

Proactive Network Anomaly Detection Using A Statistical Approach

Amar Agrawal, 04329017
M. Tech.-1 Seminar Report, Semester I-2004

Seminar Guide: Prof. Varsha Apte
Department of Computer Science and Engineering, IIT Bombay

November 10, 2004

Abstract

Network anomaly detection is the detection of abnormal conditions in the monitored network either due to a malicious attack such as DoS, network intrusions or non malicious events like equipment failure, improper network configuration, etc. The ability to detect such abnormal conditions before the actual faults occur in the network allows the network administrator to take corrective actions in order to minimize the losses due to the faults. Several approaches for network anomaly detection exist such as traffic modeling, network probes, rule based approaches, etc. One such approach uses statistical techniques to model the network behavior and analyze deviations from the normal behavior in order to detect network anomalies.

1 Introduction

The most commonly used method for network anomaly detection today is to monitor a set of network parameters for threshold violations. Any deviation beyond the threshold is treated as anomalous behavior and correspondingly alarms are generated. However using hard thresholds for alarm generation leads to undetected faults and/or high false alarm rates. Based on this model the following view can be stated : *Network anomalies are characterized by correlated transient changes in measured network data that occur prior to or during the anomalous event.*[3].

A statistical approach to the problem is to use rigorous statistical analysis of network data collected to quantify network behavior. The normal behavior of the network is modeled using signal processing techniques. Abrupt

change detection is then used to detect anomalous behavior patterns in the network traffic data. Anomaly detection is done at device level to obtain the device view of the network health. Information from all such devices is then combined together to obtain the overall network health.

In section 2, a brief overview of other approaches to anomaly detection is presented. The rest of the document discusses about the statistical approach in detail. Section 3 introduces the data sources for the system while section 4 looks into the segmentation mechanism used for obtaining stationary windows. Abrupt change detection is explained in section 5. Section 6 talks about how the variable level alarms can be combined together to generate node level alarms. Section 7 concludes the discussion with fault classification.

2 Anomaly Detection Approaches

2.1 Rule based approaches

The approach is based on expert systems where rules for faulty behavior of network are maintained in a huge database. These rules are used to detect anomalous network behavior. The method is however too slow and requires previous knowledge of network fault conditions in the network. The system is unable to detect unknown faults or faults with varying behavior and as such do not adapt well to evolving network environments.

2.2 Finite State Machines

In this approach a probabilistic finite state machine model is built for known faults using history data. The alarm sequences collected from various parts of the network are modeled as states of the system. The best mapping for the given sequence of alarms is identified. The difficulty with this approach is that the number of states increases with the number of faults and extensive offline learning is required before its deployment.

2.3 Pattern Matching

The pattern matching approach builds a traffic profile for the given network using online learning. The templates thus obtained for various network parameters are used as baselines for comparisons with instantaneous values to obtain an indication of the network state. This approach is dependent on the quality of the network profile generated and might not scale well with the network size.

2.4 Network Probes

Tools such as *ping* and *traceroute* are used to measure various network parameters like end-to-end delay and packet loss. These probes provide instantaneous measure of the network behavior and can be used to detect the network faults. However the approach assumes the existence of symmetric paths between the source and destination and that the probe packets are not differentiated from regular packets which might not always be true. Thus the network behavior modeled may not be accurate.

2.5 Statistical Analysis

Statistical techniques can be combined with on-line learning to continuously monitor the network behavior. Individual nodes in the network are monitored to obtain its view of the network health. The network health function information obtained from all the network nodes is then combined to obtain an indicator for the entire network state.

3 Data Source

Obtaining the right type of network performance data is essential for effective anomaly detection. Management Information Base (MIB) variables provide fine grained data for each network device and hence are ideal data source for network anomaly detection. The MIB variables can be classified into various groups like system interfaces, address translation, internet protocol(*ip*), internet control message protocol(*icmp*), transmission control protocol(*tcp*), user datagram protocol(*udp*), exterior gateway protocol(*egp*), simple network management protocol(*snmp*), etc. Each group of variables describes special functionality of the network. The group of MIB variables to be used depends upon the device being monitored as well as the protocol level at which the device works. No single variable provides complete information about the behavior of the network device and hence the proper choice of a subset of the variables is essential.

