Degree-based Outlier Detection within IP Traffic Modelled as a Link Stream
Audrey Wilmet, Tiphaine Viard, Matthieu Latapy, Robin Lamarche-Perrin

To cite this version:

HAL Id: hal-02172934
https://hal.archives-ouvertes.fr/hal-02172934
Submitted on 4 Jul 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Degree-based Outlier Detection within IP Traffic Modelled as a Link Stream

Audrey Wilmet*, Tiphaine Viard†, Matthieu Latapy*, Robin Lamarche-Perrin‡
*Sorbonne Université, CNRS, Laboratoire d’Informatique de Paris 6, LIP6, F-75005 Paris, France
†Institut des Systèmes Complexes de Paris Ile-de-France, ISC-PIF, UPS 3611, Paris, France
‡Discrete Optimization Unit, Riken AIP, Tokyo, Japan, tiphaine.viard@riken.jp

Abstract—This paper aims at precisely detecting and identifying anomalous events in IP traffic. To this end, we adopt the link stream formalism which properly captures temporal and structural features of the data. Within this framework, we focus on finding anomalous behaviours with respect to the degree of IP addresses over time. Due to diversity in IP profiles, this feature is typically distributed heterogeneously, preventing us to directly find anomalies. To deal with this challenge, we design a method to detect outliers as well as precisely identify their cause in a sequence of similar heterogeneous distributions. We apply it to several MAWI captures of IP traffic and we show that it succeeds in detecting relevant patterns in terms of anomalous network activity.

I. INTRODUCTION

Temporal and structural features of IP traffic are and have been for several years the subject of multiple studies in various fields. A significant part of this research is devoted to detecting statistically anomalous traffic subsets referred to as anomalies, events or outliers. Their detection is particularly important since, in addition to a better understanding of IP traffic characteristics, it could prevent attacks against on-line services, networks and information systems. Due to the temporal and structural nature of IP traffic, developed methods can be classified into two categories: those based on signal processing [7], [10], and those based on graph theory [27], [50]. However, methods within these areas lead to a loss of information: by considering interactions as a signal, the structure is aggregated; with graphs, the temporal order of interactions is lost. Hence, existing methods struggle to identify subtle outliers, which are abnormal both in time and structure. Moreover, when they succeed, the loss of information leads to a decrease of accuracy. In this paper, we model IP traffic as a link stream which fully captures both the temporal and the structural nature of traffic [28], [44]. Then, we introduce a method in order to detect subtle outliers and precisely identify which IP addresses and instants caused it. More specifically, a link stream \( L \) is defined as a set of instants \( T \), a set of nodes \( V \) (IP addresses) and a set of interactions \( E \) (communication between IP addresses over time). Within this framework, we focus on one key property: the degree of nodes. This feature is highly heterogeneous, which raises challenges for its use in outlier detection, but it is stable over time. Our method takes advantage of this temporal homogeneity: it divides the link stream into time slices and then performs outlier detection to find time slices which exhibit unusual number of nodes having a degree within specific degree classes. Then, in order to isolate responsible IP addresses and instants on which they behave unexpectedly, we design an identification method based on an iterative removal of previously detected events. Finally, we validate our method by showing that these event removals do not significantly alter the underlying normal traffic.

This paper is an extended version of the work published in [49]. In this contribution, special attention is paid to the importance of parameters involved in the method such as time slice and degree class sizes. Moreover, in addition to our previous work in which we evaluated our method on a one-hour long IP traffic trace of June 2013, we apply it on two other datasets: a one-day long IP traffic trace of June 2013 and a fifteen-minutes long IP traffic trace of November 2018 which possesses a list of abnormal events indexed by MAWILab to which we compare our results [17].

