Challenging Entropy-based Anomaly Detection and Diagnosis in Cellular Networks

P. Fiadino, A. D'Alconzo, M. Schiavone, P. Casas
FTW Vienna
surname@ftw.at

ABSTRACT

In this paper we challenge the applicability and performance of entropy-based approaches for detecting and diagnosis network traffic anomalies, and claim that full statistics (i.e., empirical probability distributions) should be applied to improve the change-detection capabilities. We support our claim by detecting and diagnosing large-scale traffic anomalies in a real cellular network, caused by specific OTT (Over The Top) services and smartphone devices. Our results clearly suggest that anomaly detection and diagnosis based on entropy analysis is prone to errors and misses typical characteristics of traffic anomalies, particularly in the studied cellular networks' scenario.

1. MOTIVATION & METHODOLOGY

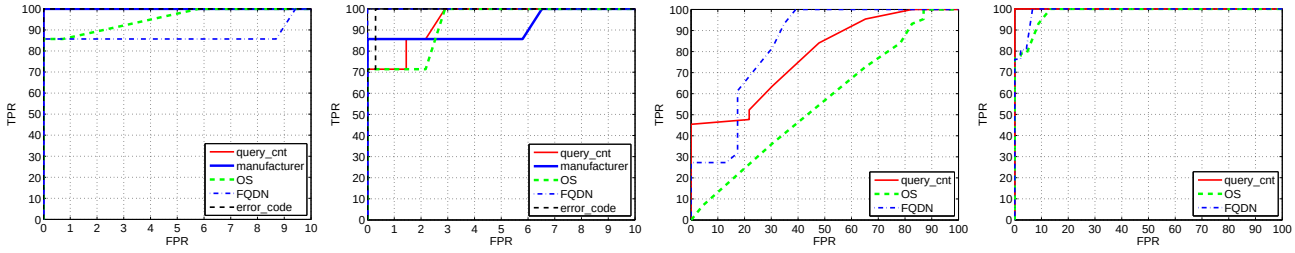
The complexity of current Internet scenario, with highly popular and omnipresent services accessed over multiple networks (cellular, WiFi, FTTH, etc.) and distributed through massive CDNs has revamped the relevance of automatically detecting and diagnosing traffic anomalies impacting large numbers of users. A particularly popular approach for detecting anomalies in network traffic is the one represented by entropy-based analysis [1]. In a nutshell, entropy-based anomaly detection consists of detecting abrupt changes in the time series of the empirical entropy of certain traffic descriptors or features, related to the specific anomaly. The entropy of a feature captures the dispersion of the corresponding probability distribution in a single number, becoming highly appealing for the analysis. However, such a compression necessarily looses relevant information about the distribution of the analyzed feature, masking in many cases the effects produced by an anomaly.

Although entropy-based approaches have been successfully applied for anomaly detection in the past [1–3], in fixed-line networks and using only transport-layer features such as IPs and ports, their application in the aforementioned context has severe limitations, given the characteristics of current traffic anomalies. In this paper we propose a generic framework to detect and diagnose large-scale network traffic anomalies, and use it to show that entropy-based approaches are not suitable for the analysis of traffic anomalies observed in an operational cellular network. To overcome this limitation, we apply an empirical distribution-based anomaly detection approach we have recently introduced in [4], which analyzes the complete distribution of the monitored features.

Given that the types of anomalies we target are related to large-scale services and application, the proposed framework is based on the analysis of DNS traffic. From our operational experience, application-specific anomalies are particularly visible in the DNS traffic. Indeed, abrupt changes in the DNS requests count can be considered as a symptom of such anomalies. From DNS transactions we derive two sets of time-series (signals from now on) denoted as *symptomatic signals* and *diagnostic signals*. All signals are checked for significant changes from their reference of "normality". However, the symptomatic signals are designed such that their changes directly relate to the presence of abnormal and potentially harmful events. On the other hand, changes in the diagnostic signals per-se do not have a negative connotation, but rather ease and guide the interpretation of the anomalous event. In the diagnosis of a detected anomaly, deviations of the symptomatic signals are correlated with the subset of simultaneously changing diagnostic signals to provide a comprehensive characterization of the event. The features we consider are the following: information about the mobile device (ID, manufacturer, OS), the network settings (APN, RAT, DNS server IP), the host name (FQDN) of the requested service, and the status of the DNS transaction (i.e., successful, time-out, retransmission).

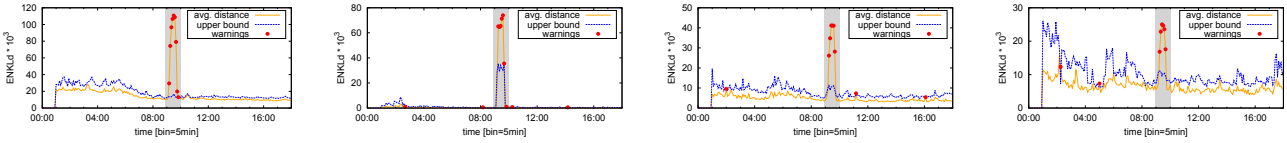
To evidence the limitations of entropy-based anomaly detection, we test and compare an entropy- and a distribution-based approach, applied to both symptomatic and diagnostic signals. The entropy of a random variable X is $H(X) = -\sum_{i=1}^{n} p(x_i)log(p(x_i))$, where $x_1, \ldots, x_n$ is the range of values for X, and $p(x_i)$ is the probability that X takes the value x_i. The entropy is normalized to a scaling factor $log(n_0)$, where n_0 is the number of distinct x_i values in a given time bin. Given the time-series of the entropy of a feature, we use any simple adaptive-threshold detector (e.g., Moving Average, EWMA) to spot abrupt changes. The distribution-based detection consists of computing the distance between the empirical distribution of a certain feature and a reference set of anomaly-free distributions previously observed. We rely on a symmetrized and normalized version of the well known Kullback-Leibler divergence to compute such a distance. The complete approach and the heuristics used for the construction of the reference set are described in [4].

