Diagnosing Device-Specific Anomalies in Cellular Networks

Mirko Schiavone, Peter Romirer-Maierhofer, Pierdomenico Fiadino, Pedro Casas
The Telecommunications Research Center Vienna - FTW
surname@ftw.at

ABSTRACT

The ever-increasing number of mobile devices is heavily modifying the traffic observed in cellular networks. From smartphones and tablets to Machine-to-Machine (M2M) devices, the traffic volumes and patterns generated by end-user and M2M applications introduce novel challenges to cellular network operators. One of these relates to the detection and diagnosis of network traffic anomalies caused by specific devices and applications. We introduce a simple yet effective approach to detect and diagnose such anomalies, applying entropy-based analysis on top of device/application-related descriptors. As case study, we present the analysis of a large scale traffic anomaly observed in a real cellular network, linked to smartphones. Our diagnosis approach promptly revealed a failure of a specific OTT (Over The Top) service not linked to the operator, showing its paramount advantage from an operational point of view.[1]

Categories and Subject Descriptors

C.2 [**Computer-Communication Networks**]: Miscellaneous; C.4 [**Performance of Systems**]: Measurement Techniques

Keywords

Anomaly Diagnosis; Entropy-based Analysis; Cellular Networks.

1. MOTIVATIONS AND METHODOLOGY

Cellular network operators have witnessed an amazing increase of heterogeneous mobile devices (smartphones, tablets, M2M devices such as telemeters, etc.) during the last decade. The applications supported by these devices introduce new traffic patterns which are potentially harmful for the network. For example, applications that provide continuous online presence (e.g., WhatsApp, Facebook, Skype) might generate a big burden on the signaling plane [1], impacting network performance. Specific devices and applications might also cause undesirable overloading events due to synchronized communication patterns, typical for M2M applications [4]. In this evolving scenario, detecting and rapidly diagnosing device-specific traffic misbehaviors becomes crucial for cellular network operators. This paper provides two main contributions: firstly, we present an approach for detecting and diagnosing anomalous traffic patterns linked to different device classes and applications; secondly, we analyze one of these anomalies in which an outage of a very popular OTT service resulted in smartphones flooding the network with connection attempts.

Our approach is articulated in two steps: **(i) Detection:** the *trigger* consists of detecting an abrupt change in the time series of specific traffic features revealing unexpected and potentially harmful behaviors. We call these time series *symptomatic signals*. 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. Therefore, we use the DNS requests count as the main symptomatic signal of the approach. The abrupt change detection is performed by a standard auto-adaptive algorithm, based on the mean and the variance of the DNS requests count.

(ii) Diagnosis: to find out the root causes of the detected anomalies, we define a set of features related to the class of problems we target, based on expert know how. In particular, we consider the following set of features, associated to each DNS request-response transaction: anonymized Mobile Device Identifier (ID), contacted DNS server IP, Radio Access Technology (RAT), Access Point Name (APN), Type Allocation Code (TAC), DNS requested Full Qualified Domain Name (FQDN), device manufacturer (obtained from the TAC code through public GSM Association databases), and error code of the DNS response (DNS rcode).

The first step of the diagnosis consists of identifying which of these features present a significant change in their probability distribution, simultaneously to the trigger. To this aim, we use the entropy as a means to condense the complete distribution of a feature into a single value for a given time bin. We refer to the time series of the entropy of these features as the *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. Entropy-based

[1] This work has been done in the framework of the EU-IP project mPlane, funded by the EC under grant 318627. The work is partially funded by the DARWIN4 COMET project.

CoNEXT Student Workshop'14, December 2 2014, Sydney, Australia.

http://dx.doi.org/10.1145/2680821.2680831.

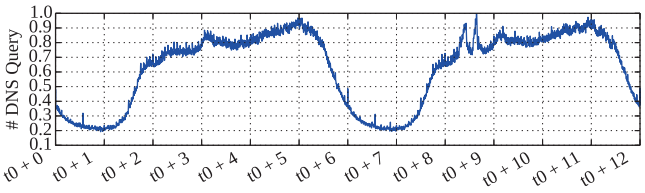

Figure 1: DNS requests count.

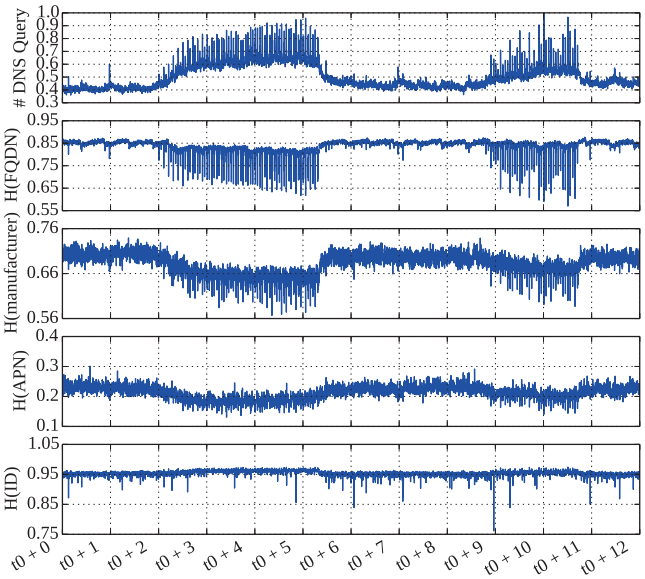

Figure 2: Entropy of selected features.

approaches have been proposed for traffic analysis in the past [2, 3], but in a fixed-line network context, and using only transport-layer features such as IPs and ports. Our analysis specifically targets cellular networks, using a much richer set of features, from network to device specific ones.