The paper is organised as follows. We overview the related work in Section II and present our contributions in Section III. We introduce IP traffic modelling as a link stream and the degree definition in Section IV. In Section V, we describe our goals and the challenges they raise. This leads to the development of our method to detect events in Section VI and to identify them in Section VII. We discuss our results in Section VIII. Subsequently, we apply our method on other datasets in Section IX. Then, we investigate the influence of different time slice sizes and different class constructions in Section X. We conclude in Section XI.

II. RELATED WORK

Techniques for anomaly detection in IP traffic are extremely diverse. Among those, methods using principal component analysis [26], [40], machine learning [48], data mining [29], signal analysis [7], [10] and graph-based techniques have been proposed. In this paper, we focus our related work on methods based on dynamic graphs. In this domain, authors traditionally study a sequence of graphs \( \{G_i\}_{i=1..k} \), such that each snapshot \( G_i = G_{i\tau..(i+1)\tau} \) contains interactions aggregated over time window \( T_i = [i\tau, (i + 1)\tau] \). Then, they attribute an abnormality score to each snapshot \( G_i \) by comparing it to others. This problem has been approached in various ways depending on the definition of the abnormality
Compression-based abnormality scores analyse the evolution of the encoding cost of each graph to detect anomalies. Sun et al. [43] and Duan et al. [15] group similar consecutive snapshots into a chain. If the adding of a graph greatly increases the description length of the chain, the corresponding snapshot is considered as abnormal. Chakrabarti et al. [12] use a similar technique but apply it on clusters of nodes to find abnormal links. Other approaches use tensor decomposition. Ide et al. [23] and Ishibashi et al. [24] build a past activity vector from the main eigenvectors associated to all snapshots within a given time window. Then, snapshot \( G_i \) is identified as abnormal when the distance between the past activity vector and \( G_i \)'s main eigenvector exceeds a certain threshold. Akoglu et al. [3] use a similar technique in which the past activity vector includes nodes local features.

Outliers can also be found by studying communities evolution. Aggarwal et al. [2] find anomalous snapshots by comparing their clustering quality. Gupta et al. [20], [19] calculate the variation in the probability of belonging to a community for each node between two consecutive snapshots. Nodes for which the variation deviates significantly from the average variation of nodes within the same community are considered abnormal. Chen et al. [14] and Araujo et al. [5], in turn, detect abnormal communities among clusters which unusually increase, merge, decrease or split. Finally, a significant amount of work quantify the distance between snapshots using graph features. Pincombe et al. [36] and Papadimitriou et al. [35] define a series of topological aggregated features to compare snapshots. Berlingerio et al. [9] use the moments of an egonet feature – e.g., degree, clustering coefficient – calculated on each node. Saxena et al. [41] use a similar method but decompose each snapshot into \( k \) cores to consider global features as well. Schieber et al. [42] use the Jensen-Shannon divergence and a measure of the heterogeneity of each graphs in terms of connectivity distance between nodes. Finally, Mongiovi et al. [33] find clusters of anomalous links by calculating, for each link, its probability to have a given weight according to its usual behaviour.

However, these techniques lead to a loss of information: by reducing interactions into a sequence of graphs, the links order of arrival within a time window is lost. To overcome this issue, other work propose to improve these methods by introducing sequences of augmented graphs. For instance, Casteigs et al. [11] and Batagelj et al. [8] use graphs in which links are labelled with their instants of occurrence. Likewise, in [6], [46] and [47], authors use causal graphs in which two nodes are linked together if there is a causal relationship between them. In this paper, we adopt a new perspective. We consider temporal interactions as a separate object called a link stream, using the formalism developed by Latapy et al. [28]. While methods for covering relevant and subtle events often go hand in hand with very complicated features, we show that we can find relevant structural and temporal outliers, as well as gain accuracy, with a very simple feature defined in the link stream formalism: the instantaneous degree of nodes over time.

