

Detecting Correlated Anomalous Flows using the Equilibrium Property

Fernando Silveira^{†*} Christophe Diot[†] Nina Taft^{*} Ramesh Govindan[‡]

[†]Thomson ^{*}University of Paris VI

^{*}Intel Research Berkeley [‡]University of Southern California

Thomson Technical Report
Number: CR-PRL-2008-09-0002
Date: September 1, 2008

Abstract: We have empirically observed that the average volume change across flows is close to zero in links carrying a large enough number of flows. This flow equilibrium property holds if the flows are nearly independent, and it is violated by traffic changes caused by several, potentially small, correlated flows. Many traffic anomalies fit this description, including scans, DDoS attacks, and routing shifts. Based on this observation, we exploit equilibrium to design a detection method for correlated anomalous flows. Our method identifies a qualitatively different set of anomalies compared with statistical techniques based on volume and entropy. It has two features that make it more practical than previously proposed methods: (1) it does *not* require a learning phase from time series data and (2) it provides an estimate of the volume of traffic involved in an anomaly.

I. INTRODUCTION

Anomaly detection methods look for *any* type of unusual traffic behavior. However, being too general about the definition of an anomaly can lead to false and missed detections. Moreover, it may be hard to identify the root cause of an anomaly or even the set of flows that triggered the detection.

In this paper, we propose a new method that detects a specific class of anomalies: those that are caused by large sets of correlated flows. In these anomalies, changes in a link’s traffic volume are caused by the collective contributions of many flows. Traffic anomalies like scans, DDoS attacks, and routing shifts, are examples of anomalies targeted by our approach. By design, our method ignores large volume changes caused by just one or a few large flows.

This method is based on the *flow equilibrium* property, which is the empirical observation that the average volume change across the flows in a link is close to zero. In links with a large number of flows but with no congestion, equilibrium arises from two simpler assumptions. First, since link capacity is limited, the traffic volume of a flow cannot always increase. Second, at time scales of up to a few minutes, flows are nearly independent from each other [2], [5]. Analyzing traffic traces, we measure the degree in which we can observe equilibrium under different time scales and flow aggregation levels. Our analysis confirms that equilibrium is found across a large parameter space.

Most of the previous anomaly detection techniques rely on the analysis of traffic volume [3], [15] or entropy of packet features [8]. Using different traces, we compare the anomalies detected by our method with those found by a statistical technique [15] that uses volume and entropy. We observe that anomaly detection with flow equilibrium has a number of attractive properties:

- Our experimental results show that equilibrium can find between four and six times more anomalies than a statistical technique that looks at volume and entropy combined. Moreover, equilibrium reveals a qualitatively different set of anomalies compared with these techniques. In order to catch these anomalies in volume and entropy, the corresponding detection threshold must be set at such a low value that the number of false alarms increases badly.
- Our method can estimate a confidence interval for the volume of anomalous traffic. We use this volume estimate to aid manual root cause analysis. Based on the relative ease with which we have been able to perform root cause analysis, we believe that our method can enable automatic classification, but have left this to future work.
- Detection can be done without any additional complexity compared to previous techniques. We rely on essentially the same traffic data as entropy based methods do. However, our method involves a simpler computation because each anomaly is detected with information from a single snapshot of traffic in time, as opposed to previous methods which require calibration on time series data.

The rest of the paper is organized as follows. In Section II,

we define the notion of flow equilibrium, explore its causes through simple models, and present empirical evidence that it is observed in real traffic data. Section III presents a simple method to detect if equilibrium is violated on a set of flows, and discusses some properties of this method. Specifically, we show that equilibrium is violated by a well-defined class of anomalies, namely volume changes caused by correlated flows. Section V explores the parameter space where equilibrium violation can be effectively used as an anomaly detector. In Section VI, we show on traffic traces that the anomalies found by equilibrium are substantially different from those detected by a statistical method based on traffic volume and entropy of packet features. For example, in traces from Internet2 and GEANT2, our method finds a systematic measurement anomaly caused by a specific router vendor implementation. We measure the complexity of our approach in Section VII. Finally, we discuss related work in Section VIII, and summarize our conclusions in Section IX.

II. FLOW EQUILIBRIUM

In this section, we introduce *flow equilibrium*, an empirically observed traffic property. We define equilibrium and discuss its possible causes. We argue that it arises when a large number of nearly independent flows traverse an uncongested link, which is the case in backbone networks. This observation forms the basis for our anomaly detection method.

A. Definition of Flow Equilibrium

A traffic *flow* is a set of packets that shares the same values for a given set of traffic features (e.g., source and destination IP addresses, source and destination ports, and protocol number). To study the evolution of a flow, time is usually divided into fixed sized intervals called *bins*. The *volume* of a flow f during bin i , denoted by $x_{f,i}$, is the number of packets or bytes in the flow during the corresponding bin. Denote by:

$$\delta_{f,i} = x_{f,i+1} - x_{f,i}, \quad (1)$$

the *volume change* of flow f in bin i .

Now, consider F flows traversing a link L . On this link, the average volume change between bins i and $i + 1$ is given by:

$$\hat{\delta}_i = \sum_{f=1}^F \frac{\delta_{f,i}}{F}. \quad (2)$$

For notational simplicity, we omit L from $\hat{\delta}_i$ as the link under consideration is evident from the context.

