Comparing Anomaly Detection Methods in Computer Networks

Andreas Löf
Department of Computer Science
The University of Waikato
Hamilton, New Zealand
andreas.lof@cs.waikato.ac.nz

Richard Nelson
Department of Computer Science
The University of Waikato
Hamilton, New Zealand
richardn@cs.waikato.ac.nz

***Abstract*—This work in progress outlines a comparison of anomaly detection methods that we are undertaking. We are comparing different types of anomaly detection methods with the purpose of achieving results covering a broad spectrum of anomalies. We also outline the datasets that we will be using and the metrics that we will use for our evaluation.**

***Keywords*-Network fault diagnosis; Computer network security;**

I. Introduction

Recent years have seen a multitude of new anomaly detection methods for computer networks. Many of these methods specialise in detecting only a few different types of anomalies and have therefore been evaluated against methods detecting the same anomaly types. We are currently undertaking a broad comparison of anomaly detection methods. This comparison focuses on anomaly detection methods from several different approaches. A range of methods representative of their approaches have been selected. These methods will be evaluated using several different publicly available datasets. The evaluation of the methods will establish overall performance of the different methods and establish their strengths and weaknesses in terms of the types of anomalies they can detect. The primary aim of this work is to establish a base set of methods and their evaluations to which other methods can be added and reliably compared in the future.

A. Motivation

There is a lot of research going into inventing new or improving upon existing anomaly detection methods. While new methods have normally been compared to previous work, this has frequently been a comparison only to other methods of the same type, primarily the older methods that they extend. There are two main reasons for this. The first is that there a lack of standardised implementations of the methods that can be used in the evaluations. It is difficult to find published code for anomaly detection and our enquiries have revealed few authors willing to share their code with us. The second major problem in the field is that there are no common datasets evaluation of anomaly detection methods beyond the Defense Advanced Research Projects Agency (DARPA) datasets [1] and these have well known problems. Most evaluations of new methods use the DARPA datasets together with a second, different dataset. However due to privacy reasons, these secondary sets are not normally able to be shared and are therefore different between each evaluation.

To alleviate the first problem, we are performing a comparison of anomaly detection methods in networks. Methods from a range of different approaches will be evaluated against each other in the comparison. The goal is to make the framework and the implementations public to the research community to facilitate easier comparisons.

To alleviate the second problem we will use publicly available datasets, and if new datasets are required, we will make every effort to ensure it can be freely published.

B. Paper Organisation

The following section outlines the related work to this paper. Section III contains a definition of the anomalies we use, an outline of the methods that is used for the comparison, the data sets that will be used and a brief explanation of the evaluation criteria. The final section of the paper provides a brief view of where we intend to focus our future research.

II. Related Work

There have been other comparisons of network anomaly methods. These studies focus on comparing a few similar methods.

Cottrell et al. [2] compared different time series forecasting methods to detect anomalies in end to end bandwidth. The data that they used was obtained through active measurements and they were looking for anomalies in available bandwidth over end-to-end links. The paper is valuable because of the comparison of forecasting methods and shows that the forecasting methods used are applicable to several different traffic features.

Lazarevic et al. [3] surveyed several clustering based anomaly detection methods and a support vector machine classifier. They evaluated their data on the DARPA 98 dataset and real time internet traffic. The real time traffic was labelled using Snort. The comparison does, however, focus on only one type of anomaly detection method and

978-0-7695-4023-8/10 $26.00 © 2010 IEEE
DOI 10.1109/ICIMP.2010.9

does not explore the limitations of the methods nor what they are best suited for.

Cárdenas et al. [4] created a framework for evaluating intrusion detection systems. The paper suggests using a different measure to compensate for the skewed distributions of the problem. We will draw upon their experiences in our evaluation.

We focus on doing a broader comparison of different types of anomaly detection methods that detect different types of anomalies.

III. Comparison of Anomaly Detection Methods

Methods will be classified based on their ability to detect different types of events and the performance of the method. Computational performance will not be taken into account in this comparison but the ability to detect anomalies in near real time is requirement for the longer term aims of our research project.

