Multidimensional and Adaptive Non-Intrusive Anomaly Detection in Network Services

Kurnia Hendrawan, Thomas Sinnwell, Dirk Leinenbach

***Abstract*—We propose the analysis of signal-based metrics with a combination of wavelet transform and Mahalanobis distance to automatically detect anomalies in network services. In contrast to conventional detection methods like thresholding, our technique adapts automatically to gradual changes in the measured signals and deals well with periodical load patterns. It supports multidimensional analysis to improve reliability and significance of the detection and provides confidence values, which are the base for judging the anomaly.**

***Keywords*—Anomaly detection, Mahalanobis distance, networks, wavelet**

I. Introduction

Performance monitoring and anomaly detection is critical for todays complex network services. However, conventional detection mechanisms like thresholding for single parameters do not deal well with distributed services where many different performance indicators might influence the overall performance of the combined system. Estimating good thresholds for single performance values is difficult, especially in situations of rapidly varying request rates.

Distinguishing a harmless increase in request rates from failures or active attacks is difficult. Braford et al. [1] classify network traffic anomalies from the causal point of view into three sets: 1) flash crowd events, i.e., sudden increase in the network service request rate caused by changes in the public interest (e.g., software releases), 2) network operation anomalies, e.g., failure of one mirror server followed by higher load on the remaining ones, and 3) network abuse, e.g., denial of service (DoS) attacks. Flash crowd events [2] are often misinterpreted as DoS attacks. Upon wrong classification, protection systems might shut down servers even if they were capable of handling the requests [3].[1]

In this paper, we describe a technique that is able to detect such events with high confidence and to distinguish them from DoS attacks. We analyze signal-based metrics in the networks with the help of wavelet transform [4] and Mahalanobis distance [5], [6] to reveal hidden anomalies in the signal. The method supports *multidimensional* analysis, i.e., is not restricted to a single performance indicator.

A *static* anomaly detection system would potentially produce many false positives due to evolution of the network or (periodical) changes in the request rate [7]. In contrast, an *adaptive* method should be able to react within short notice to changes in the normal behavior of a system. Our method does so by relating the observed (current) values with recent values from a sliding reference window without requiring an explicit learning phase.

The rest of the paper is organized as follows. The next section compares our method with related work. Section III describes the design and implementation including the wavelet transform and sliding windows method with the Mahalanobis distance while Section IV evaluates the applicability of our technique. Section V concludes.

II. Related Work

Most current studies in anomaly detection deal with intrusion detection and network abuse. Using a *statistical* approach, the following papers observe the activity of subjects and generate profiles to represent their behavior. Haystack [8] is based on behavioral constraints imposed by security policies and on models of typical behavior of users. NIDES [9] uses statistical algorithms for anomaly detection and an expert system to encode known intrusion scenarios. The tool SPICE [10] maintains a statistical model of the total activity on the network and uses nested self-balancing binary trees in order to approximate how anomalous a given packet is. Ye et al. [11] use Hotelling's T^2 test to detect counter-relationship and mean-shift anomalies.

Machine-learning based techniques try to build a system that improves its performance based on previous results. Forrest et al. [12] and Eskin et al. [13] compare actual traces of system calls with normal behavior to detect intrusion. Valdes et al. [14] and Shyu et al. [15] also use machine-learning algorithms to analyze network traffic.

With *data-mining* techniques, anomalies can be distinguished from common traffic by identifying bounds of valid network activities. Lee et al. [16] use association rules algorithm and frequent episodes algorithm to compute the intra- and inter-audit record patterns (program / user behavior). The intrusion detection system (IDS) FIRE [17] creates fuzzy input parameters by selecting specific features in the raw input data to provide strong indication of anomaly in a network. The ANDSOM module [18] of the INBOUNDS IDS detects anomalous network traffic based on *self-organizing maps* which are trained with normal traffic. The Minnesota Intrusion Detection System (MINDS) [19] assigns a *local outlier factor* to each data point and summarizes anomalous connections using association rules.