Flow equilibrium is the observation that the average volume change $\hat{\delta}_i$ on a link tends to zero when the number of flows F in a time bin is large enough. Stated formally, flow equilibrium holds if:

$$\lim_{F \rightarrow \infty} \hat{\delta}_i = 0, \quad \forall i. \quad (3)$$

Before we discuss possible causes of flow equilibrium, we illustrate this property in an OC48 link between a commercial ISP and the AMES Internet Exchange [12]. On this peering link, we select a specific 1-minute time bin, and measure

the volume change of each 5-tuple flow between this bin and the next. We gradually increase the set of flows that are used to compute the average volume change. We do this by first computing a random permutation of all the flows seen in the time bin, and then computing the average and 95% confidence interval for the first n flows, for increasing values of n . Figure 1 shows that as more flows are included in the computation, the closer this average gets to zero together with its confidence limits. Although we illustrated this property for a specific time bin and a specific trace, we have observed the same behavior on all other time bins in this trace, as well as in other traces.

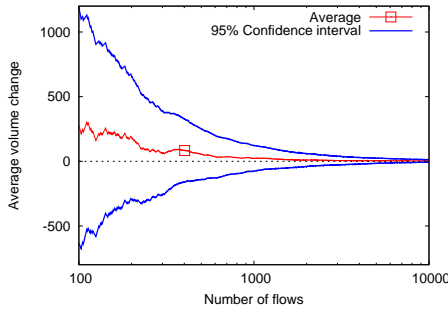


Fig. 1. An example from a traffic trace illustrating flows in equilibrium.

Flow equilibrium may not hold at all time scales or for all flow definitions. For example, diurnal variations in traffic usage would violate flow equilibrium over hour-long time bins. In Section V, we analyze traffic data and show that flow equilibrium holds across many flow definitions for time bins in the order of a few minutes. Furthermore, equilibrium may not hold at all locations in the network since it requires a large enough collection of flows. For example, it may not hold on access links from end hosts.

B. Causes of Flow Equilibrium

When equilibrium holds, we can explain it in two ways. First, consider the case of a congested, fully utilized link. If one flow manages to increase its throughput (for example, by malicious behavior), then the volume of other flows on the link must decrease to compensate. Likewise, if a flow lowers its volume (when the corresponding application is idle), then TCP will re-distribute the available capacity to other flows. Thus, on a fully-utilized link, flow equilibrium holds trivially.

Now consider the case of an over-provisioned link with a large number of flows (e.g., a backbone link). On such links, equilibrium holds if the flows traversing the link are independent and thus their volume changes cancel each other out, making the average close to zero. To understand this more precisely, consider the following argument.

Suppose a flow f is only active between time intervals 1 and N , i.e., $x_{f,i}$ is zero outside of this range. The *drift* of f is the rate of change across consecutive time intervals, which

is given by:

$$\begin{aligned} \Delta_f &= \sum_{i=0}^N \frac{\delta_{f,i}}{N} = \sum_{i=0}^N \frac{x_{f,i+1} - x_{f,i}}{N} \\ &= \frac{x_{f,N+1} - x_{f,0}}{N} = 0. \end{aligned} \quad (4)$$

Thus, for any finite flow, the total drift is zero.

Now consider the case when N goes to infinity, in order to model long-lived flows. Note that the difference between $x_{f,N+1}$ and $x_{f,0}$ is at most $x_{f,N+1}$, which is limited by the maximum volume that can be transferred in one time bin. If this limit is X , then $\Delta_f \leq X/N$. The value of X is proportional to the link capacity at time $N+1$. If we assume that the link capacity is constant or it increases slower than a linear function of N , then the drift Δ_f goes to zero when N tends to infinity.

If a flow f has zero drift and we randomly pick a time bin i while f is active, then $\delta_{f,i}$ is a zero mean random variable. Thus if we have F flows with zero drift at time i , the average volume change $\hat{\delta}_i$ also has a zero mean. Finally, if the volume changes $\delta_{f,i}$ are independent across flows, then the *law of large numbers* [7] leads to equilibrium as it is stated in Equation (3).

Independence across flows is a common assumption in models of backbone traffic [2]. Even when independence does not hold strictly, small correlations between flows become insignificant compared with the randomness in large traffic aggregates. In fact, the law of large numbers holds even if the flows are weakly dependent [7]. In this situation, the convergence of $\hat{\delta}_i$ to zero may be slower, requiring a larger aggregate than in the perfect i.i.d. case.

It is important to understand that flow equilibrium does not contradict the observation that network traffic is self-similar [9]. One way in which self-similarity manifests is in slowly decaying correlations in measurements of *total* traffic volume across *time*. On the other hand, we have explained flow equilibrium by arguing that, in an uncongested link, there is little correlation across *flows* within a single time bin. Our explanation does not say anything about the temporal correlation of flows. Indeed, self-similar traffic can be explained by the superposition of a large number of independent flows [16], albeit with a specific characteristic (heavy tailed on-off times).

In this paper, we use flow equilibrium as a model for normal traffic behavior. On uncongested links, if equilibrium is violated, then all of its possible underlying causes must have been violated too, including the assumption that flows are independent or weakly correlated. This happens if several flows correlate by increasing (or decreasing) their individual volumes during a single time bin. We exploit this property in the rest of the paper, since many real-world attacks (e.g., DDoS, port scans, worms) exhibit such correlation.

III. EQUILIBRIUM-BASED ANOMALY DETECTION

In this section, we show how equilibrium can be used to detect correlated anomalous flows. We discuss two complementary problems that give insights into our method and illustrate its power: (1) given the volume of anomalous traffic,

what properties it needs to have in order to be detected by our method; and (2) given an anomaly that has been detected, how to estimate the volume of traffic involved in it.

A. Checking for Equilibrium in Traffic Data

Equilibrium arises from the law of large numbers, but since that is only an asymptotic result, the value of $\hat{\delta}_i$ is unlikely to be exactly zero on real data. A practical method to check for equilibrium has to specify how close $\hat{\delta}_i$ must be to zero so that a set of flows is considered in equilibrium.