Other authors detect outliers using this modelling. Yu et al. [51] calculate the main eigenvector of the ego-network of each node and find abnormal nodes among those experiencing a sudden change in the amplitude and/or direction of their vectors. Manzoor et al. [31] use a similar technique. They store the link stream in a sketch built from the ego-networks patterns of each node and label a new edge as abnormal if the difference between the sketch before its arrival and the one after is significant. Ranshous et al. [38] also use sketches. They store the link stream in a Count Min sketch that approximates the frequency of links and nodes. From this sketch, they assign an abnormality score to each link \((u, v, t)\) based on prior occurrences, preferential attachment and mutual neighbours of nodes \(u\) and \(v\). Eswaran et al. [16] also rely on approximations and attribute a score to every new edge arriving in the stream by relying on a sub-stream \(L'\) sampled from past edges. If the new edge connects parts of \(L'\) which are sparsely connected, then it is considered as abnormal. Finally, Viard et al. [44] find anomalous bipartite cliques using the link stream formalism developed in [28].

A large proportion of the methods cited above is devoted to find globally anomalous instants (as abnormal snapshots). Among those extracting local features on nodes or links, either authors use similarity functions which aggregate local information, or they rely on approximations as samples or sketches. In the first case, instants are abnormal based on their local patterns but information about which sub-graphs are responsible is lost. In the second case, approximations allow a fast processing but lead to a decrease of accuracy. In contrast, our method identify abnormal couples \((t, v)\) exactly, without any information loss and still exhibits fast and efficient processing.

### III. Contributions

We model IP traffic with a link stream and study one of its most important properties, the degree of nodes over time. We show that, although this property follows a very heterogeneous distribution that is hard to model, this distribution is stable over time. We then design a method that exploits the stability of this heterogeneity for anomaly detection, and may be applied in various such situations. This method first splits traffic into time slices and computes the degree distribution in each slice. By comparing these distributions, the method then points out degree classes and time slices such that having a degree in this class during this slice is anomalous. Using this information, we identify IP addresses and time periods involved in anomalies, as well as the corresponding traffic. By removing this traffic from the original data, we validate our identification by noticing that we turn back to a normal
traffic with respect to the degree. We illustrate the method and its outcome on MAWI public IP traffic.

IV. TRAFFIC MODELLED AS A LINK STREAM

IP traffic consists of packet exchanges between IP addresses. We use here one hour of IP traffic capture from the MAWI archive\(^1\) on June 25\(^b\), 2013, from 00:00 to 01:00. We denote this trace by a set \(\mathcal{D}\) of triplets such that \((t, u, v) \in \mathcal{D}\) indicates that IP addresses \(u\) and \(v\) exchanged a packet at time \(t\). The set \(\mathcal{D}\) contains 83,386,538 triplets involving 1,157,540 different IP addresses.

We model this traffic as a link stream \(L\) in order to capture its structure and dynamics [28]. Nodes are IP addresses involved in \(\mathcal{D}\) and two nodes are linked together from time \(t_1\) to time \(t_2\) if they exchanged at least one packet every second within this time interval. Formally, \(L = (T, V, E)\) is defined by a time interval \(T \subseteq \mathbb{R}\), a set of nodes \(V\) and a set of links \(E \subseteq T \times V \times V\) where \(V \times V\) denotes the set of unordered pairs of distinct elements of \(V\), denoted by \(uv\) for any \(u\) and \(v\) in \(V\) (thus, \(uv \in V \times V\) implies that \(u, v \in V\) and \(u \neq v\), and we make no distinction between \(uv\) and \(vu\)). If \((t, uv) \in E\), then \(u\) and \(v\) are linked together at time \(t\). In our case, we take \(E = \bigcup_{(t,u,v) \in \mathcal{D}} [t - \frac{\Delta}{2}, t + \frac{\Delta}{2}] \times \{uv\}\) with \(\Delta = 1s\). Other choices can be made. For instance, we can set a value of \(\Delta\) that is different for each \(uv\), or each link \((t, uv)\), using external knowledge. We can also use a value of \(\Delta\) that changes over time (see for instance the work of Léon et al. [30]). These operations are depicted in Figure 1.a.