To detect changes in the diagnostic signals we use the same auto-adaptive algorithm of the detection step. In the practice, different anomalies cause significant changes only in a subset of diagnostic signals. This subset allows to build-up a signature for the detected anomaly. Finally, by further drilling down into the features that correlate the most to the trigger, the approach permits to narrow down the causes for a certain anomaly: e.g., anomalies linked to a specific device type, a specific service failure, and so on.

2. RESULTS

We present now a case study based on the detection and diagnosis of a large scale anomaly occurred in a real cellular network. Fig. 1 shows the time series of the total DNS requests count observed in the network for two consecutive days. Two significant and anomalous spikes are observed on the second day, which are easily spotted by the abrupt-change detection algorithm.

Fig. 2 provides a closer look into the anomaly, comparing the time series of the total DNS requests count and the entropy of 4 selected features: FQDN, manufacturer, APN, and ID. These features are extracted for each DNS request-response transaction (to preserve user privacy, any user related data are removed on-the-fly). The other diagnostic signals are omitted for brevity, as they show a behavior similar to the reported next. We notice that some of the observed diagnostic signals are correlated in a minor way to the anomaly. This is the case for ID, TAC, RAT, and DNS rcode, therefore we can exclude the cases in which the anomaly is caused by few users, a specific RAT, etc.. On the contrary, dimensions such as FQDN and manufacturer present a very high correlation with the spikes in the DNS count, suggesting that the issue might be due to specific devices (manufacturer) querying for certain services (FQDN). Features such as APN and server IP show partial correlation to the anomaly, thus need to be further cross-checked.

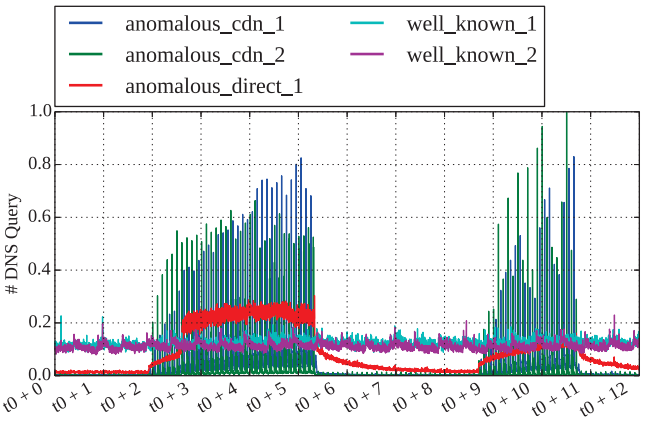

Figure 3: DNS requests per FQDN class.

The next step of the diagnosis is to drill down each of the dimensions that are highly correlated with the anomaly. This can be achieved, e.g., by comparing the heavy hitters before and during the anomaly. Fig. 3 reports the specific case for the FQDN. The plot shows the time-series of the most requested FQDNs during the anomaly. We observe that, while some of the top FQDNs associated to well-known services present a stable behavior (`well_known_1/2`), the FQDNs `anomalous_cdn_1/2` and `anomalous_direct_1` show a significant increase. The first two refer to content of a specific popular OTT service delivered via a major Content Delivery Network (CDN), whereas the third one points directly to the specific OTT service, showing that the problem is actually related to this service.

The mapping of the TAC codes to the manufacturer of the devices requesting the FQDNs related to the anomaly also reveal a specific smartphone type involved in the anomaly. In particular, the specific anomalous service runs on all these devices, but not on the other smartphone types. W.r.t. the dimensions presenting partial correlation, we found that all the different APNs are affected by the anomaly but in a different manner, suggesting that different APNs are configured for different customers. Indeed, different APNs are normally linked to different default DNS servers.

As a main conclusion, the proposed approach is helpful in highly reducing the time spent by the network operator in the diagnosis of unexpected traffic behaviors. In particular, this service outage resulted in an abrupt increase in the number of connection attempts from a large number of devices, and its fast diagnosis was paramount to understand the nature of such an anomaly.

3. REFERENCES

[1] A. Aucinas, N. Vallina-Rodriguez, Y. Grunenberger, V. Erramilli, K. Papagiannaki, J. Crowcroft, and D. Wetherall. Staying online while mobile: The hidden costs. In *Proceedings of the Ninth ACM Conference on Emerging Networking Experiments and Technologies*, CoNEXT '13, pages 315–320, New York, NY, USA, 2013. ACM.

[2] A. Lakhina, M. Crovella, and C. Diot. Mining anomalies using traffic feature distributions. In *Proceedings of the 2005 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications*, SIGCOMM '05, pages 217–228, New York, NY, USA, 2005. ACM.

[3] G. Nychis, V. Sekar, D. G. Andersen, H. Kim, and H. Zhang. An empirical evaluation of entropy-based traffic anomaly detection. In *Proceedings of the 8th ACM SIGCOMM Conference on Internet Measurement*, IMC '08, pages 151–156, New York, NY, USA, 2008. ACM.

[4] M. Z. Shafiq, L. Ji, A. X. Liu, J. Pang, and J. Wang. A first look at cellular machine-to-machine traffic: Large scale measurement and characterization. In *Proceedings of the 12th ACM SIGMETRICS/PERFORMANCE Joint International Conference on Measurement and Modeling of Computer Systems*, SIGMETRICS '12, pages 65–76, New York, NY, USA, 2012. ACM.