Raw Network Analyser (RNA): A DNS & Netflow Correlation System (Abstract)

Danny Lachos
dlachos@benocs.com
BENOCS GmbH
Berlin, Germany

Ingmar Poese
ipoese@benocs.com
BENOCS GmbH
Berlin, Germany

ABSTRACT

Network data correlations are widely used to obtain multi-dimensional traffic information [1–4]. Netflow streams, for instance, are correlated with BGP to generate enriched information about the ASes, IP addresses, and traffic volume. BENOCS is already doing correlations through our BENOCS Flow Analytics (See Figure 1). This tool uses a set inputs such as Netflow, BGP, SNMP, and other network protocols, and then performs correlations, data reduction, and data aggregation (amongst others), to finally generate an easy-to-read information of the network topology and transport traffic. However, this is not enough to know the applications and infrastructures behind those IP addresses or ASes. Currently, we have different applications (*e.g.*, content providers) using different infrastructures (*e.g.*, CDNs) to distribute their content. In this context, two key questions are missing to answer: What are those applications and infrastructures and how much traffic comes from them?. The DNS correlation is to annotate (and extend) the traffic information with the domain name they came from, so that two new dimensions can be obtained: infrastructure and application dimensions. With those new dimensions (See Figure 2), it is possible to know, for example, the amount of traffic which the *Content Provider 1* is providing over different infrastructures such as *akamaiedge.net*, *cloudfront.net*, etc.

The DNS correlation, however, is not an easy task and there are a set of research and technical challenges:

- The DNS records need to be tracked in order to know if they are still valid or not. For example, when a DNS record is received, it includes a TTL value that indicates how long the DNS record is valid. Based on this condition, not all the Netflow records will include DNS data because sometimes flow records are coming from the past or will come after the DNS record expired.
- The domain name resolution process includes a recursive lookup into the DNS records. For example, when a Netflow record is received, it will make a lookup into the DNS ARecords to obtain the infrastructure domain, then it needs to make a recursive lookup into the CNAME records. This recursive lookup process, in theory, may have infinity loops.
- Over the time, it is possible to have multiple domain names or URLs for a single IP address. In this case, it is necessary to identify mechanisms to reduce this ambiguity and at the same time increase the accuracy.
- After the DNS and netflow mapping, an infrastructure and application classification is necessary to reduce the large number of domains that are obtained in the form of fully qualified domain names (FQDNs).

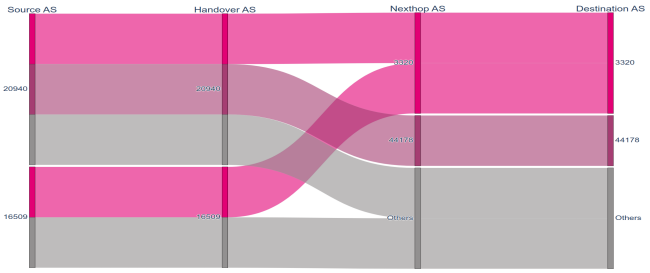

Figure 1: BENOCS Flow Analytics (simplified view).

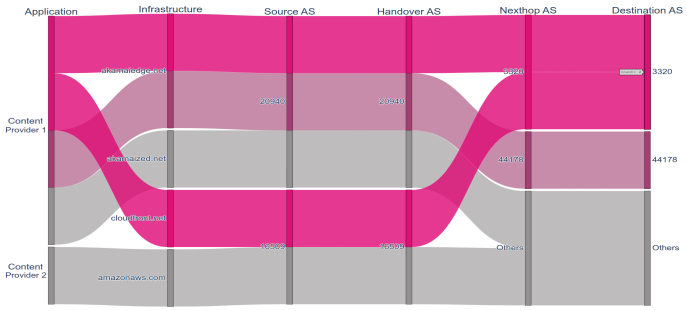

Figure 2: BENOCS Raw Network Analyser (RNA).

- Another challenges include the live processing with multiple DNS and Netflow records running in parallel, DNS warm-up periods of more than 12 hours both for DNS A records and DNS CNAME records, data might be corrupted and come in different formats, and others related to have a good trade-off between speed and memory used to reduce the time-to-frontend.

In order to try to overcome the previous challenges, BENOCS has developing a methodology that includes the analysis, design, and implementation of a real-time DNS-Netflow correlation system called Raw Network Analyser (RNA). RNA, running in a large European ISP, is taking advantage of the high performance and scalability of our flow analytic tools while extending its flow mapping functionality to be able to provide infrastructure and application information. This presentation will cover the main aspects of how the RNA is dealing with the DNS & Netflow correlation: from the motivation, going through the challenges, until its implementation and deployment to finally generate a new product that extends our current analytics and visualizations tools.

DNS and Internet Naming Research Directions virtual workshop, DINR, 2023
2023.

REFERENCES

[1] Martin Fejrskov Andersen. 2022. Detecting malware and cyber attacks using ISP data. (2022).

[2] Yoshiaki Harada, Koji Okamura, Takashi Chiyonobu, and Youngseok Lee. 2006. Analyzing correlation between flow data and as paths in BGP routing. In *Frontiers of High Performance Computing and Networking–ISPA 2006 Workshops: ISPA 2006 International Workshops, FHPCN, XHPC, S-GRACE, GridGIS, HPC-GTP, PDCE, ParDMCom, WOMP, ISDF, and UPWN, Sorrento, Italy, December 4-7, 2006. Proceedings 4*. Springer, 1126–1135.

[3] Atsushi Kobayashi, Yutaka Hirokawa, and Hiroshi Kurakami. 2010. Traffic Shift Monitoring System based on Correlation between BGP Message and Flow Data. *IEICE Technical Report; IEICE Tech. Rep.* 109, 438 (2010), 271–276.

[4] Enric Pujol, Ingmar Poese, Johannes Zerwas, Georgios Smaragdakis, and Anja Feldmann. 2019. Steering Hyper-Giants' Traffic at Scale. In *Proceedings of the 15th International Conference on Emerging Networking Experiments And Technologies (CoNEXT '19)*. Association for Computing Machinery, New York, NY, USA, 82–95. https://doi.org/10.1145/3359989.3365430