The degree of \((t, v) \in T \times V\), denoted by \(d_t(v)\), is the number of distinct nodes with which \(v\) interacts at time \(t\):

\[d_t(v) = |\{u : (t, uv) \in E\}|.\]

\(^1\)http://mawi.wide.ad.jp/mawi/ditl/dit2013/ [25]

Figure 1.b shows the degree of node \(b\) over time. Notice that the degree of \(b\) is not its number of exchanged packets over time; it accounts for its number of distinct neighbours over time.

V. HETEROGENEITY OF DEGREES

In order to find outliers in a link stream using the degree, we first need to characterize the normal behaviour of the set of observations \(O = \{d_t(v) : (t, v) \in T \times V\}\). Then, an outlier is a couple \((t, v) \in T \times V\) which has a significantly different degree from others.

For this purpose, we call degree distribution of \(L\) the fraction \(f(k)\) of couples \((t, v) \in T \times V\) for which \(d_t(v) = k\), for all \(k \in \mathbb{N}\):

\[f(k) = \frac{|\{(t, v) \in T \times V : d_t(v) = k\}|}{|T \times V|}.\]

Figure 2 shows that the degree distribution is very heterogeneous, which discards the hypothesis of a normal behaviour. In this situation, one may hardly identify values of degree that could be considered anomalous.

A solution is to fit this distribution, and then find values which deviates from the model. Given its heterogeneity, one may think that it is well fitted by a power law distribution \(P(k) \propto k^{-\alpha}\) where \(\alpha > 1\) and \(k_{\max} \geq k \geq k_{\min} > 0\). However, we show that this is not the case following the procedure proposed by Virkar et al. [45]. Results show that differences between the empirical distribution and the estimated model cannot be attributed to statistical fluctuations, which leads us to reject the hypothesis that the degree is distributed according to a power law distribution.

\[\text{Figure 2: Degree distribution and complementary cumulative degree distribution in } L. \text{ For all } (t, v) \in T \times V, \text{ we compute the degree } d_t(v) \text{ and plot the distribution of the set of values } O = \{d_t(v) : (t, v) \in T \times V\}. \text{ The fraction expresses the probability to draw a time instant } t \in T \text{ and a node } v \in V \text{ such that } d_t(v) = k.\]

This shows that finding outliers in this type of distribution is not trivial. In order to circumvent this issue, we observe degrees on sub-streams corresponding to IP traffic during time.
the sizes of the two samples. With a significance level of two-sample Kolmogorov-Smirnov (KS) tests on all pairs have similar shapes. To quantify this similarity, we perform Nonetheless, Figure 3 also shows that degree distributions that these distributions still are heterogeneous.

Degree distribution and complementary cumulative distribution of the set of values \( O_i = \{ d_i(v) : (t, v) \in T_i \times V \} \), for \( i = \{0, 1, 2, 3\} \). The fraction expresses the probability to draw a time instant \( t \in T_i \) and a node \( v \in V \) such that \( d_i(v) = k \).

slices of duration \( \tau = 2.0 \)s. Formally, we call \( T_i = [2i, 2i + 2[ \) the \( i^{th} \) time slice, for all \( i \in \{0, \ldots, 1799\} \), and we define 

\[
 f_i(k) = \frac{|\{(t, v) \in T_i \times V : d_i(v) = k\}|}{|T_i \times V|},
\]

the degree distribution of the \( i^{th} \) time slice. Figure 3 shows that these distributions still are heterogeneous.