Some techniques use *signal processing* to isolate anomalous patterns that match certain frequency spectrums. Thottan et al. [20] developed a technique very similar to wavelet transform using a linear operator to produce confidence values over a vector of abnormality indicators gained with the *generalized likelihood ratio*-test. Huang et al. [21] tried to use the Haar wavelets to detect periodic components in the network traffic, which might indicate certain performance

All authors are with consistec Engineering & Consulting GmbH, Science Park 1, D-66123 Saarbrücken, Germany (phone: +49 681 95904 400, email: firstname.surname@consistec.de)

[1] In 2009 the *German Federal Office for Economic and Export Control* launched an online reservation system to apply cash bonus for new cars. Due to the huge interest there was a rapid increase in the inbound traffic and the protection system wrongly interpreted it as DoS attack and took the server down while it was actually capable of handling the requests.

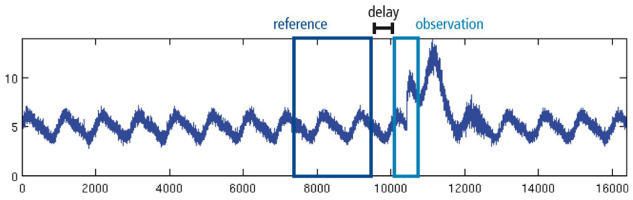

Fig. 1. Reference and observation windows in the signal

problems. Barford et al. [22] developed a tool based on wavelet transform to capture spikes at certain frequency and produce deviation score to detect anomalies.

To the best of our knowledge, none of the techniques above matches our technique in terms of features, i.e., is system independent and scalable, relatively noise resistant, online, unsupervised, adaptive, and multidimensional. Some of the techniques use only a static set of parameters (e.g., [9], [21], [22]), some are off-line techniques (e.g., [8], [12]), others (e.g., [11]–[13], [18]) are very specific to certain types of systems. [14], [21] are restricted to very elementary algorithms without any preprocessing (in our experiments, we were not able to reproduce some examples from [21]). Several techniques are potentially expensive in the deployment due to the collection of huge historical data and models [10], packet header keydata or system calls [13], [15], [16], [19], or due to rules generation [17]. [20] is comparable to our technique but seems to have a high false positive rate from bursty signals as the authors admitted.

III. Design and Implementation

The platform used in this work is Matlab version 7.8.0 (R2009a). We have used synthetic signals based on [22] and [2] for modelling daily traffic load and flash crowd events. These traffic signals are generated by overlaying a sinus signal with one third of the second harmonic and adding white Gaussian noise as can be seen in Fig. 1.

The x- and y-axis represent the time-line and the power of the measured parameter (for synthetic signals we omit units in the figures). We also imported datasets from a real network with distributed architecture [23].

A. Noise Filtering

Before analyzing the signals, noise needs to be maximally reduced while preserving information about the anomalies in the signal. For this task we use the wavelet transform instead of the conventional linear low-pass filter. According to Dolabdjian et al. [24] wavelet-based denoising is clearly superior to linear low-pass filtering in terms of effectiveness (higher SNR) and adaptivity. We have evaluated different wavelet variations and are favoring the *Spline Wavelet 2.6* which produced the best contrast between anomalies and normal traffic in our tests [23].

B. Multidimensional Analysis With Confidence Value

To have quantifiable results we compute confidence values, adopting an approach from learning algorithms by defining reference and observation segments of the signal. The observation data is compared with a profile of normal network behavior from a sliding reference window.

Both groups are compared by means of the Mahalanobis distance [6], that reflects how strong the observation segment deviates from the rest of the signal. The Mahalanobis distance is a very fast classification technique and can be applied multidimensionally without specifying explicitly how the dimensions correlate. It performs very good as long as noise is minimally present [6].

