Journal of Computer Applications ISSN: 0974 – 1925,
Volume-5, Issue EICA2012-5, February 10, 2012

A Levy Alpha Stable Model for Anomaly Detection in Network Traffic

Diana A
Dept of IT,
KalasalingamUniversity,
Tamilnadu, India
E-mail: arul.diana@gmail.com

Mercy Christial T
Asst. Prof I/IT,
Dept of IT,
Kalasalingam University,
Tamilnadu, India

Abstract-The detection of anomaly in real time traffic is challenging and important task. The statistical method tracks the marginal distribution of traffic flow which is controlled by several driving factors and complicated behavior. This paper proposes the alpha stable model and statistical hypothesis to analyse the traffic pattern. Since the model is a parametric approach which includes the skewness, scatter, centrality and shape as parameters in detecting the traffic windows as anomaly. To classify the traffic patterns by means of GLRT Generalized Likelihood Ratio Test which is used to automatically choose real traffic window to compared with reference under test , with no human intervention .The use of the GLRT is to assess the similarity of a particular trace to reference normal and anomalous traffic windows. Two attacks are focused in this paper namely Flood and Flash crowd . Flood anomalies distinct as having one or more relatively constant sources(DDos attacks). Where as , Flash Crowd deals with the encompass traffic patterns caused by net growth of users trying to access the network resources(web fileserver).Each anomaly detection performance have been measured through Receiver Operating Characteristic (ROC) Curve. The goal is to achieve the self adaptation of reference traffic windows to a new real traffic. And also, the network traffic that tends to vary over time, it may be desirable that training traffic is periodically updated to fit new circumstances and recalculate the reference window.

Keywords: *Alpha Stable , GLRT, Flood, Flashcrowds, ROC.*

I. INTRODUCTION

Anomaly detection aims at finding the presence of anomalous patterns in network traffic. Automatic detection of such patterns can provide network managerial with an additional source of information to diagnose network behavior or finding the root cause of network faults. Internet is today the fundamental component of the worldwide communication infrastructure, playing a crucial role in education, entertainment, business, and social life.

Traffic monitoring is certainly one of the paramount tasks for network operators that will be affected by the strong development of network traffic, simply because capturing and analyzing large volumes of heterogeneous traffic network-wide can be extremely costly. Network and traffic anomalies may arise from equipment failures, misconfigurations, and outages, unusual customers behavior (e.g., sudden changes in demand, flash crowds, high volume flows), external routing modifications, network attacks (e.g., DOS attacks, scans, worms), Four stages of anomaly detection in traffic

1.1Data Acquisition

In the data acquisition stage, Data collection is done by querying the routers via SNMP periodically for accumulated byte counters at each physical port. These routers should be representative of heavily and lightly loaded networks.

The traffic samples are taken at intervals of t seconds, so that data windows of W seconds are continuous. W should be large enough to have a minimum amount of data when trying to fit a statistical model to them, and short enough to have (at least) local stationarity.To ensure the model adequately fits the data ,so chose a time window length W = 30 minutes.

1.2Data Analysis

With traffic windows of W/t = 360 samples each, the sample size is rather small, so the use of α-stable model, the use of α-stable distributions, unrestricted in their parameter space, as a model for traffic marginals is desirable.

α-stable distributions : It is characterized by four parameters.The first two of them, (α) and (β), provide the properties of heavy tails (α) and asymmetry(β), while the remaining two, (σ) and (μ), have analogous senses scatter and center. The allowed values for α lie in the interval (0; 2), being α= 2 ,while β must lie inside[-1,1] (-1 means totally left-asymmetric and 1totally right-asymmetric). The scatter parameter (σ) must be a Nonzero positive number and (μ) can have any real value. If (α=2), the distribution does not have heavy tails .

1.3Inference

A common approach is to define anomalies as a sudden change in any quantity measured in the analysis phase or a significant deviation between the current traffic window and a earlier chosen reference to fix an arbitrary threshold which marks the boundary between normal and anomalous traffic, and trigger an alarm when the threshold is exceeded. So,the use of sets of synthetic anomalies (one set per anomaly type),analogous to the normal traffic windows but known to be anomalous, and setting the threshold so that a given traffic window is classified as normal or anomalous based on its similarity to the normal and anomalous sets and informs via 'abnormality index'.

A hypothesis, p(normal traffic) >> p(anomalous traffic) in any correctly behaved network at any circumstance. For these windows, estimate the parameters of an α -stable PDF which fits the data and store those parameters in catalogs which test windows are compared with stored training windows within the corresponding catalog.

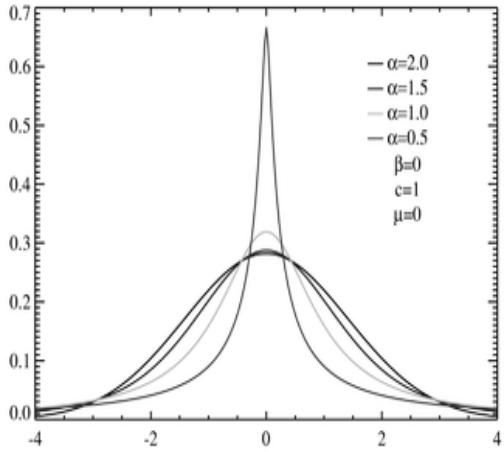

Figure 1 :Probability Density Function with Different α ValuesTwo Anomalies to Detect

Flood anomalies : attacks, which result in a net growth of instantaneous traffic DDoS attacks typically give rise to anomalies

Flash-crowd anomalies : traffic patterns, caused by a net growth of (usually human) users trying to access a network resource. overwhelming web server usage patterns
Generalized Likelihood Ratio Test (GLRT): The classification algorithm as a parametric test.