If equilibrium is a consequence of independence among flows, there exists a simple test to check if a set of flows is in equilibrium, namely that, with high confidence, the volume changes of these flows are drawn from a zero mean distribution. For this, we simply compute the confidence interval for the average volume changes across flows, and check if that confidence interval includes zero. If this condition does not hold for a given time bin, we mark that time bin as anomalous. The rest of this section formalizes this intuition.

Consider a sample of F flows with volume changes given by $\delta_{f,i}$. Let $\hat{\sigma}_i$ be the sample standard deviation of the flow volume changes:

$$\hat{\sigma}_i = \left[\sum_{f=1}^F \frac{(\delta_{f,i} - \hat{\delta}_i)^2}{F-1} \right]^{\frac{1}{2}}. \quad (5)$$

If the volume changes are independent across flows, then for large F , $\hat{\delta}_i$ has a p -confidence interval given by the central limit theorem:

$$I_{\hat{\delta}_i} = [\hat{\delta}_i - K(p)\hat{\sigma}_i/\sqrt{F}, \hat{\delta}_i + K(p)\hat{\sigma}_i/\sqrt{F}]. \quad (6)$$

where $K(p)$ is the percentile $1 - (1 - p)/2$ of the standard normal distribution. We say that a set of flows is in equilibrium if $I_{\hat{\delta}_i}$ contains zero. Otherwise, we say that equilibrium is violated.

Clearly, the efficacy of the algorithm depends upon the choice of $K(p)$. As we increase the confidence level p , we also increase the size of the confidence interval $I_{\hat{\delta}_i}$. For a given set of flows, it is clear from (6) that the size of the interval is characterized by the value of $K(p)$. The smallest value of $K(p)$ such that the interval contains zero is:

$$K' = \left\lceil \frac{\hat{\delta}_i}{\hat{\sigma}_i} \sqrt{F} \right\rceil. \quad (7)$$

We call K' the *equilibrium assessment value* of a time interval. Note that equilibrium is violated if and only if K' is larger than $K(p)$.

When equilibrium is violated there are two possibilities. First, the confidence interval $I_{\hat{\delta}_i}$ is supposed to contain zero only for a fraction p of the time bins. Thus, in a fraction $1 - p$ of the times, we should expect equilibrium to be violated by normal traffic. This is the *false positive rate* of our anomaly detection method and it can be minimized by increasing the confidence level p , or the detection threshold $K(p)$.

The second possibility is that some set of flows violates the independence assumption. For instance, if many flows increase

(or decrease) their volumes at the same time, then these flows are no longer independent of each other. Such flows that cause equilibrium violations are termed *equilibrium anomalies*. On the other hand, a high-volume flow does not violate the independence assumption and cannot be detected using our method. More generally, a small number of correlated flows may not constitute an equilibrium anomaly. We explore this issue in greater detail in Section III-B.

These results are valid when the number of flows F is large, namely to get a good estimate of the equilibrium assessment value. According to a rule of thumb in statistics [1], we should consider time bins with at least 30 flows so that the central limit theorem provides a good approximation of the confidence interval $I_{\hat{\delta}_i}$. We make sure in all the traces at least 95% of the bins contain a minimum of 100 flows. Usually, the number of flows is much higher than that, i.e., in the order of thousands or tens of thousands.

Our method to verify whether equilibrium holds for a given pair of successive time bins is summarized in four steps:

- 1) Given a target false positive rate $1 - p$, determine the detection threshold $K(p)$ as discussed above.
- 2) For each flow f , measure its volume change between the two time bins, $\delta_{f,i}$.
- 3) Compute the equilibrium assessment value defined in Equation (7).
- 4) If the assessment value is larger than the detection threshold, then equilibrium has been violated.

B. Characterizing Equilibrium Anomalies

To better understand the types of flows that constitute equilibrium anomalies, consider N normal flows, whose volume changes have mean $\hat{\delta}_N$ and standard deviation $\hat{\sigma}_N$. In addition to the normal flows, consider a volume of anomalous traffic V that spreads across A flows. Thus $\delta_A = V/A$ is the volume change by each anomalous flow. We consider that there is no intersection between the normal and the anomalous flows, and we denote the total number of flows by $F = N + A$. It is easy to see that if we mix normal and anomalous flows, the mean volume change is:

$$\hat{\delta} = \frac{N\hat{\delta}_N + A\delta_A}{F}, \quad (8)$$

and the sample standard deviation is given by:

$$\hat{\sigma} = \left[\frac{N-1}{F-1} \hat{\sigma}_N^2 + \frac{NA}{F} (\hat{\delta}_N - \delta_A)^2 \right]^{\frac{1}{2}}. \quad (9)$$

Together, Equations (8) and (9) can be used to compute the equilibrium assessment value (Equation (7)) of the mixed set of flows.

We use the trace from Figure 1 to illustrate what anomalies can be detected by our method. Consider time intervals of one minute, flows defined as 5-tuples, and volume changes measured in packets. In a particular time bin, we observe around 550,000 5-tuples whose volume changes have average 0.04 packets and standard deviation of 47 packets. Thus the

equilibrium assessment value K' for these flows is 0.63. We call this the *background traffic*.

Figure 2 shows how the assessment value increases (i.e., equilibrium is violated) when we add different amounts of anomalous traffic to the background traffic. Each curve in the plot corresponds to a specific number of packets in the anomaly. On the x axis we vary the number of flows in the anomaly. Suppose we apply a threshold of 3 on the assessment value to trigger an equilibrium anomaly; this corresponds to a target false positive rate of 0.2%. Even if the anomaly contains half a million packets, the assessment value only exceeds the threshold (thus triggering a detection) if the packets are spread across 10 or more anomalous flows.