Let $\bar{x}_\mathrm{r} = (x_{\mathrm{r},1}, \ldots, x_{\mathrm{r},n})^T$ and $\bar{x}_\mathrm{o} = (x_{\mathrm{o},1}, \ldots, x_{\mathrm{o},n})^T$ be means of reference and observation groups respectively with n dimensions each, and S be the covariance matrix. The Mahalanobis distance is defined as

$$d\left(x_\mathrm{r}, x_\mathrm{o}\right) = \sqrt{\left(\bar{x}_\mathrm{r} - \bar{x}_\mathrm{o}\right)^T S^{-1} \left(\bar{x}_\mathrm{r} - \bar{x}_\mathrm{o}\right)} \,.$$

As we can observe, the Mahalanobis distance is not linear to the spatial distance of the objects. Instead, it also considers the correlation of the data set in each group (cf. [23, Section 3.3.1] for more details). Moreover, the computation is independent from the scale of measurements (i.e., scale invariant), thus the signals produced by the wavelet transform do not need to be normalized.

The confidence value is computed using two discrete sliding windows as reference and observation sets (cf. Fig. 1). Usually, both windows are slided synchronously to the arrival of new data in order to maximize response to anomalies. The reference window contains data from the past, while the observation window consists of the new incoming data. The ideal expectancy is that the reference window contains only data with normal behavior while the observation window contains only anomalous data.

C. Parameterization

We can tune several parameters when applying our technique to a particular system: 1) the set of signals to be monitored, 2) the level of wavelet transform, and 3) the lengths of the windows and the delay between them.

Our method assumes signal-based datasets in which anomalies produce abrupt changes. Often, these are time-stamped data like network load or CPU utilization. In the case of multidimensional analysis, the choice of the signals is essential. The Mahalanobis distance produces a maximum confidence value for signals with orthogonal behavior during anomalies (e.g., increasing request rate together with decreasing output bandwidth).

The Mahalanobis distance computation requires low noise levels. On the other hand, higher levels of wavelet transform, which guarantee low noise levels, reduce time resolution of the analysis. For signals with periodical usage pattern, the reference window should contain at least one complete period of the lowest level pattern (e.g., the daily pattern) so that the fluctuation is recorded as normal behavior. However, computation of the Mahalanobis distance takes significantly more time if more samples are contained in the transformed reference window; 128 to 256 samples turned out to be a good trade-off between accuracy and effort. The observation window, on the other hand, should be as small as possible to maximize response to anomaly; in our experiments, setting it to contain only a single sample turned out to be a good choice. The delay between reference and observation window must be chosen sufficiently long such that anomalies do not enter the reference window before they have generated an alarm. However, the delay should not be too long, because the bigger the delay, the later the reference windows learns about the current signal behavior.

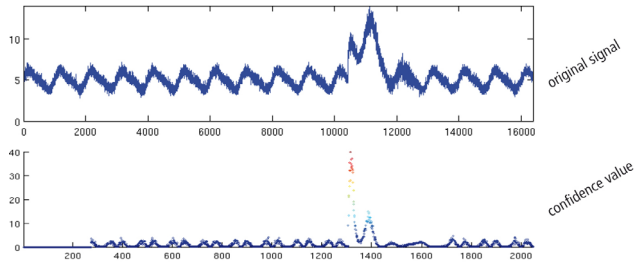

Fig. 2. The distance value of flash crowd anomaly

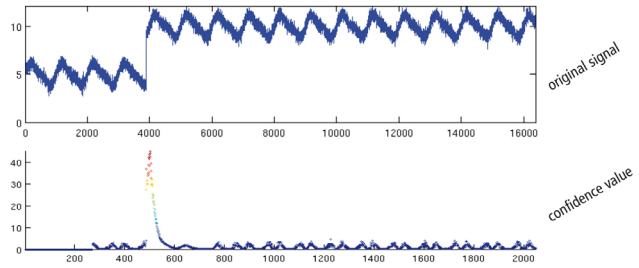

Fig. 3. The distance value of network operation anomaly

Let s be the sampling rate of the original signal and t the time interval contained in the reference window. The dependency between the level of wavelet transform l and the number w of samples in the transformed reference window can be approximated with $l \approx \log_2 \frac{t \cdot s}{w}$.

IV. Evaluation

A. Singularity Detection

Before demonstrating how the technique deals with multidimensional scenarios, we present its reaction to simple synthetic signals mimicking flash crowd events and network operation anomalies.