Nonetheless, Figure 3 also shows that degree distributions \( f_i \) have similar shapes. To quantify this similarity, we perform two-sample Kolmogorov-Smirnov (KS) tests on all pairs of distributions \( (f_i, f_j)_{i \neq j} \) [37]. According to the relative position between the KS distance \( D_{i,j} \) and a critical value \( c \), this test assess whether two samples may come from the same distribution or not. Let \( m = k_{\text{max}} \) and \( n = k_{\text{max}} \) be the sizes of the two samples. With a significance level of 0.1, \( c = 1.073 \sqrt{\frac{n + m}{nm}} \) [1]. Figure 4 shows the ratio between \( D_{i,j} \) and \( c \). Most \( D_{i,j} \) are below \( c \). This means that most samples \( O_i = \{ d_i(v) : (t, v) \in T_i \times V \} \) are drawn from the same distribution. On the contrary, some of them are different from all others which in turn, indicate changes in the overall behaviour on particular sub-streams.

Using these observations, we design below an outlier detection method based on the temporal homogeneity of heterogeneous degree distributions.

VI. LEVERAGING TEMPORAL HOMOGENEITY TO DETECT EVENTS

The above observations lead to the following conclusion: degree distributions are heterogeneous in the same way on most, if not all, time slices. In other words, in each time slice, the fraction of couples \((t, v)\) that have a given degree is similar to this fraction in other time slices. This is what we will consider as normal. Anomalies, instead, correspond to significant deviation from the usual fraction of nodes having a given degree. In this section we describe our method to compare degree distributions on all time slices and its use for outlier detection.

First, notice that it makes little sense to consider the fraction of couples \((t, v)\) having a degree exactly \( k \) when \( k \) is large: having degree \( k - 1 \) or \( k + 1 \) makes no significant difference. Therefore, to increase the likelihood of observing values in the tail of the distribution, we define logarithmic degree classes \( C_j \) and consider the fraction of couples \((t, v)\) having degrees in \( C_j \), for all \( j \):

\[
 f_i(C_j) = \frac{|\{(t, v) \in T_i \times V : d_i(v) \in C_j\}|}{|T_i \times V|}.
\]

We define the \( j^{th} \) degree class, \( C_j = \{k_j, \ldots, k_{j+1}\} - 1 \) such that \( k_1 = 1 \) and \( \log(k_{j+1}) = \log(k_j) + r \) where \( r = 0.1 \) is the degree class size. This leads to \( C_1 = \{1\}, C_2 = \{2\}, C_3 = \{3\}, C_4 = \{4, 5\}, \) etc., until \( C_{41} = \{19953, \ldots, 25117\} \). Then, to compare degree distributions, we plot for a given degree class \( C \), the distribution on all time slices \( T_i \) of the fraction \( f_i(C) \) in \( C_j \).

In other words, we study how the fraction of couples \((t, v)\) which have a degree within \( C \) during \( T_i \) is distributed among all time slices.

Figure 5 shows the distributions for classes \( C_1, C_2, C_{19}, C_{22}, C_{31} \) and \( C_{41} \). In accordance with temporal homogeneity, we can see that most fractions are distributed around the mean and that only a few are distant from it. As expected
In order to validate fractions \( f_i \) homogeneity over time slices within each degree class, we fit their distributions with a normal distribution model \( P(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2} \left( \frac{x-\mu}{\sigma} \right)^2} \) where values are normally distributed around a mean \( \mu \) with a standard deviation \( \sigma \). Deciding whether a given distribution is homogeneous with outliers or not may be done as follows [27]: (1) Iteratively remove outliers from the distribution with Grubbs test [18]; (2) Fitting the resulting distribution with the normal model; (3) Evaluate the goodness of the fit. We use Maximum Likelihood Estimation (MLE) to determine which model parameters fit the best the empirical distribution and evaluate the goodness of the fit with the KS test between empirical and estimated distributions. In this framework, we find 37 distributions that are homogeneous with outliers among the 41 corresponding to each degree class (see Figure 6). The remaining 4 classes are discarded from the study, we call them \( R \)-classes for rejected classes. One may use more complex and accurate techniques to automatically perform the fit, see for instance the work performed by Motulsky et al. [34].