For example, if the network device to be monitored is a router which works at the network layer then the group of MIB variable relevant to the router are the *ip* group of variables. Some redundancy exists even within the group and hence proper subset of these must be chosen so that the node information is adequately obtained. For the router the following MIB variables could be chosen

1. *ipIR* - Number of datagrams received by the ip layer of the router.

2. *ipIDe* - Number of datagrams forwarded to the higher layers.
3. *ipOR* - Number of datagrams received from the higher layers.

4 Segmentation

The MIB variables are sampled periodically to obtain a time series which is processed independently using a sequential change point algorithm. The statistical properties of the MIB variables change in response to the network fault. These changes are subtle and must be differentiated from normal traffic variations in the network. The challenge here is to detect these subtle changes that precede any fault conditions in the network inspite of the non-stationary nature of the MIB variables. The non-stationary time series is segmented into piecewise stationary segments and signal processing techniques are used to detect abrupt changes in the statistical properties of the the variables. Within each segment the MIB data is modeled using a linear first order Auto-Regressive(AR) process that takes into consideration the auto-correlation in the time series.

5 Abrupt Change Detection

The input MIB data is sequentially processed by considering time series of each variable over piecewise stationary windows. Using two adjacent windows, the learning window $L(t)$ and the test window $T(t)$ a sequential hypothesis test is performed using the Generalized Likelihood Ratio(GLR) test [6].

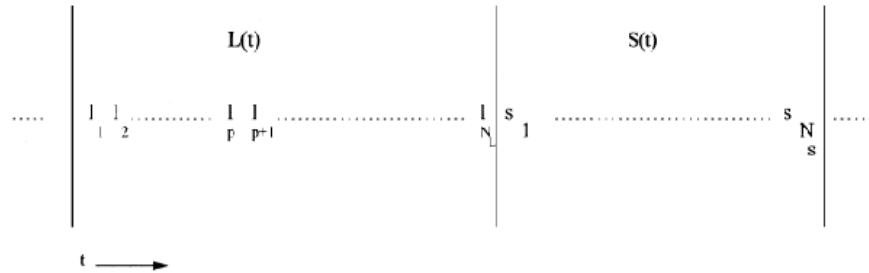


Figure 1: Piecewise stationary windows $L(t)$: Learning window , $T(t)$: test window

The joint likelihood l of the residual errors in the two windows $L(t)$ and $T(t)$ of lengths N_L and N_T respectively is given by

$$l = \left(\frac{1}{\sqrt{2\pi\sigma_L^2}} \right)^{N_L} \left(\frac{1}{\sqrt{2\pi\sigma_T^2}} \right)^{N_T} \exp\left(\frac{-N_L\hat{\sigma}_L^2}{2\sigma_L^2}\right) \exp\left(\frac{-N_T\hat{\sigma}_T^2}{2\sigma_T^2}\right) \quad (1)$$

where

- σ_L^2, σ_T^2 : Variance of the residuals in windows L(t) and T(t)
- $\acute{N}_L = N_L - p, \acute{N}_T = N_T - p$: Covariance estimates of σ_L^2 and σ_T^2 .

A binary hypothesis test is performed using the statistic l with
 H_0 : No change is observed between the two windows

$$l_0 = \left(\frac{1}{\sqrt{2\pi\sigma_P^2}} \right)^{N_L + \acute{N}_T} \exp\left(\frac{-(N_L + \acute{N}_T)\hat{\sigma}_P^2}{2\sigma_P^2}\right) \quad (2)$$

H_1 : Change is observed between the two windows

$$l_1 = l \quad (3)$$

In order to obtain the value of likelihood ratio η between [0 1] we define η as

$$\eta = \frac{l_1}{l_1 + l_0} \quad (4)$$

Using maximum likelihood estimates(MLEs) for the variance terms in equation (1) and (2)

$$\eta = \frac{\hat{\sigma}_L^{-N_L} \hat{\sigma}_T^{-\acute{N}_T}}{\hat{\sigma}_L^{-N_L} \hat{\sigma}_T^{-\acute{N}_T} + \hat{\sigma}_P^{-(N_L + \acute{N}_T)}} \quad (5)$$

The value of η thus obtained is the abnormality indicator for the MIB variable and is updated every N_T lags. The length of the learning window must be optimized according to the MIB variable under consideration.