Flash crowd events, which are characterized by rapid increase in the traffic flow of a particular type followed by a gradual drop off over time, have been modelled with beta distribution and then convolved with the original segment of the signal before added onto it. This convolution has been used because flash crowd events are caused by human interaction and thus follow the usage pattern. Network operation anomalies, which are characterized by a single spike (up or down) followed by stable load but at a different level, have been modelled by shifting a portion of the signal.

The result of the technique can be seen in Fig. 2 and Fig. 3 respectively. The *contrasts* – i.e., the ratio between the average value of the signal in normal behavior and the anomalous segment – produced by the technique lie around $1:8$, which is much better than $1:2$ in the original signals.

B. Multidimensional Analysis

In the multidimensional scenario, we demonstrate how the technique reacts to two signals which show similar behavior in the *normal phase*. Regarding their behavior at the time the *anomaly* occurs we distinguish three cases: 1) both signals increase (or decrease) abruptly, 2) one signal increases and the other one remains unchanged following the normal course, and 3) one signal increases and the other one decreases. We see in Figs. 4, 5 and 6 that the

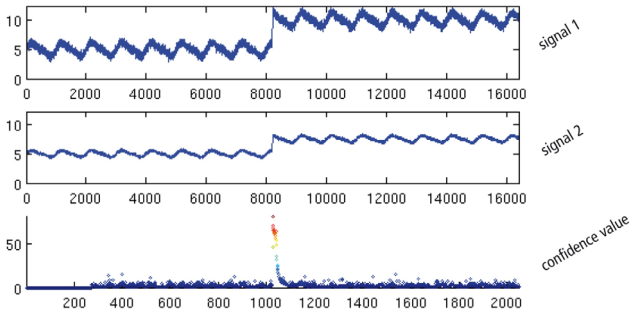

Fig. 4. Signals and distance metric of scenario 1

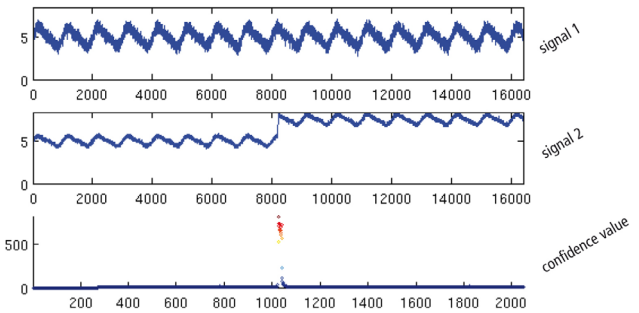

Fig. 5. Signals and distance metric of scenario 2

confidence value is considerably higher when both signals show different behavior.

The confidence values for a particular kind of anomaly can be maximized by selecting an appropriate set of measured signals, e.g., the inbound and outbound traffic of a system. In scenario 1, the system experiences sudden increase in the number of requests but it can still handle them without a hitch. Scenario 2 indicates a more suspicious situation: either there is a request buffering/discarding or the system has reached its saturation stage, i.e., runs at maximum throughput. In case of scenario 3 the system is likely to experience the *thrashing effect* [25], where the throughput decreases above a certain input load.

Scenarios 2 and 3 are also typical for DoS attacks. Because the main purpose of DoS attacks is to stress the attacked system, this kind of attacks usually cause increase at inbound traffic but regular or even lower throughput. This can be distinguished from flash crowd events where both signals increase as long as the server is capable of handling the requests. When simultaneously analyzing the signals from inbound and outbound rate of the system with our technique, we would see significantly higher confidence values for the DoS scenario than for flash crowd events.