Given an homogeneous distribution with outliers, we use here the classical assumption that a value is anomalous if its distance to the mean exceeds three times the standard
deviation [13], [21]. In classes displayed in Figure 5, we obtain 151 time slices flagged as anomalous in the first class containing degree 1 only, 5 anomalous time slices in $C_2$ and 12 in $C_{19}$. In $A$-classes, peaked on 0, anomalous values correspond to all values greater than 0.

All in all, our method for event detection from degree distributions is the following: we group degree values into degree classes of logarithmic width. For a given degree class $C$, we look at the distribution on all time slices of the fraction $f_i(C)$. This distribution indicates anomalous values which mean that there are anomalous high numbers of couples $(t, v)$ having degree within $C$ during specific time slices $T_1$. We then call an anomalous value of this kind a detected event and denote it $(C, T_i)$.

A detected event gives two pieces of information: the time slice $T_i$ on which the anomalous value has been observed, and the degree class $C$ in which the couples responsible for the high fraction are located. At this stage, we detected 1,358 such events. However, a time slice and a degree class are not sufficient information to accurately characterize the anomaly. We now address the goal of identifying couples $(t, v)$ in $T \times V$ responsible for these detected events.

VII. ITERATIVE REMOVAL TO IDENTIFY EVENTS

A detected event $(C, T_i)$ is a degree class $C$ and a time slice $T_i$ such that the fraction $f_i(C)$ is unusually high compared to the ones in other time slices. Identifying this event means recovering the set $\mathcal{I}(C, T_i)$ of couples $(t, v)$ responsible for this anomaly. In this section, we introduce an iterative removal method and show that it leads to such identification.

Let us take event $(C_2, T_{1080})$ as an example. We have access to the set of couples $(t, v)$ which have a degree in $C_2$ during $T_{1080}$. However, we cannot directly identify the event by this set. Indeed, let us consider the new link stream $L'$ in which we removed the corresponding interactions: $L' = (T, V, E')$ with $E' = E \setminus \{(t, uv) : t \in T_{1080} \text{ and } d_i(v) \in C_2\}$. We see in Figure 7 that the removal of this set of interactions from the link stream causes the appearance of a negative outlier in the distribution of fractions on $C_2$. Thus, by removing all interactions $(t, uv)$ such that couples $(t, v)$ have degree in $C_2$ during $T_{1080}$, we removed anomalous traffic but also normal traffic. Therefore, identifying the detected event $(C_2, T_{1080})$ as the set $\mathcal{I}(C_2, T_{1080}) = \{(t, v) : t \in T_{1080} \text{ and } d_i(v) \in C_2\}$ is not accurate enough.

This suggests that one cannot directly identify couples acting abnormally in $AN$-classes. Indeed, in these classes, the normal fraction is greater than zero. Hence, an anomalous fraction consists in anomalous couples but also normal ones, which prevents us from identifying responsible couples without disrupting normal traffic.

On the contrary, in $A$-classes the expected fraction is zero. Therefore, couples $(t, v)$ contributing to non-zero fractions are clearly anomalous. Events detected in such degree class $C$ can therefore be correctly identified with the set $\mathcal{I}(C, T_i) = \{(t, v) : t \in T_i \text{ and } d_i(v) \in C\}$. Thus, we now consider a class $C_{41}$. Its larger anomalous fraction corresponds to time slice $T_{315}$. Hence, this event can be identified by the set $\mathcal{I}(C_{41}, T_{315}) = \{(t, v) : t \in T_{315} \text{ and } d_i(v) \in C_{41}\}$. Figure 8 shows the consequences of the removal of these abnormal couples activities. As expected, the anomalous fraction in $C_{41}$ vanishes without creating a negative outlier. Additionally, we notice the disappearance of event $(C_1, T_{315})$. Indeed, removed nodes $v$ were linked to an unexpectedly large number of nodes having degree 1 before the removal. Then, the removal of the set $\mathcal{I}(C_{41}, T_{315})$ leads to the identification of event $(C_1, T_{315})$ such that $\mathcal{I}(C_1, T_{315}) = \{(t, u) : t \in T_{315}, u \in N_i(v), d_i(v) \in C_{41} \text{ and } d_i(u) \in C_1\}$, where $N_i(v)$ is the set of neighbours of $v$ at time $t$.