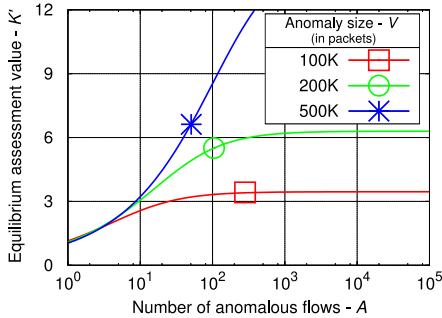


Fig. 2. A volume change violates equilibrium if it spreads across many flows.

We compute the minimum number of correlated flows needed to trigger an anomaly of a given size. Figure 3 shows this metric for different values of the detection threshold. The plot shows that given a threshold value, equilibrium cannot be violated by less than a certain amount of flows. The intuitive explanation is that if the anomalous flows are too few, then the correlation between them is too small to be detected (Section II-B). While this is just a qualitative description based on one example trace, we validate this intuition with more traffic data on a later section.

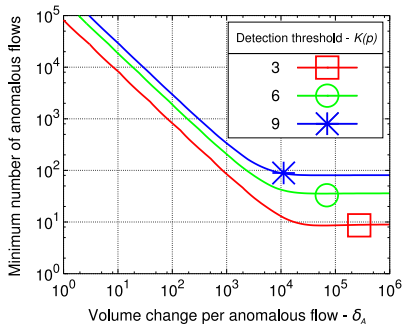


Fig. 3. Minimum number of flows in the anomaly to trigger a detection.

C. Estimating the Volume of an Equilibrium Anomaly

Our detection technique has an interesting property. The confidence interval in Equation (6) is an estimate of the

average volume change across all flows in the link. When an anomaly is detected, we can multiply the limits of this interval by the number of flows in the link and estimate the total volume of traffic involved in an equilibrium anomaly. This estimate is given by:

$$I = [\hat{\delta}_i - K(p)\hat{\sigma}_i\sqrt{F}, \hat{\delta}_i + K(p)\hat{\sigma}_i\sqrt{F}]. \quad (10)$$

This measure can be useful during root cause analysis. If an operator needs to track down the flows responsible for the anomaly (e.g., in the case of an attack) the list of alternatives can be narrowed down to those sets whose total traffic is within the lower and upper bounds of the confidence interval. This estimate can also be used as an input to automated procedures for root cause analysis, which we have left for future work. However, in Section VI, we use this estimate to help in manually identifying anomalous flows.

IV. TRAFFIC DATA

We study flow equilibrium and anomaly detection in three flow traces, summarized in Table I. These traces include research, academic, and corporate traffic. The first two traces are publicly available for research purposes from their sources.

TABLE I
SUMMARY OF TRACES USED IN THIS PAPER.

Trace	Collection period	Average rate	Duration
Internet2	Aug 2007	289 Mbps	31 days
GEANT2	Nov 2007	895 Mbps	30 days
Corporate	Sep-Dec 2007	2 Mbps	122 days

Our traces come from three network links: (1) a link connecting several customers of Internet2 to a backbone router in New York¹; (2) a peering link between a transit provider and the Frankfurt router at GEANT2²; (3) an access link between a corporate site and the rest of a worldwide MPLS enterprise network. Both Internet2 and GEANT2 use Juniper routers, and generate sampled J-Flow statistics at rates of 1/100 and 1/1000, respectively (the average rates in Table I are multiplied by the inverse of the sampling rates). The corporate network collects non-sampled Cisco NetFlow.

When binning the flow records, we assume that packets in a record arrive uniformly spaced in time. This assumption is necessary because flow traces keep timestamps only for the first and last packet in each record, so we cannot know the exact volume of the flow's traffic that should be attributed to each bin. Previous research has shown that this is a reasonable approximation for large time bins [14]. In our traces, this assumption affects less than 10 percent of the flows at bins of one minute, and less than one percent for bins of five minutes and larger. Thus we restrict the analysis of these traces to time bins larger than one minute.

Finally, in the Internet2 trace, the last 11 bits of all IP addresses are set to zero for anonymization. IP addresses

¹<http://www.internet2.edu/>

²<http://www.geant2.net/>

with a common prefix longer than /21 are aggregated into a single address. This gives a different meaning to the flow aggregation levels in the Internet2 trace. We take into account this difference during our analysis in the following sections.

V. PARAMETERIZING FLOW EQUILIBRIUM

Our goal in this section is to identify bin widths and flow definitions for which our method can detect traffic anomalies most effectively. We take an empirical approach to parametrize our model. We examine the traces described in Section IV, analyze time bins between one minute and one hour, and consider six flow aggregation levels: (1) *5-tuples*, (2) *host pairs*, (3) *source IPs*, (4) *destination IPs*, (5) *source ports*, (6) *destination ports*.

Ideally, parameter values should be chosen to minimize the numbers of false and missed detections, but this inevitably requires a “ground truth” set of anomalies, i.e., the precise list of events that must be detected. However, complete ground truth for anomaly detection is impossible to obtain and the common alternatives, such as event logs [3] or synthetically generated anomalies [8], [15] provide only *partial* ground truth. In our case, using the output of another anomaly detection method as a *referential* ground truth [17] is equally flawed, since our method targets a specific type of anomalies (as seen in Section III-B) which might be different those detected by the reference technique.

Therefore, instead of relying on imperfect ground truth, our approach consists in studying the sensitivity of our method to its parameters. Given a combination of trace and parameter values, we compute the fraction of time bins in the trace that are considered anomalous by our method. We choose a target false positive rate of 1%, which corresponds to a detection threshold of 2.57 in our method. Thus, even if our method cannot find any real traffic anomalies for a given parameter setting, it should still trigger detections for around 1% of the bins. Conversely, if it marks significantly more than 1% of the bins as anomalous, we consider that our method does not work well for that choice of bin width and flow aggregation.