6 Spatial Correlation Using Combiner

The abnormality indicators thus obtained for each individual MIB variable are collected to form an abnormality vector $\vec{\psi}(t)$ which is a measure of the abrupt changes in the normal network behavior. A network health indicator is obtained from the abnormality vectors $\vec{\psi}(t)$ by incorporating the spatial dependencies between the abrupt changes in the individual MIB variables. A linear operator based on the correlation between the chosen MIB variables is used to generate a continuous scalar indicator of the network health.

$$f(\vec{\psi}(t)) = \vec{\psi}(t)A\vec{\psi}(t) \quad (6)$$

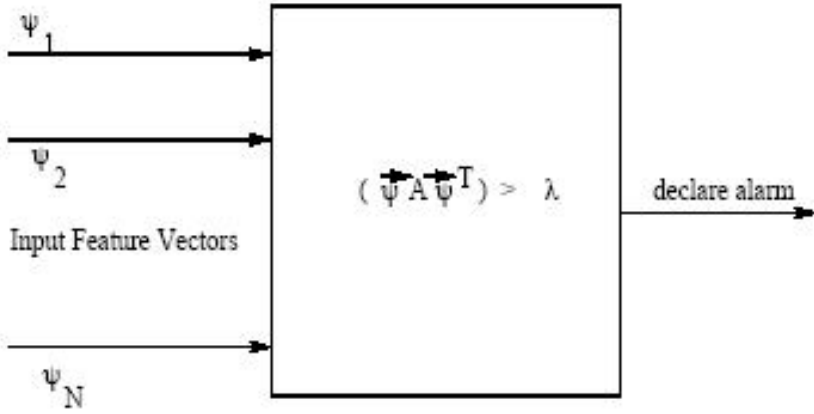


Figure 2: Discriminant function scheme for N inputs using the linear operator A.

In order to complete the basis set so that all the states of the system are represented, an additional component ψ_0 is added that corresponds to the probability of normal functioning of the network.

$$\vec{\psi} = \alpha[\psi_1 \dots \psi_n \psi_0] \quad (7)$$

The new input vector is normalized using α as the normalization constant.

The elements a_{mn} of the operator matrix are assigned based on the cross correlation between the variables so that the element a_{ij} gives the correlation between the i th and the j th MIB variable. The matrix is symmetric so that $a_{21} = a_{12}$, $a_{23} = a_{32}$, etc. The main diagonal values are chosen such that the sum of rows and columns is 1. Taking the normal state to be uncoupled

from the abnormal states we get a block diagonal matrix with a $M \times M$ upper block A_{upper} and 1×1 lower block.

The linear operator A is designed to be Hermitian so that the corresponding eigenvectors are orthogonal with real eigen values. The eigenvectors are normalized to form an orthonormal basis set. The first M components of $\vec{\psi}(t)$ can be represented as linear combination of the eigenvectors of A_{upper}

$$A_{upper}\vec{\phi} = \lambda\vec{\phi} \quad (8)$$

$$\vec{\psi}(t) = \sum_{i=1}^M c_i \vec{\phi}_i \quad (9)$$

Incorporating the spatial dependencies using the operator A the abnormality vector is transformed as

$$A_{upper}\vec{\psi}(t) = \sum_{i=1}^M c_i \lambda \vec{\phi}_i \quad (10)$$

As there are M different values corresponding to M variable components, an average scalar value of the transformation is obtained by using the function

$$f(\vec{\psi}(t)) = \vec{\psi}(t)A\vec{\psi}(t) \quad (11)$$

$$= \sum_{i=1}^M c_i^2 \lambda_i \quad (12)$$

$$= E(\lambda) \quad (13)$$

Since the abnormality vector $\vec{\psi}(t)$ is normalized the following condition holds

$$\sum_{i=1}^M c_i^2 = 1 \quad (14)$$

Using this condition a bound for $E(\lambda)$ can be obtained

$$\min_{\lambda_i \in \{\lambda_1, \dots, \lambda_N\}} (\lambda_i) \leq E(\lambda) \leq \max_{\lambda_i \in \{\lambda_1, \dots, \lambda_N\}} (\lambda_i) \quad (15)$$

The measure $E(\lambda)$ is the measure of the average abnormality in the network as perceived by the node. Using these bounds on the functional $f(\vec{\psi}(t))$ an alarm is declared when

$$E(\lambda) > \min_{\lambda_i \in \{\lambda_1, \dots, \lambda_N\}} (\lambda_i) \quad (16)$$

7 Fault Classification

Identifying the fault after an alarm is declared is equally important. For this purpose the MIB data from the network to be monitored is investigated and the behavior of the abnormality vectors prior to the faults is observed. The faults tend to cluster together in the fault space and hence, using prior knowledge of the average values for the abnormality vectors for known faults, these can be identified.