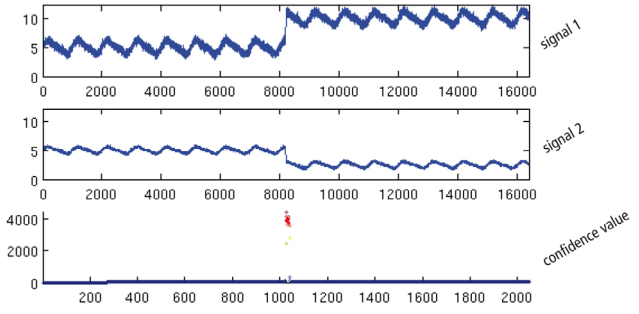

Fig. 6. Signals and distance metric of scenario 3

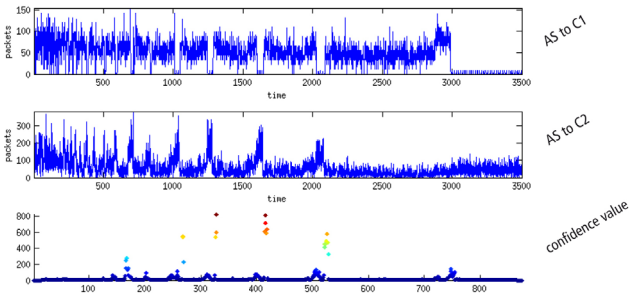

Fig. 7. Computation of the confidence value of the CWS signals

However, our method is not tailored to detecting malicious attacks.

C. Real World Example

In addition to the synthetic tests from before, we have applied our technique to analyze data from an overload situation of a real web service from one of our customers [23]. The system consists of three components: one main application server (AS) and two satellite components, an SMS gateway (GW) and a database server (DB). Requests trigger the system at high rate through AS which responds by sending answers via GW. For some incomplete requests AS has to query DB before answering the requests via GW.

We have monitored the number of packets from AS to DB and from AS to GW for our analysis. The input signals and the computed confidence value can be seen in Fig. 7. As desired, the confidence value stays relatively low most of the time although the input signals are very noisy. At around 1000, 1250, 1650 and 2100 time units of the input signals, we see very high confidence values. They indicates correlations between the two signals: the packet rate from AS to GW increases while the rate from AS to DB drops to nearly zero. This had not been noticed before by the system's administrators and has been considered *important* enough to require further investigation.

V. Conclusion

We have presented a new technique for detecting anomalies in networks using wavelet transform and Mahalanobis distance. Comparing current and reference data with the Mahalanobis distance allows for 1) comparison of arbitrary multidimensional signals, 2) computation of confidence values indicating *how* abnormal the current values are, 3) adapting automatically to evolution of the load pattern, and 4) dealing with periodic fluctuations of the normal load. Taking multiple input signals into account reduces the probability of false alarms and can be used to distinguish flash crowd events from DoS attacks. Besides monitoring systems for excessive load or attacks, the technique can also be used to detect hidden correlation in composite systems.

References

[1] P. Barford and D. Plonka, "Characteristics of network traffic flow anomalies," in *IMW '01: Proceedings of the 1st ACM SIGCOMM Workshop on Internet Measurement*. New York, NY, USA: ACM, 2001, pp. 69–73.

[2] I. Ari, B. Hong, E. L. Miller, S. A. Brandt, and D. D. E. Long, "Managing flash crowds on the internet," in *11th International Workshop on Modeling, Analysis, and Simulation of Computer and Telecommunication Systems (MASCOTS 2003)*. IEEE Computer Society, 2003, pp. 246–249.

[3] BAFA, "BAFA bedauert technische Schwierigkeiten ...." http://www.bafa.de/bafa/de/presse/pressemitteilungen/2009/10_umweltpraemie.html, 2009, [Online; accessed 13-Sept-2010].

[4] S. Mallat, *A wavelet tour of signal processing - Third Edition: The Sparse Way*. Academic Press, 2008.

[5] P. C. Mahalanobis, "On the generalised distance in statistics," in *Proceedings National Institute of Science, India*, vol. 2, no. 1, April 1936, pp. 49–55.

[6] M. Wölfel and H. Ekenel, "Feature weighted Mahalanobis distance: Improved robustness for Gaussian classifiers," in *13th European Signal Processing Conference (EUSIPCO)*, Antalya, Turkey, September 2005.

[7] A. Patcha and P. Jung-Min, "An overview of anomaly detection techniques: Existing solutions and latest technological trends," *Comput. Netw.*, vol. 51, no. 12, pp. 3448–3470, 2007.

[8] S. E. Smaha, "Haystack: an intrusion detection system," in *Proceedings of the Fourth Aerospace Computer Security Applications Conference*. Washington, DC, USA: IEEE Computer Society, 1988, pp. 37–44.