Figure 4 shows the fraction of anomalous bins in all traces for each combination of time interval and flow aggregation level, when volume changes are measured in packets. We omit results for bytes as they are numerically similar and qualitatively the same.

A. Binning Time

The bin length is an important parameter in our algorithm. If it is too small, some anomalies may be missed. Intuitively, the smaller the time bin the more difficult it is for a large set of flows to synchronize and become correlated. If the anomalous traffic spreads across several small time bins, each bin by itself may not contain enough correlated flows to violate equilibrium. For instance, in GEANT2 with one minute time bins, the fraction of anomalous bins is around 2.5% across all flow aggregation levels, which is not much higher than the target false positive rate of 1%. Thus, if there are more true

anomalies in this trace, we might be missing them at such a small time scale.

On the other hand, at large time scales, the daily traffic patterns look like anomalies to the model. For example, if the time bins are one hour long, then at each hour in the morning, most flows are likely to start sending more traffic. This deterministic behavior makes flows strongly correlated with each other, and thus our method triggers alarms. These detections are *not* false positives for our method, since the volume changes of flows are indeed correlated. However, the root cause of these correlated flows (i.e., the daily usage patterns) is irrelevant to anomaly detection. The impact of the daily patterns for one hour time bins is most visible in Internet2 and GEANT2, where the contrast between peak and low usage times is greater due to the larger number of users in these networks than in the corporate one.

We observe that the sweet spot of the curves in Figure 4 is between 5 and 10 minutes, and the largest increases happen beyond 10 minutes. For this reason, we choose a bin length of five minutes for detecting anomalies in these traces.

B. Flow Aggregation

The flow aggregation level influences the types of anomalies that can be detected by our method. Namely, a given anomaly may violate equilibrium in one aggregation level and not in another. For instance, a port scan spreads traffic across several destination ports, but it targets a single destination host. The former type of flows are correlated and thus violate equilibrium, while the latter one does not.

Among the six types of flows we consider, 5-tuples are the ones with the smallest aggregation level. Clearly, any anomaly made of A flows in another flow aggregation level corresponds to at least A 5-tuples. Thus we expect to find more anomalies at the 5-tuple level than in other flow types. This is confirmed by the plots in Figure 4.

While it seems intuitive that we can detect any correlated anomalous flows at the 5-tuple level, this is not true in practice. Among the anomalies found by any of the more aggregated flow types (i.e., source IPs, destination IPs, source ports, and destination ports) we compute the fraction of those which is also detected on 5-tuples. This value is high in the Internet2 and GEANT2 traces (respectively 94% and 91%) but relatively lower in the corporate trace (74%).

We observe in Figure 4 that, in the Internet2 trace, the destination port flows detect less anomalies than all other flow aggregations. We found that this happens because 13% of the packets in this link are destined to TCP port 80 (HTTP traffic). The presence of this dominant flow reduces the number of detections in one specific flow definition, i.e., destination ports, but not in the others. To verify this, we removed from the trace all port 80 packets and detected the anomalies on the remaining traffic. The number of equilibrium anomalies on the destination port flows becomes very close to that on other flow definitions.

Instead of using one or just a few flow definitions, we choose a conservative approach and, in the next section, we

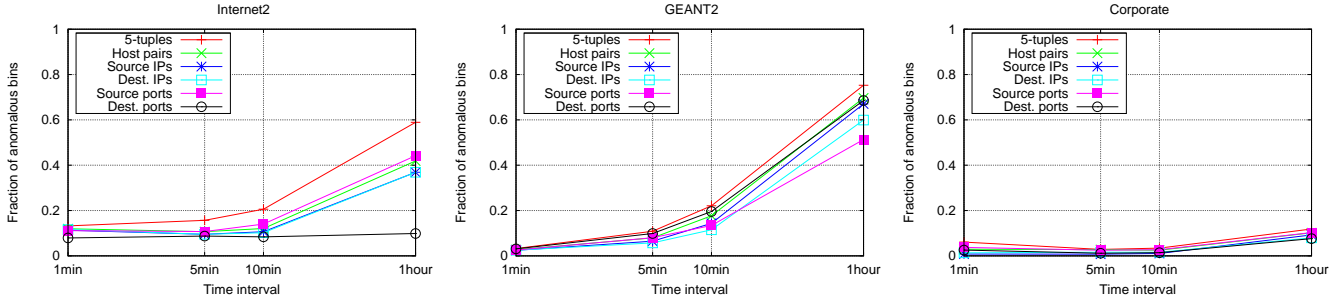


Fig. 4. Fraction of anomalous bins in each trace for different time intervals and flow aggregation levels.

evaluate our method on *all* the flow aggregation levels. In Section VI-C, we exploit the differences among the detections of all aggregation levels to help in identifying the anomalous flows, in the middle of all traffic.

VI. METHOD EVALUATION

In this section, we analyze the anomalies detected by flow equilibrium and compare them with those detected on volume and entropy, two traffic measures for finding general purpose traffic anomalies. We use the Kalman detection method [15] as a reference technique. For each anomaly found by either method, we manually identify its root cause by carefully inspecting the traffic flows in the time bin where it has been detected. Manual identification allows us to explore differences between equilibrium and the reference method.

A. Anomaly Detection with Volume and Entropy

Anomaly detection techniques often look for abnormal patterns in either traffic volume [4], [3], [15] or the entropy of packet features like IP addresses and ports [8]. They rely on statistical signal processing methods (e.g., frequency analysis, linear models, PCA) to extract outliers from time series data. Feature entropies have been shown to enable more accurate detections than traffic volume only.