All in all, our approach for event identification is
detected ones, hence more than the removals on the average degree per second.
which they were acting abnormally.

of 4 nodes which have been removed for time periods during all outliers disappeared. Figure 10 shows the degree profiles C

7  

Fig. 9 the final shape of classes

1205 events in A classes. Altogether, we directly identified and removed outliers (this is not what we observe for all datasets, see Section IX). Altogether, we directly identified and removed 205 events in A-classes. These removals allowed us to identify a total of 1,163 outliers on the 1,358 previously detected ones, hence more than 85% of detected outliers. To do so, we removed 7.4% of all the traffic. We can see in Figure 9 the final shape of classes C1 and C2 in which almost all outliers disappeared. Figure 10 shows the degree profiles of 4 nodes which have been removed for time periods during which they were acting abnormally.

In our dataset, none of the removals generated negative outliers (this is not what we observe for all datasets, see Section IX). Altogether, we directly identified and removed 205 events in A-classes. These removals allowed us to identify a total of 1,163 outliers on the 1,358 previously detected ones, hence more than 85% of detected outliers. To do so, we removed 7.4% of all the traffic. We can see in Figure 9 the final shape of classes C1 and C2 in which almost all outliers disappeared. Figure 10 shows the degree profiles of 4 nodes which have been removed for time periods during which they were acting abnormally.

VIII. VALIDATION

The IP traffic trace we use does not have a ground truth dataset listing abnormal IP addresses and instants. However, we can validate our results by looking at the consequences of the removals on the average degree per second.

Let \( d_i \) be the instantaneous average degree at time \( t: d_i = \frac{1}{t} \sum_v d_i(v) \). The average degree during second \( s_i = [i, i+1] \), denoted by \( d(s_i) \), is the average of \( d_i \) from \( t = i \) to \( t = i + 1 \), for all \( i \in \{0, \ldots, 3599\} \):

\[
d(s_i) = \int_{i}^{i+1} d(t) \, dt .
\]

We see in Figure 11.a that the average degree per second is homogeneously distributed with outliers. After removing events identified with our method, we see in Figure 11.b that peaks as well as sudden changes in the trend disappear, while the average over time stays unaltered. Likewise, in

Fig. 9: Distributions of fractions on all time slices for degree classes C1 and C2 after event removals - Before event removal there were 151 anomalous values in C1 and 5 in C2. After removal, it only remains 10 unidentified anomalous values in C1 and 2 in C2.

Fig. 10: Degree profiles of four identified nodes - \( v_1 \) is responsible for outliers observed in \( C_{22} = \{252, \ldots, 2510\} \). The set \( \{(t, v_1): t \in [712, 940] \text{ and } d_t(v) \in C_{22}\} \) has been identified and removed. \( v_2 \) has a normal activity with a degree around 160 and a sharp variation on \( T_{23} = [446, 448] \). The set \( \{(t, v_2): t \in T_{23} \text{ and } d_t(v) \in C_{22}\} \) has been identified and removed. Sets \( \{(t, v_3)\} \) where \( v_3 \) is active have all been identified and removed. For node \( v_4 \), the four peaks corresponding to degree values higher than 300 have been identified and removed. The degree profiles of these nodes suggest that they constitute malicious activity [22], [32]. Node \( v_3 \) reaches several powers of two indicating that it is running network scans. We observe a similar behaviour around a degree of 256 for node \( v_1 \).

