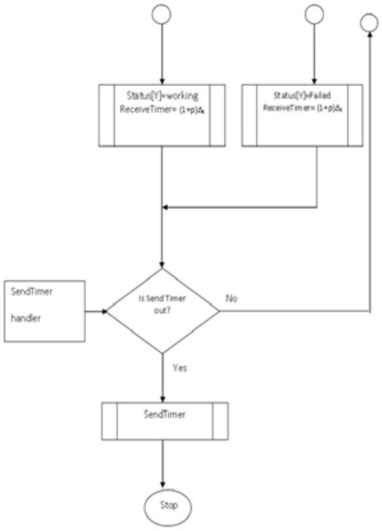


Fig. 8. Node Exchanging heartbeats anomaly identification and diagnosis messages by nodes.

V. EXPERIMENTS & RESULTS

Whenever say node A enters in working state that means the status of this node is working and initially status of all other nodes is unknown. It also sets send message timer for sending messages and receive message timer for receiving messages. It sends messages periodically. On receiving message from node B or C, node A sets status of corresponding nodes as working and again resets timer for Receive. On expire of timer for Send, it sends message to all network nodes and sets send timer with predefined interval period. On expire of receive timer, host node sets status of corresponding node as failed.Node status information is as showen with node status information table

Sender	Reciever	Timer set	Respond time	Ack	Status
Node A	Node B	60 Sec	within 60 Sec	Recieved	Node B working
Node A	Node B	60 Sec	More than 60 Sec	Timer Expire	Node B failed

TABLE I

NODE STATUS INFORMATION

VI. CONCLUSION

When a system is not working as per expectation , system related information is collected for troubleshooting. it is always one of the challenging task to identify anomalies from the amount of noisy and high dimensional data[1]. The traditional manual anmaly identification approach is timeconsuming, error-prone, and even worse and not scalable.In this proposed system, we present an automated mechanism for node-level anomaly identification in large scale systems. A set of techniques are presented to automatically analyze collected data,perform data transformation to construct a uniform data format for data analysis and unsupervised learning to detect the nodes acting differently from others.We can effectively identify faulty nodes with high accuracy and low computation overhead. System proposed in the paper should identify anomalies with highest probability and identifying nodes under failure, Making fault tolerant system.

REFERENCES

[1] Zhiling Lan, Member, IEEE Computer Society, Ziming Zheng,Student Member, IEEE, and Yawei Li, Member, IEEE To ward Automated Anomaly Identification in Large-Scale Systems,*IEEE Transaction on parallel and distributed system*,Vol. 21,No. 2, pp.174-187 February 2010.

[2] Li Zonglin, Hu Guangmin, Yao Xingmiao Multi-dimensional traffic anomaly detection based on ICA ,*IEEE conference, pp.*333-336 , 2009.

[3] A. Hyvarinen and E. Oja, Independent Component Analysis Algorithms and Applications, Neural Networks, vol. 13, nos.4/5, pp. 411430, 2000.

[4] Y. Rao and J. Principe, A Fast, On-Line Algorithm for PCA and Its Convergence Characteristics, *Proc.IEEE Signal Processing Soc.Workshop* , 2000.[5]Anomaly Detection System Based on PCA wuhan university journal of naturascience vol.11 no.06,pp.17691772, 2006.

[5] Preparata F., Metze G., and Chien R., On the connection assignment problem of diagnosable systems, *IEEE Trans. Elect.Comput. EC-16, 6 (Dec.)*, pp. 848-854,1967.

[6] Kuhl J. and Reddy S., Distributed fault-tolerance for large multiprocessor systems, In Proceedings of the 7th Annual Symposium on Computer Architecture, pp. 23-30, 1980

[7] R. Bianchini and R. Buskens, Implementation of On-Line Distributed System-Level Diagnosis Theory, *IEEE Trans. Com-puters, vol. 41 , pp.* 616-626, May 1992.

[8] E.P. Duarte Jr. and T. Nanya, A Hierarchical Adaptive Dis-tributed System-Level Diagnosis Algorithm,*IEEE Trans. Com-puters , vol. 47, pp. 34-45,* Jan. 1998.

[9] E.P. Duarte Jr., A. Brawerman, and L.C.P. Albini, An Algo-rithm for Distributed Hierarchical Diagnosis of Dynamic Fault and Repair Events, Proc. Seventh Intl Conf. Parallel and Distributed Systems, pp. 299-306, 2000.