Several such techniques have been proposed for extracting anomalies out of these metrics. We use the Kalman filter [15] as our reference technique, since it has been shown to detect a large variety of anomalies and it is relatively simpler to calibrate than other techniques [11]. Kalman was originally proposed as a network-wide anomaly detection technique. To provide a fair comparison with equilibrium, we run Kalman on the time series of volume and the four entropies (i.e., source and destination IPs and ports) of a single link. We have validated this single-link Kalman by comparing the anomalies it finds with the network-wide version of Kalman [13]. We found that single-link Kalman catches more than 90% of the anomalies found by the network-wide version.

Another reason for using Kalman as a reference technique is that its detection threshold can be set using the same rationale that we use in our equilibrium-based approach (Section III). Given a target false positive rate $1 - p$, we set the Kalman threshold to the percentile $1 - (1 - p)/2$ of the standard normal

distribution. This allows us to target the same amount of false positives on both methods.

B. Detection Settings for both Methods

We consider each trace is binned at five minute intervals. We extract all the anomalies found by equilibrium on any of the six flow aggregation levels: 5-tuple flows, host pairs, source IPs, destination IPs, source ports, and destination ports. We also detect anomalies using the single-link version of Kalman on the traffic volume and the four feature entropies: source IPs, destination IPs, source ports, and destination ports.

Our main goal is to compare the true detections on each method and confirm our intuition (Section III-B) that: (1) equilibrium finds anomalies that spread traffic across several correlated flows, and (2) it ignores large volume changes if they happen in a few flows. To do this, we set an extremely low false positive rate on both methods: approximately 2×10^{-9} . This corresponds to a threshold value of 6 for both methods.

C. Manual Root Cause Analysis

We have inspected the anomalies found by each method for all traces to understand their root causes. For the anomalies found by equilibrium, we designed a strategy that worked very well in practice. It is motivated by the observation (Section V-B) that anomalies often violate equilibrium in some flow aggregation levels but not in others. For each anomaly with this characteristic, we perform two steps:

- 1) For the flow aggregation levels where equilibrium is violated, we estimate the confidence intervals for the volume of traffic involved in the anomaly, as discussed in Section III-C.
- 2) For the flow aggregations where equilibrium is *not* violated, a possible explanation is that the anomalous traffic is concentrated in one or a few large flows (see Section III-B). Thus in these aggregation levels, we look among the largest flows, for a small set whose volume change fits at least one of the confidence intervals estimated in the previous step.

This procedure allows us to tag IP addresses and ports involved in the anomalies. If equilibrium is violated on all the flow aggregation levels at the same time, then the above heuristic cannot be applied. In this case, we classify the

anomaly using other sources of information. Specifically, for the anomalies found only by Kalman on volume and entropy, we relied on another observation. If an anomaly is not found by equilibrium on *any* flow aggregation level, then a plausible explanation is that it is concentrated on a few large flows on all flow aggregations. Thus equilibrium also helps in finding the root causes of Kalman anomalies.

Using the strategies described above, we have been able to identify the root causes of all, except five anomalies among all the detections in our traces. We specifically marked these five anomalies as having unknown root causes. These unknown detections can be either due to false positives or to real anomalies that could not be identified by visual inspection.

We validate our manual anomalous flow identification as follows. We remove from the anomalous time bin all the packets that we have manually identified as causing the anomaly and we check whether equilibrium is still violated on the remaining traffic. We found a single anomaly that could not be validated by this procedure. This happened in the Internet2 trace because at the same time of a port scan, there was also a measurement gap. After removing the first anomaly, we still detect the second one.

D. Comparing Equilibrium with Volume and Entropy

Having identified the flows responsible for a large set of anomalies, we can characterize the detection bias of each method. In Figure 5, for each detection where we identified the anomalous traffic, we plot the number of flows involved in the anomaly against the total number of packets in these flows. As expected, the anomalies detected by equilibrium always involve a large number of correlated flows. Interestingly, Kalman tends to detect events on few flows and large volume. Note also that the overlap between the two methods is small, and only in GEANT2 there are points found by both methods.

To better understand what type of anomalies equilibrium detects, we analyzed each individual anomaly and labeled it with a meaningful description from an operator’s perspective. For the anomalies where we identified the responsible flows, we investigate these flows looking for features like average packet sizes, application-level protocols, number of sources and destinations. For the anomalies for which we could not identify a set of anomalous flows, we looked at the whole traffic and used other sources of data (e.g., IP prefixes) to infer the type of anomaly. Table II summarizes this labeling in each trace. We now describe in more detail the criteria used to label all anomalies.

We label anomalies as *Denial-of-Services (DoS)* attacks when one or more sources send many small packets to a single destination. If the packets are large (above 1024 bytes) and use TCP, then we label the anomaly as a *file transfer*. When one or more sources send small packets to several destination ports of a single target host, we label it as a *port scan*.

In the corporate trace, we found anomalies related to a series of *enterprise applications*. For instance, we observed that once a day, hundreds of hosts simultaneously send broadcast packets for name service. Another similar case is an instant

TABLE II
ANOMALIES FOUND ON EACH TRACE

Internet2			
Anomaly type	Total	Equilibrium	Kalman
DoS	38	1	37
File transfer	2	2	0
Port scan	198	198	0
Routing change	1	1	0
Link outage	15	12	12
Measurement gap	136	135	4
Unknown	5	2	3
Total	395	351	56

GEANT2			
Anomaly type	Total	Equilibrium	Kalman
DoS	16	0	16
File transfer	4	4	0
Port scan	8	8	4
Routing change	37	36	3
Measurement gap	51	51	1
Total	116	99	24

Corporate			
Anomaly type	Total	Equilibrium	Kalman
File transfer	13	0	13
Enterprise applications	38	37	1
Link outage	12	11	2
Upstream link outage	13	13	0
Total	76	61	16

messaging (IM) application used between employees. During the anomaly, we found a number of clients continuously trying to establish TCP connections to a server which was probably not reachable. We identified anomalies due to *routing changes* by aggregating the IPs per source or destination prefixes³. Namely, we check if the total volume change across the IP addresses in a given prefix falls inside one of the confidence intervals estimated by our detection method. We classify an anomaly as a *link outage* if the time bin where it is detected contains no packets. We identify *measurements gaps* if there is a period of at least 10 seconds between two flow arrivals inside the time bin. The anomalies for which we could not discover their root causes were labeled as *Unknown*.