The analysis is performed on top of two different example anomalies which we have observed in the cellular network of a national wide ISP in EU; we refer to them as E_1 and E_2 respectively. Both anomalies were manually analyzed and diagnosed, thus we have a solid ground truth describing both events. The anomaly E_1 is a short lived (i.e., hours) high intensity anomaly (e.g., about 10% of devices repeating a request every few seconds), where all the involved devices are produced by a single manufacturer and run the same



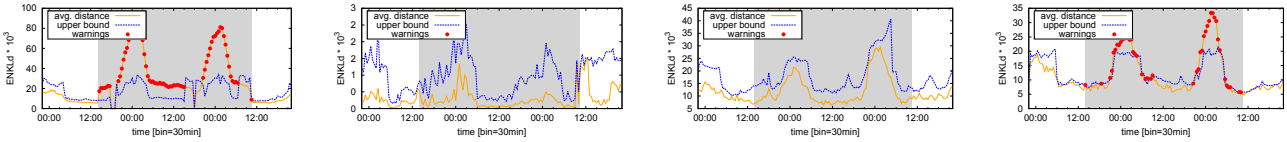
(a) E_1 - Entropy (b) E_1 - Distribution (c) E_2 - Entropy (d) E_2 - Distribution

Figure 1: ROC curves for the detection of abrupt changes in the corresponding symptomatic and diagnostic signals. Entropy-based detection performs properly with anomaly E_1, but completely fails with E_2.



(a) FQDN (b) Error Code (c) Manufacturer (d) OS

Figure 2: Output of the distribution-based detector for the diagnostic signals in E_1. All the signals exhibit distribution changes during the event.



(a) FQDN (b) Error Code (c) Manufacturer (d) OS

Figure 3: Output of the distribution-based analysis for the diagnostic signals in E_2. Distributions of FQDN and OS exhibit changes during the event, while manufacturer and Error Flag are unaffected.

OS. In this case, the number of involved terminals and the overall number of additional queries is such to overload the local DNS servers, increasing the number of time-out codes in the Error Flag field. The anomaly is related to the Apple Push Notification service on iPhone devices (FQDN = `*.push.apple.com`). The event E_2 corresponds to a long lasting (i.e., days) low-intensity anomaly (e.g., about 5% of devices repeating requests every few minutes). Differently from E_1, the involved terminals are produced by multiple manufacturers, even if they share the same OS. In this case, the anomaly is related to Android devices accessing the mtalk service (FQDN = `mtalk.google.com`).

2. WHEN ENTROPY FAILS

Fig. 1 depicts the ROC curves obtained in the detection of E_1 and E_2. The curves reflect the True Positive and False Positive Rates (TPR and FPR) obtained when changing the detection thresholds of both approaches. Each anomalous sample corresponds to a 5 minutes time bin, during the entire span of the anomaly (about 1hr for E_1, and about 2 days for E_2). The symptomatic signal is in all the cases the DNS query count per device, using either its entropy or the full distribution. Given the characteristics of E_1, there are four relevant diagnostic signals which show an abrupt change at the time of the anomaly: the manufacturer and OS (same type of devices are impacted), the FQDN (points to the requested, unavailable service) and the error code (the local DNS servers get overloaded and time-outs heavily increase). Figs. 1(a) and 1(b) show that both approaches are capable of detecting the abrupt changes induced by this anomaly, resulting in almost perfect detection for the impacted signals. Fig. 2 depicts the output of the distribution-based detector for the aforementioned diagnostic signals.

Fig. 1(c) shows how the entropy-based approach completely fails to capture the characteristics of the E_2 anomaly, as the FPR becomes too high to be applicable in the practice. In the case of E_2, only the OS and the FQDN diagnostic signals are impacted. Fig. 3 shows how both signals are detected as anomalous by the distribution-based approach during the almost the 2 days that the anomaly lasted. In addition, it shows how the other 2 signals (Error Code and Manufactures) correctly remain non-flagged by the approach. Finally, Fig. 1(d) shows a very similar performance to the one attained by the distribution-based approach in the previous scenario, reinforcing the evidence of its supremacy against entropy-based analysis.

To sum up, evaluations show that lower intensity anomaly E_2 is not correctly captured by the entropy-based detector, as the entropy results in this case into a too coarse metric, failing to reveal the effects of this type of anomalies. This limitation calls for the adoption of a distribution-based approach, which is perfectly suited for both tested scenarios.

3. REFERENCES

[1] A. Lakhina et al., "Mining Anomalies using Traffic Feature Distributions", in *ACM SIGCOMM*, 2005.

[2] G. Nychis et al., "An Empirical Evaluation of Entropy-based Traffic Anomaly Detection", in *ACM IMC*, 2008.

[3] Y. Kanda et al., "ADMIRE: Anomaly Detection Method using Entropy-based PCA with three-step Sketches", in *Computer Communications*, vol. 36(5), pp. 575-588, 2013.

[4] P. Fiadino et al., "On the Detection of Network Traffic Anomalies in Content Delivery Network Services", in *ITC*, 2014.