Anomaly Identification & Failure Diagnosis

Mr.Swapnil B.Kadam
[1]*(D.Y.Patil college of engineering, kolhapur, India)*

Abstract: *When wework in large scale network ,number of problems arises, the total time required to deal with these type of problemsdepends on how severe the problemis. As system takes more time to recover from failures, maintenance costgoes on increasing,italso 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. Intraditional system thisprocess of dealing with failureswashandled by humans, but such manual processing was leading to various problems such as 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***.***

Keywords -*Abnormal,Anomaly ,Communication,Detection,Failure, Timer.*

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 consequences of faults in systems could be disastrous in terms of human mortality and environmental impact. To a less extent, fault detection in process and manufacturing industries is also crucial in order to improve production efficiency, quality of the product and cost of production. Today’s computer networks are becoming much larger and more complex. One single fault occurred in one network component might cause considerably high volume of alarms to be reported to network operators, which is called alarm burst.Alarm burst may be a result of fault re-occurrence, Multiple invocations of a service provided by a faulty component,generating multiple alarms by a device for a single fault, detection of and issuing a notification about the same network fault by many devices simultaneously, error propagation to other network devices causing them to fail and, as a result, generate 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 algorithms implemented are based on adaptive distributed system-level diagnosis. The term adaptive indicates that at different stages of algorithm decisions are taken dynamically according to the situation. The term System-level indicates that the system is considered as set of indivisible units called processing elements. Each node in the system works independently and shares information with others by message passing.Existing diagnosis algorithms like ADSD [4] (Adaptive distributed system-level Diagnosis) and Hi-ADSD [5], [6] (Hierarchical Adaptive distributed system-level Diagnosis) compromise minimization of diagnosis latency to reduce network bandwidth consumption. Both algorithms have considered fully connected network topologies.

Performance of a diagnosis algorithm 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 each event (node failure or repair) within a bounded time.
2. Bounded start-up: Recovered nodes must de-termine 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

An important problem in distributed systems that are subject to component failures is the distributed diagnosis problem. In distributed diagnosis, each working node must maintain correct information about the status (working or failed) of each component in the system. Distributed systems are the systems in which hardware or software components of networked computers communicate and coordinate their actions only by message passing. In such systems, it is difficult to predict the behavior of system under various faulty conditions. When systems are working in a faulty environment, it becomes necessary to handle the faults in a graceful manner and keep systems running. To handle faults in such environment it is required to diagnose these nodes. 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. The goal of fault diagnosis research is improving the security, efficiency, maintainability and reliability. A fault is called intermittent if its effects on the system are hidden for discontinuous periods of time. Although a fault is tolerable at the moment it occurs, 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. The cause of an error is called a fault. 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 ar-bitrary 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 processs 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 explicit testing and the crash fault assumption could then be removed. 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.To reduce overhead, the heartbeat algorithm is implemented with multicast, for completely connected networks and with unicast for not completely connected networks.Both mechanisms are implemented over UDP/IP

IV. System Modules

Fig1.below shows 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[1][8][9][10][11].System has been consist of,

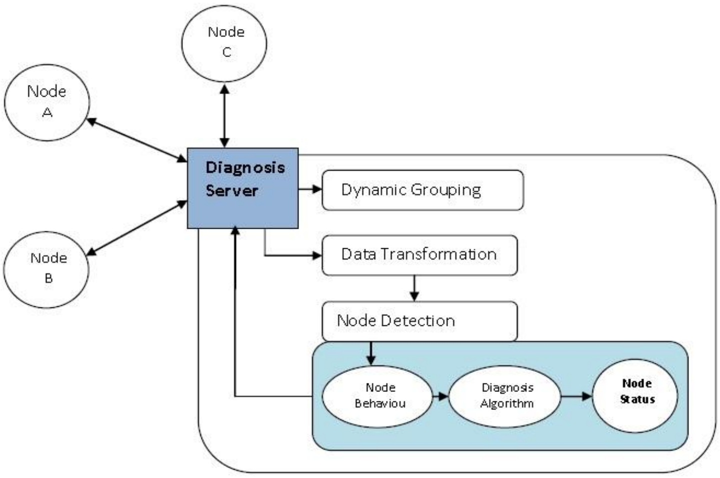

Fig.1.Overview

1. Dynamic Grouping : Grouping of node[1] dynamically can be performed either unicast or multicast broadcasting in a network.It is possible to implement the algorithm using any type of communication to discover a 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 every pair of nodes. This is a requirement to achieve bounded correctness with an arbitrary number of node failures. In not-completely connected networks,intermediate nodes relay messages between some source-destination pairs. Hence,the number of node failures is limited such that the network remains connected at all times.To reduce over head, the grouping algorithm is implemented with multicast, for completely connected networks and with unicast for not completely connected networks.
2. 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[1], the algorithms could be easily converted to ones that use explicit testing and the crash fault assumption could then be removed.3.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.
3. Feature Extraction :A feature extraction[1] 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.
4. Node detection : Node detection [1]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 algo-rithm 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.

V. Experiments & Results

Whenever say node A enters in working state that means the status of this node is working and initiallystatus 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.

VI. Conclusion

When a system fails to function properly, health related data are collected for troubleshooting. However, it is challenging to effectively identify anomalies from the voluminous amount of noisy, hig-dimensional data. The traditional manual approach is timeconsuming, error-prone, and even worse, 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

Journal Papers:

[1] ZhilingLan, Member, IEEE Computer Society, ZimingZheng,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] M Ozaki, Y. Adachi, Y. Iwahori, and N. Ishii, Application of fuzzy theory to writer recognition of Chinese characters, *International Journal of Modelling and Simulation, 18(2),* 1998, 112-116.

[3] 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.*

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

[5] *E.P. Duarte Jr. and T. Nanya, "A Hierarchical Adaptive Distributed System-Level Diagnosis Algorithm", IEEE Trans. Computers, vol. 47, pp. 34-45, Jan. 1998.*

Proceedings Papers:

[6] E.P. Duarte Jr., A. Brawerman, and L.C.P. Albini, "An Algorithm for Distributed Hierarchical Diagnosis of Dynamic Fault and Repair Events", *Proc. Seventh Int'l Conf. Parallel and Distributed Systems*, pp. 299-306, 2000.

[7] 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*

[8] Li Zonglin, Hu Guangmin, Yao Xingmiao Multi-dimensional traffic anomaly detection based on ICA ,*IEEE conference,pp.333-336, 2009.* W.J. Book, Modelling design and control of flexible manipulator arms: A tutorial review, *Proc. 29th IEEE Conf. on Decision and Control*, San Francisco, CA, 1990, 500-506.

[9] A. Hyvarinen and E. Oja, Independent Component Analysis Algorithms and Applications, Neural Networks, v*ol. 13, nos.4/5,pp. 411-430, 2000.*

[10] *Y. Rao and J. Principe, A Fast, On-Line Algorithm for PCA and Its Convergence Characteristics, Proc.IEEE Signal Processing Soc.Workshop , 2000.*

[11] *Anomaly Detection System Based on PCA wuhan university journal of naturascience vol.11 no.06,pp.1769-1772, 2006.*