In Table II there are many of these anomalies in Internet2 and GEANT2. We have investigated these gaps in detail [6], and we discovered that they are caused by a bug in one of the implementations of J-Flow (Juniper’s equivalent of NetFlow), used in both Internet2 and GEANT2⁴. Note that the corporate trace is collected with Cisco NetFlow and does not have gaps.

Note that equilibrium detects considerably more anomalies than Kalman. The only way to catch those anomalies with the Kalman filter is to lower the detection threshold until the bins containing those anomalies get triggered. However, when doing so, Kalman will inevitably detect other time bins as anomalous, and these may be false positives.

We perform the following experiment. We define the set of anomalies found by equilibrium with the threshold value

³Prefixes are not available in the packets but are recorded in flow traces.

⁴The bug and its solution have been recently described in Juniper’s Problem Report 277942.

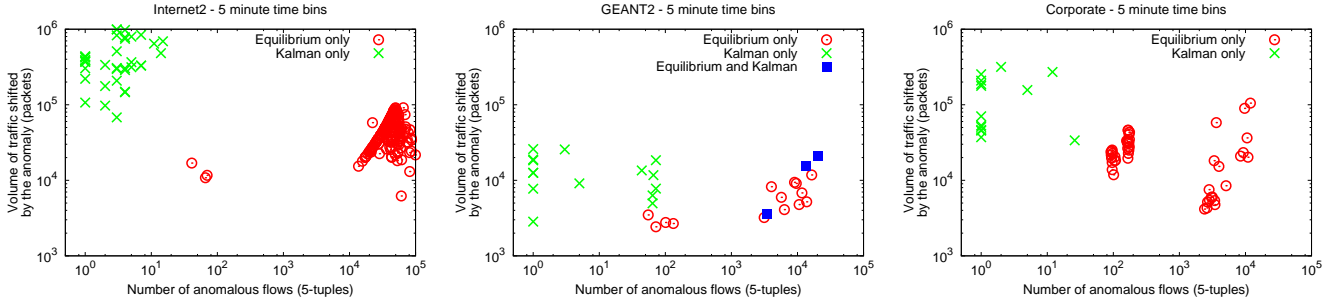


Fig. 5. Number of anomalous flows against the total volume change within these flows.

of six as our *target set* of anomalous time bins. Then, we lower Kalman’s detection threshold and measure the fraction of the target set that gets detected. We also measure the number of Kalman detections that are outside the target set. If this extra “noise” found by Kalman grows too large, then it is not practical to use that method to detect the anomalies in the target set. Figure 6 shows this result, with one curve per trace.

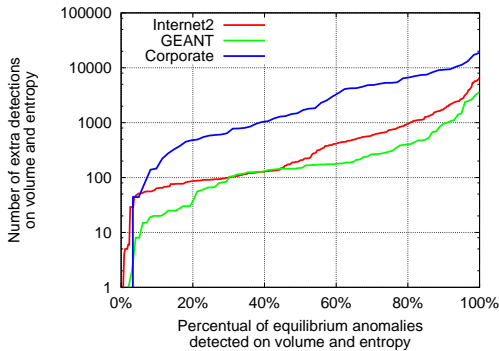


Fig. 6. Number of extra detections by Kalman in order to detect a fraction of the equilibrium anomalies.

The plot shows that if the Kalman threshold is tuned to catch even moderate fractions of the equilibrium anomalies in each trace, the amount of noise is indeed very high. In the corporate trace, to catch at least 80% of the equilibrium anomalies, Kalman triggers 6,600 extra detections, which is more than a hundred times the size of the target set. It is very likely that most of these detections contain no meaningful traffic anomalies.

VII. COMPUTATIONAL COMPLEXITY

The computational cost of our method is fairly small. There are two main steps to verify if a set of flows is in equilibrium. First, we need to compute the average and standard deviation of the volume changes across flows. This requires inspecting each packet header (or flow record in a flow trace) once. Note that any method based on entropy needs the same processing step to compute that metric [8]. In addition, multiple flow definitions can be analyzed in parallel with no extra cost as

each packet (or flow record) needs to be read only once. Table III shows the time in seconds to process all the flow records in the GEANT2 trace and to bin the flow volumes at each aggregation level in 5 minute time bins.⁵

TABLE III
DATA PROCESSING TIME PER FLOW AGGREGATION LEVEL
FULL GEANT2 TRACE - 5 MINUTE TIME BINS

Flow aggregation	Time (seconds)
5-tuples	491.78
Host pair	442.41
Source IPs	398.59
Source Ports	348.23
Dest. Ports	327.29
Dest. IPs	311.39

The second step of checking for equilibrium is computing the assessment value of Equation (7) and comparing it against the detection threshold. This involves only a square root, a multiplication and a division. Other anomaly detection methods in the literature are based on time series analysis, and perform either learning phases [8], [15] or frequency analysis [3] which require at least looking at each time bin within a calibration window (usually of a few days or weeks). For example, it takes 21 seconds to calibrate the Kalman filter of Section VI on the entire GEANT2 trace, while the equilibrium assessment values can be computed in less than one second.