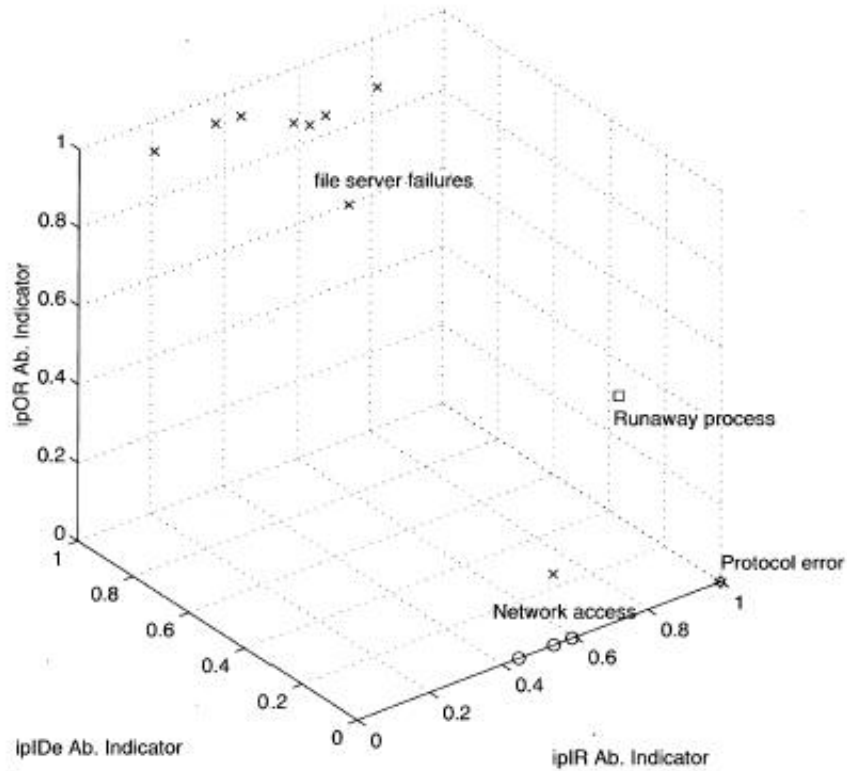


Figure 3: Classification of faults using average of the feature vectors.

8 Conclusion

The statistical approach for anomaly detection maintains a statistical usage profile of the network. It uses this profile to detect abnormal behavior in the network as statistically significant deviations from the normal. The network profile is continuously updated with new data collected as part of the change detection process. Thus the system adapts well to changes in the network behavior. The use of likelihood ratios and hypothesis test instead of hard thresholds greatly improves the detection sensitivity of the system and reduces false alarm rates. When used with other detection approaches, it can provide a robust network anomaly detection tool that is able to detect network anomalies proactively.

References

- [1] Kishor S. Trivedi. *Probability and statistics with reliability, queuing and computer science applications*. John Wiley and Sons Ltd., New York, 2001.
- [2] William Hines et al. *Probability and Statistics in Engineering, 4th edition*. John Wiley and Sons Ltd., Singapore, 2003.
- [3] Chuanyi Ji Marina Thottan. Anomaly detection in ip networks. *IEEE Transactions on Signal Processing*, 51(8):2191–2204, 2003.
- [4] M. Thottan and C. Ji. Using network fault predictions to enable ip traffic management. *J. Netw. Syst. Manage.*, 9(3):327–346, 2001.
- [5] Chuanyi Ji Marina Thottan. Fault prediction at the network layer using intelligent agents. In *IFIP/IEEE International Symposium on Integrated Network Management*, pages 745–759. IEEE, 2000.
- [6] M. Thottan and C. Ji. Adaptive thresholding for proactive network problem detection. In *Proceedings of the IEEE Third International Workshop on Systems Management*, page 108. IEEE Computer Society, 1998.