The α-stable model there is no optimality associated with the GLRT, asymptotically it can be uniformly most powerful (UMP) among all tests that are invariant .To determine whether the current traffic window arrives from the normal set or one of the anomalous sets, Let H0: current traffic is normal versus H1: current traffic is not normal.The GLRT will decide H1, when

$LG(x) = \frac{p(\max 1j, \theta 1j; H1)}{p(\max 0j, \theta 0j; H0)} \geq \varepsilon$ where x is the vector of traffic samples in the current window, ε is the chosen threshold, and $\theta 0j$, $\theta 1j$ are, respectively, the normal and anomalous sets of α-stable parameter

1.4Validation

Classifier and Performance

1. For all collected traffic windows belonging to a particular combination of port, hour, and weekday,repeat steps 2 to 8.
2. Prepare three consecutive traffic windows. The first one is the reference window; second and third onesact as normal and anomalous traffic windows.
3. Inject a synthetic flood anomaly of a given intensity into the third window.
4. For time resolutions 5, 10, 20, 40, and 80 seconds per sample, repeat step 5.
5. Estimate the shape parameter of three Gamma distributions, one fitted to each traffic window, and compute the following quadratic distances: a) reference to normal windows and b) reference to anomalous windows.

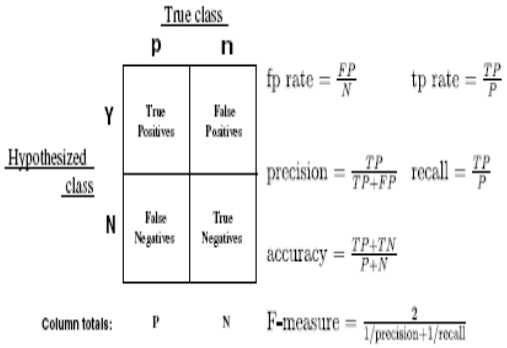

Table 1:Confusion Matrix and Performance Metrics

6. Calculate mean quadratic distances over all time resolutions.
7. For a sufficiently dense, logarithmically spaced set of thresholds from 0 to þ1, repeat step
8. Accumulate number of false positives/negatives for each threshold.
9. Calculate false positive/negative ratios.
10. Plot ROC curve and calculate its AUC.

First, the data are sampled at 5-second intervals, so no other option than making the multiresolution calculations at a larger scale. Second, choose the window preceding normal and anomalous ones as reference traffic. Third, do as to fairly compare classification performance between approaches.

II. CONCLUSION

In this paper, an anomaly detection method based on statistical inference and an α-stable first-order model has been studied. a four-phase approach to describe each of the pieces from which to build a functional detection system (data acquisition, data analysis, inference, and validation), yielding the final classification results the uses of aggregated traffic as opposed to packet-level sampling so that any dedicated hardware is not needed.In the data analysis stage, use α-stable distributions as a model for traffic marginals. Since these distributions are able to adapt to highly variable data, and due to the fact that they are the limiting distribution of the generalized central limit theorem, they are a natural candidate to use in modeling aggregated network traffic.

REFERENCES

1. A. Scherrer, N. Larrieu, P. Owezarski, P. Borgnat, and P. Abry, “Non-Gaussian and Long Memory Statistical Characterizations for Internet Traffic with Anomalies,” IEEE Trans. Dependable and Secure Computing, vol. 4, no. 1, pp. 56-70, Jan. 2007.
2. C. Manikopoulos and S. Papavassiliou, “Network Intrusion and Fault Detection: A Statistical Anomaly Approach,” IEEE Comm. Magazine, vol. 40, no. 10, pp. 76-82, Oct. 2002.
3. S. Stolfo et al., “The Third International Knowledge Discovery and Data Mining Tools Competition,” http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html, 2011
4. A. Lakhina, M. Crovella, and C. Diot, “Diagnosing Network-WideTraffic Anomalies,” Proc. ACM SIGCOMM ’04, pp. 219-230, Aug.2005.

5. A. Papoulis, Probability, Random Variables, and Stochastic Processes,third ed., McGraw-Hill, 1991.
6. P. Barford, J. Kline, D. Plonka, and A. Ron, "A Signal Analysis ofNetwork Traffic Anomalies," Proc. Second ACM SIGCOMMWorkshop Internet Measurement, pp. 71-82, Nov. 2002.
7. P. Embrechts and M. Maejima, Selfsimilar Processes. Princeton Univ. Press, 2002.
8. S"ApacheJMeter," The Apache Jakarta Project, Apache Software Foundation, http://jakarta.apache.org/jmeter/, 2011.
9. G.F. Cretu-Ciocarlie, A. Stavrou, M.E. Locasto, and S.J. Stolfo,"Adaptive Anomaly Detection via Self-Calibration and Dynamic Updating," Proc. 12th Int'l Symp. Recent Advances in Intrusion Detection (RAID), Sept. 2009.
10. Cisco Systems, "Cisco IOS NetFlow," http://www.cisco.com/web/go/netflow, 2011.