Every method will be evaluated on several datasets and attempts will be made to establish what the different methods are most suitable for.

A. Anomaly Classification

In this paper, we focus on two different types of events, network faults and security anomalies. Network faults indicate an error in the network, while security anomalies indicate that one or more sources are adversely affecting the network in a negative manner. We make this distinction because the different anomaly detection methods that are investigated are demonstrated as more suitable for different types of anomalies by their authors.

1) Network Faults: Network faults are considered to be non-malicious in their origin but can still have an adverse effect on the network.

We divide network faults into two main categories: hardware failures and configuration errors.

Hardware Faults: Hardware faults cause an abrupt cessation of traffic over the affected area. Examples are faulty network interfaces or a physical link being severed.

Configuration Errors: Configuration errors are more difficult to detect than hardware faults. They are generally caused by the operator choosing an improper setting for some or all of the networking equipment. This can cause symptoms such as higher-than-normal latencies, oversaturated links, underused links and inefficient routing.

2) Security Anomalies: Security anomalies are malicious in nature. No distinction is made between Internet background radiation [5] and specific attacks or probes [6].

Security anomalies can be divided into two types, probes and attacks.

Probes: Probes are attempts to map the network or the services running on a specific node in a network. A well known example of these are port scans.

Attacks: Attacks attempt to disrupt one or more nodes on the network or attempt to saturate the links in the network itself. Examples are worms [7] and Denial of Service attacks [8].

B. How the Methods Were Chosen

Only methods using data on a flow level or higher are included in this comparison. Flow level means that we look at the network flows as the most fine grained input data for the anomaly detection methods. Higher levels of data mean either aggregation of collected data, for example traffic volumes, or metrics extracted from a series of flows.

Methods that examine individual packets tend to rely on heuristics or signatures and are extremely computationally expensive on fast links. This makes such methods unsuitable since it will be difficult to use those methods in near real time detection. An example of this is Snort's hardware requirements to keep up with a 1 gigabit link.

In addition to this, only data from passive measurements is used. It is possible that active measurement data might be included in the future. Passive measurement means that traffic has been captured at a critical link in a network, normally where it will provide a good view of the behaviour of the network.

A small subset of all of the available methods have been chosen due to practical reasons. We have attempted to choose at least two methods using approaches from the same field. The methods chosen from each field are expected to detect the same types of anomalies. Since several of the methods will be able to detect both security anomalies and network faults, the methods will be classified according to the different types of anomalies that they can detect.

The methods chosen in each category had to have a good description and be straightforward to implement and have published performance results. Methods were preferred if they had been the basis for modifications by other researchers.

C. Methods Chosen

Out of the vast multitude of available anomaly detection methods only a small subset has been chosen for the comparison. The aim of the comparison is not to be exhaustive but rather to establish a base of standardised implementations of anomaly detection methods. Future research can then add to the established base.

The following methods will be included in the comparison. Further methods might be added during the course of the comparison.

1) Forecasting Methods: Brutlag [9] introduced using the Holts Winters Exponential forecasting method in networks. The method is good at detecting volume anomalies while still adapting to the diurnal trends in a network.

Soule et al. [10] uses Kalman filters to forecast the network traffic using traffic matrices.

Currently only a simple test based on standard deviation is used to flag events. More sophisticated tests will be implemented in the future.

Based on the authors' own evaluations of these methods, they can be expected to detect hardware faults and configuration errors in the network and attacks upon the network. Probes that does not significantly affect the network might not be detected by these methods.

2) Signal Analysis Methods: Kim and Reddy [8] created a technique that looks at the frequency of IP addresses in a trace over an egress router. Their method combines correlation computations with a wavelet transform to create data in which they then look for outliers. The detector that they use is a simple threshold based detector based on standard deviation with a trigger buffer.

Kim and Reddy's evaluation of the method shows that it deals well with both traffic anomalies and attacks upon the network.

We are still searching for another suitable method in this category.

3) PCA based methods: Lakhina et al. [11] introduced a new set of Principal Component Analysis (PCA) based anomaly detection methods that we will use in our comparison.