[9] D. Anderson, T. Frivold, and A. Valdes, "Next-generation intrusion detection expert system (NIDES)," SRI International, 333 Ravenswood Avenue Menlo Park, CA 94025-3493, Tech. Rep., May 2005.

[10] S. Staniford, J. A. Hoagland, and J. M. McAlerney, "Practical automated detection of stealthy portscans," *J. Comput. Secur.*, vol. 10, no. 1-2, pp. 105–136, 2002.

[11] N. Ye, S. M. Emran, Q. Chen, and S. Vilbert, "Multivariate statistical analysis of audit trails for host-based intrusion detection," *IEEE Trans. Comput.*, vol. 51, no. 7, pp. 810–820, 2002.

[12] S. Forrest, S. A. Hofmeyr, A. Somayaji, and T. A. Longstaff, "A sense of self for Unix processes," in *SP '96: Proceedings of the 1996 IEEE Symposium on Security and Privacy*. Washington, DC, USA: IEEE Computer Society, 1996, pp. 120–128.

[13] E. Eskin, S. J. Stolfo, and W. Lee, "Modeling system calls for intrusion detection with dynamic window sizes," in *Proceedings of DARPA Information Survivabilty Conference and Exposition II (DISCEX)*, vol. 1. Los Alamitos, CA, USA: IEEE Computer Society, 2001, p. 0165.

[14] A. Valdes and K. Skinner, "Adaptive, model-based monitoring for cyber attack detection," in *RAID '00: Proceedings of the Third International Workshop on Recent Advances in Intrusion Detection*. London, UK: Springer-Verlag, 2000, pp. 80–92.

[15] M. Shyu, S. Chen, K. Sarinnapakorn, and L. Chang, "A novel anomaly detection scheme based on principal component classifier," in *Proceedings of the IEEE Foundations and New Directions of Data Mining Workshop, in conjunction with the Third IEEE International Conference on Data Mining (ICDM'03)*, 2003, pp. 172–179.

[16] W. Lee and S. J. Stolfo, "Data mining approaches for intrusion detection," in *SSYM'98: Proceedings of the 7th USENIX Security Symposium*. Berkeley, CA, USA: USENIX Association, 1998, p. 6.

[17] J. Dickerson and J. Dickerson, "Fuzzy network profiling for intrusion detection," in *Proceedings of the 19th International Conference of the North American Fuzzy Information Processing Society (NAFIPS)*, 2000, pp. 301–306.

[18] M. Ramadas, S. Ostermann, and B. Tjaden, "Detecting anomalous network traffic with self-organizing maps," in *Recent Advances in Intrusion Detection*, ser. LNCS. Springer, 2003, pp. 36–54.

[19] L. Ertös, E. Eilertson, A. Lazarevic, P. Tan, V. Kumar, P. Dokas, and J. Srivastava, "Minds - Minnesota intrusion detection system," 2004.

[20] M. Thottan and C. Ji, "Anomaly detection in IP networks," *IEEE Trans. Signal Process.*, vol. 51, no. 8, pp. 2191–2204, 2003.

[21] P. Huang, A. Feldmann, and W. Willinger, "A non-intrusive, wavelet-based approach to detecting network performance problems," in *IMW'01: Proceedings of the ACM/SIGCOMM Internet Measurement Workshop*, November 2001.

[22] P. Barford, J. Kline, D. Plonka, and A. Ron, "A signal analysis of network traffic anomalies," in *IMW '02: Proceedings of the 2nd ACM SIGCOMM Workshop on Internet Measurement*. New York, NY, USA: ACM, 2002, pp. 71–82.

[23] K. Hendrawan, "Anomaly detection on network service with wavelet technique," Master's thesis, Saarland University, May 2010.

[24] C. Dolabdjian, J. Fadili, and E. H. Leyva, "Classical low-pass filter and real-time wavelet-based denoising technique implemented on a DSP: a comparison study," *The European Physical Journal*, vol. 20, pp. 135–140, 2002.

[25] H.-U. Heiss and R. Wagner, "Adaptive load control in transaction processing systems," in *VLDB '91: Proceedings of the 17th International Conference on Very Large Data Bases*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1991, pp. 47–54.