VIII. RELATED WORK

Significant attention has been devoted to anomaly detection in recent years. For space reasons, we only briefly discuss this large body of work. Early anomaly detection techniques relied only on volume metrics such as packet and byte counts [4], [3]. Lakhina et al. [8] showed that the entropy of feature distributions (e.g., IP addresses and ports) extends the set of detections by volume metrics. However, Nychis et al. [10] have recently shown that the entropy of packet features has some limitations. They showed on traffic traces that the entropies of flow size and degree distributions can detect low-volume anomalies that go unnoticed in the entropies of addresses and ports. They have confirmed these results through controlled experiments with synthetically generated anomalies, and shown

⁵All times were measured on an 3 GHz AMD Opteron processor.

that most of the anomalies found by feature entropies are large enough to be detected as volume anomalies. Our main result is consistent with this since equilibrium reveals a larger variety of anomalies than volume and entropy. Interestingly, the two conclusions were obtained using different detection methods: while we have used the Kalman filter, Nychis et al. have used both Wavelets and deviation scores.

Prior work has explored correlation across flows in backbone links. Barakat et al. [2] have shown empirically that some features of the arrival process of 5-tuples, such as the inter-arrival time, display little or no correlation across different flows in highly aggregated backbone links. However, correlated flows can result from network events and attacks. These violate a condition for flow equilibrium and we are able to identify them using our methodology.

Finally, a large body of work in the literature has been devoted to understanding self-similarity [9] in network traffic. Self-similar behavior is associated with long-range dependence, i.e., significant *temporal* correlations in traffic. In this paper, however, we investigate correlation between two or more concurrent individual *flows*. We emphasize that these two types of correlation are orthogonal. In fact, one can also find self-similarity in large aggregates of independent sources [16]. Therefore, the conditions we study to observe flow equilibrium do not contradict the long-range dependence of traffic.

IX. CONCLUSIONS

We have introduced an anomaly detection technique based on a traffic property named flow equilibrium. We have shown through simple models that large sets of correlated flows violate this traffic property, and we presented a practical method for detecting these events. A drawback of this method is that it applies only to uncongested links with a large number of flows. We compared the anomalies found by equilibrium with those detected by a statistical technique based on volume and entropy. Our results show that equilibrium finds different types of anomalies compared to the reference technique. A distinguishing feature of our method is that it does not require a learning phase and thus is very simple to compute. We exploit this simplicity to compute our method on multiple flow aggregation levels, and this enables us to manually identify the root causes of a large number of traffic anomalies. Our results have given us insights for designing an automated technique for root cause analysis, which we leave for future work.

REFERENCES

- [1] NIST/SEMATECH e-Handbook of Statistical Methods (Online version). <http://www.itl.nist.gov/div898/handbook/>.
- [2] C. Barakat, P. Thiran, G. Iannaccone, C. Diot, and P. Owezarski. A flow-based model for Internet backbone traffic. In *Proceedings of IMW*, 2002.
- [3] P. Barford, J. Kline, D. Plonka, and A. Ron. A signal analysis of network traffic anomalies. In *Proceedings of IMW*, pages 71–82, 2002.
- [4] J. D. Brutlag. Aberrant behavior detection in time series for network monitoring. In *Proceedings of LISA*, pages 139–146, 2000.
- [5] J. Cao, W. S. Cleveland, D. Lin, and D. X. Sun. On the nonstationarity of Internet traffic. In *Proceedings of SIGMETRICS*, June 2001.
- [6] I. Cunha, F. Silveira, R. Oliveira, R. Teixeira, and C. Diot. Do you trust what flow measurement tools tell you? Technical report, Thomson, 2008.
- [7] W. Feller. *An introduction to probability theory and its applications (2 vols)*. John Wiley & Sons, 3 edition, 1968.
- [8] A. Lakhina, M. Crovella, and C. Diot. Mining anomalies using traffic feature distributions. In *Proceedings of SIGCOMM*, August 2005.
- [9] W. E. Leland, M. S. Taqqu, W. Willinger, and D. V. Wilson. On the self-similar nature of Ethernet traffic (extended version). *Transactions on Networking*, 2(1):1–15, 1994.
- [10] G. Nychis, V. Sekar, D. G. Andersen, H. Kim, and H. Zhang. An empirical evaluation of entropy-based anomaly detection. In *Proceedings of IMC (to appear)*, 2008.
- [11] H. Ringberg, A. Soule, J. Rexford, and C. Diot. Influence of data-reduction techniques on traffic anomaly detection. Technical report, Thomson, January 2007.
- [12] C. Shannon, E. Aben, kc claffy, D. Andersen, and N. Brownlee. CAIDA OC48 Traces Dataset. <http://www.datcat.org>.
- [13] F. Silveira, C. Diot, N. Taft, and R. Govindan. Empirical evaluation of network-wide anomaly detection. Technical report, Thomson, 2008.
- [14] R. Sommer and A. Feldmann. NetFlow: information loss or win? In *Proceedings of IMW*, pages 173–174, 2002.
- [15] A. Soule, K. Salamatian, and N. Taft. Combining filtering and statistical methods for anomaly detection. In *Proceedings of IMC*, 2005.
- [16] W. Willinger, M. S. Taqqu, R. Sherman, and D. V. Wilson. Self-similarity through high-variability: statistical analysis of Ethernet LAN traffic at the source level. *Transactions on Networking*, 5(1):71–86, 1997.
- [17] Y. Zhang, Z. Ge, A. Greenberg, and M. Roughan. Network anomography. In *Proceedings of IMC*, pages 1–14, 2005.