Li et al. [12] created a method that relies on sketch subspaces to detect anomalies. The method can detect individual flows that are anomalous.

The methods show a good ability to detect network faults successfully and some security anomalies.

4) Clustering: Breunig et al. [13] created a density based clustering method. Lazarevic et al. included it in their survey and it was the best overall clustering method.

A second clustering algorithm will be used but it is not decided which at the moment of writing.

5) Other Methods: Eimann et al. [14] have successfully used T-entropy to detect network anomalies. The method extracts an entropy measure across various attributes in a network flows and creates an aggregate number for the network that changes during anomalies.

This method is the only exception to the criteria mentioned in the previous subsection. The method was chosen because the source code was available to us and we think that the approach is sufficiently different from most other approaches to warrant our attention.

D. Data sets

These are the datasets that we use in the comparison of the different anomaly detection methods.

1) WITS Archive: The WITS archive [15] contain a collection of traces captured at the University of Auckland and University of Waikato in New Zealand. The traces have been captured with DAG cards developed by Endace [16]. The network users' privacy anonymising the addresses and truncating packets.

2) Internet2 Observatory: The Internet2 observatory [17] provides a complete view of the core Internet2 network in the U.S.A. They do however not provide full capture data like the two other datasets that we use. Instead, they capture 1% of the data that flows through the network and complete NetFlow data.

User privacy is ensured by only retaining 21 bits of the address fields.

Because this dataset does not contain a full passive capture, it is possible that some methods will not perform as well on this dataset as they would on a full dataset. The dataset is still valuable since it contains captures from several core routers.

3) DARPA Datasets: There are three DARPA datasets created in 1998, 1999 and 2000. They are artificially created datasets that contain synthetic background data and real attacks against hosts. They are the only fully labelled datasets we have, but there has been criticism put forward towards the dataset and it is 10 years old.

4) Possible New Datasets: We are investigating the possibility of creating a new passive dataset using available network capture facilities. We will label the data with Snort [18], which has become a de facto standard, before anonymising the dataset. The ambition is that the wider community can use this dataset to detect both security anomalies and network faults and that it will be adopted in future evaluations.

E. Evaluation

We will use the labelled DARPA datasets to establish the Receiver Operating Characteristics (ROC) of the different anomaly detection methods. ROC curves are useful for determining how effective a classifier is without taking the class distribution into account. Area Under the ROC Curve (AUC) is a very useful metric [19] of how well an algorithm performs. Unlike accuracy, the AUC does not depend of the distribution of the data. However, labelled data is necessary to establish these two metrics.

The two unlabelled datasets will be used for a comparative evaluation and each detected anomaly will be examined by an expert to establish whether it is a true positive or a false positive. We will try to avoid false negatives by running the methods on as high sensitivity as possible before applying more optimal settings. It is however unlikely that all anomalies will be detected since the other datasets are not synthetic.

All of the methods examined will be classified according to the anomalies they can detect. They will be classified according to the types of anomalies that we established earlier in this section. Some methods can also work on more than one type of in data. An example of this is Holts Winters forecasting algorithm. Brutlag introduced it in a paper where he detected volume anomalies whereas Cottrell et al. used

it to detect changes of available bandwidths over paths in networks.

Methods that can take different types of traffic features as their input will be evaluated on the features suggested in literature.

IV. Future Work

We will continue to add methods to our comparison framework until we have a suite of anomaly detection methods that can detect different categories of anomalies with a certain overlap. We will then investigate different approaches of data fusion to improve on the original results derived from the methods themselves.

Acknowledgment

This work is partially funded by New Zealand FRST project "UOWX0705 Autonomous Networks". It has also benefited from the use of measurement data collected on the Internet2 network as part of the Internet2 Observatory Project.

References

[1] MIT Lincon Laboratory, "DARPA Intrusion Detection Challenge," 1998, last accessed on 2010-02-22. [Online]. Available: www.ll.mit.edu/mission/communications/ist/corpora/ideval

[2] R. Cottrell, C. Logg, M. Chhaparia, M. Grigoriev, F. Haro, F. Nazir, and M. Sandford, "Evaluation of Techniques to Detect Significant Network Performance Problems using End-to-End Active Network Measurements," in *IEEE/IFIP Network Operations and Management Symposium NOMS*. IEEE, 2006, pp. 85–94.

