BGP Anomaly Detection
Using Wavelet Analysis

Lihua Yuan
Jianning Mai

I. MOTIVATION AND OBJECTIVE

Current works on BGP anomaly detection require extensive processing on collected routing data. Consequently, these techniques are limited to offline post-processing which has restricted applicability. In this work, we propose an BGP anomaly detection mechanism that requires only a simple count of BGP UPDATE messages collected over a certain period. We first investigate the self-similarity in BGP UPDATE traffic and present a preliminary validation. The strength of wavelet analysis in handling signals with scaling phenomenon and the earlier success in using it to detect network layer anomaly motivated us to apply the same technique to BGP routing traffic. Since this method does not process the content of routing messages, we consider it light-weight and believe it has the potential to be applied to anomaly detection on-the-fly. Additionally, our method can be used as an auxiliary tool to reduce the candidate data set for root cause analysis.

This is a continuation of our earlier work in last quarter’s EEC 273 class. In the previous work, we proposed to apply wavelet analysis techniques for BGP anomaly detection based on three observations. First, most known BGP anomalies has increased number of BGP UPDATE message as a byproduct. Second, BGP UPDATE traffic exhibits scaling phenomenon typically observed in self-similar and long-range dependent traffic. Last, wavelet analysis has its advantage in studying the scaling behavior, and earlier works applying it to anomaly detection in network layer traffic has been successful.

II. PLAN AND APPROACH

In our earlier work, we performed a preliminary validation on the self-similarity based on visual inspection of burstiness. In this project, we plan to vigorously study the self-similarity of BGP UPDATE traffic. We can find the Hurst parameter by study its variance plot, R/S plot, periodogram and Hurst parameter estimation graphs. Recent study indicates that Internet traffic
is non-stationary at multi-second scale. We plan to check if this is the case for BGP traffic. Meanwhile, the BGP replay tool provided by Sharad Agarwal might be useful to study the root cause of long-range dependence.

Our earlier work on wavelet analysis is based on simple Daubechies family wavelets and Matlab program. In order to achieve true on-the-fly anomaly detection, we need to find a more suitable wavelet family and write our own program in either C or Python. In addition, we plan to design a combinational “deviation score” mechanism for automated anomaly detection. Our results should be compared with available works to check its efficiency in terms of false positives and false negatives produced.

In addition to study the aggregated BGP UPDATE traffic, we plan to extend our study to segregated BGP traffic based on origin AS number. This could further reduce the candidate data set for root cause analysis and offer a clearer picture on anomalies. One associated question will be which AS do we look at? We certainly want to avoid monitoring every origin AS constantly since that will affect the scalability of our tool. Therefore, how to identify the potential problematic origin AS (or a list of candidates) remains a challenging task.