[3] A. Lazarevic, A. Ozgur, L. Ertoz, J. Srivastava, and V. Kumar, "A comparative study of anomaly detection schemes in network intrusion detection," in *In Proceedings of the Third SIAM International Conference on Data Mining*. Philadelphia, PA, USA: SIAM, 2003, pp. 25–36.

[4] A. A. Cárdenas, J. S. Baras, and K. Seamon, "A Framework for the Evaluation of Intrusion Detection Systems," in *Security and Privacy, IEEE Symposium on*. Los Alamitos, CA, USA: IEEE Computer Society, 2006, pp. 63–77.

[5] R. Pang, V. Yegneswaran, P. Barford, V. Paxson, and L. Peterson, "Characteristics of internet background radiation," in *Proceedings of the 4th ACM SIGCOMM conference on Internet measurement*. New York, NY, USA: ACM, 2004, pp. 27–40.

[6] P. Barford and D. Plonka, "Characteristics of network traffic flow anomalies," in *Proceedings of the 1st ACM SIGCOMM Workshop on Internet Measurement*. New York, NY, USA: ACM, 2001, pp. 69–73.

[7] S. Singh, C. Estan, G. Varghese, and S. Savage, "Automated worm fingerprinting," in *Proceedings of the 6th conference on Symposium on Opearting Systems Design & Implementation*. Berkeley, CA, USA: USENIX Association, 2004, pp. 4–4.

[8] S. S. Kim and A. L. N. Reddy, "Statistical techniques for detecting traffic anomalies through packet header data," *IEEE/ACM Transactions on Networking*, vol. 16, no. 3, pp. 562–575, 2008.

[9] J. D. Brutlag, "Aberrant Behavior Detection in Time Series for Network Monitoring," in *Proceedings of the 14th USENIX conference on System administration*. Berkeley, CA, USA: USENIX Association, 2000, pp. 139–146.

[10] A. Soule, K. Salamatian, A. Nucci, and N. Taft, "Traffic matrix tracking using Kalman filters," *SIGMETRICS Performance Evaluation Review*, vol. 33, no. 3, pp. 24–31, December 2005.

[11] A. Lakhina, M. Crovella, and C. Diot, "Diagnosing network-wide traffic anomalies," in *Proceedings of the 2004 conference on Applications, technologies, architectures, and protocols for computer communications*. New York, NY, USA: ACM, 2004, pp. 219–230.

[12] X. Li, F. Bian, M. Crovella, C. Diot, R. Govindan, G. Iannaccone, and A. Lakhina, "Detection and identification of network anomalies using sketch subspaces," in *Proceedings of the 6th ACM SIGCOMM conference on Internet measurement*. New York, NY, USA: ACM, 2006, pp. 147–152.

[13] M. M. Breunig, H.-P. Kriegel, R. T. Ng, and J. Sander, "LOF: identifying density-based local outliers," *ACM SIGMOD Record*, vol. 29, 2000.

[14] R. Eimann, U. Speidel, N. Brownlee, and J. Yang, "Network Event Detection with T-Entropy," University of Auckland, Tech. Rep., 2005.

[15] Waikato Internet Traffic Storage, "WITS Website," last accessed on 2010-02-22. [Online]. Available: www.wand.net.nz/wits

[16] Endace, "Endace Website," last accessed on 2010-02-22. [Online]. Available: www.endace.com

[17] Internet2 Observatory, "Internet2 Observatory Website," last accessed on 2010-02-22. [Online]. Available: www.internet2.edu/observatory

[18] Sourcefire, "Snort Website," last accessed on 2010-02-22. [Online]. Available: www.snort.org

[19] A. P. Bradley, "The use of the area under the ROC curve in the evaluation of machine learning algorithms," *Pattern Recognition*, vol. 30, pp. 1145–1159, 1997.