Fig. 11: Consequences of event removals on the average degree per second - Our method succeeds in removing identified anomalies with no significant impact on the underlying normal traffic.

a) Distribution of the average degree per second

b) Time evolution of the average degree per second
the distribution, all outliers disappear but the bell curve stays the same. Quantitatively, we find 33 outlying seconds before removals versus 5 after applying our method. The average over time of the average degree per second is equal to $3.39571 \times 10^{-4}$ before removals, and to $3.31795 \times 10^{-4}$ after removals. These results show that our method succeeds in removing abnormal traffic without altering the underlying normal traffic.

These results highlight another important fact. An outlier in distribution of Figure 11.a means that there is a second during which the average degree is larger than usual. This event is detected but not identified since we cannot trace back responsible nodes with this aggregated feature. By using the instantaneous degree on couples $(t, v)$, our method is able to identify events unidentified with the average degree per second. This last result is particularly promising: it shows that by using more complex and less aggregated features, it is possible to identify events previously detected but unidentified with simpler metrics. Then, the 195 events that we were not able to identify with the instantaneous degree of nodes could therefore be identified in future works by using other link stream features.

IX. OTHER DATASETS

To test the generality and applicability of our method, we test it on other datasets from the MAWI archive. In this section, we present the main results and differences we observe with these datasets.

A. One day long IP traffic trace from June 2013

We use here a one day long IP traffic capture from the MAWI archive from June 25th, 2013, at 00:00 to June 26th, 2013, at 00:00. The set $D$ contains 2,196,079,591 triplets involving 15,390,298 different IP addresses. This dataset is larger than the first one and covers one day of IP traffic with its circadian cycle. We keep identical time slices size and degree classes size.

Figure 12.a (left) shows, for $C_2 = \{2\}$, the distribution on all time slices $T_i$ of the fraction $f_i(C_2)$. Partly as a result of circadian cycles, we see that this distribution consists of three normal distributions. To address this issue, we normalize the degree with the average degree per second and consider the normalized degree, denoted by $\bar{d}_i(v)$, such that

$$\bar{d}_i(v) = \frac{d_i(v)}{d_i([t])},$$

where $[t]$, the floor function of $t$, is the second to which $t$ belongs. We see in Figure 12.a (right) that local distributions on time slices are similar and in Figure 12.b that the global normalized degree distribution is heterogeneous. Thus, the two constraints required to apply our method are met. We find 34 degree classes, from $C_1 = \{1\}$ to $C_{34} = \{3982, \ldots, 5011\}$. Among these, 3 classes are rejected because they do not fit with an homogeneous distribution with outliers. We find 11 $AN$-classes from $k = 1$ to $k = 26$ and 20 $A$-classes from $k = 51$ to $k = 5011$. We detect 22,669 outliers and succeed in identifying 63% of them. To do so, we removed 8.7% of all the traffic. Once again, we see in Figure 12.c that these removals lead to the cleaning of the average degree per second: 1,672 abnormal seconds before removals and 10 after. Likewise, normal traffic stays unchanged: the average over time of the average degree per second is equal to $1.53199 \times 10^{-5}$ before removals and to $1.48824 \times 10^{-5}$ after.

B. Fifteen-minute IP traffic trace from November 2018: comparison to MAWILab

We use here a fifteen-minute IP traffic capture from the MAWI archive on November 3rd, 2018, from 14:00 to 14:15. The set $D$ contains 64,913,871 triplets involving 16,453,608 different IP addresses. This dataset is more recent than the first one and has a list of anomalies indexed by MAWILab [17] to which we can compare our results. Given the shorter temporal extent, we take time slices of size $\tau = 1.0s$ instead of $\tau = 2.0s$, in order to keep a significant number of time slices. Degree classes stay unchanged.

We observe an heterogeneous global degree distribution and similar local degree distributions on time slices of 1.0 seconds (see Figures 13.a and 13.b). We find 43 degree classes, from $C_1 = \{1\}$ to $C_{43} = \{31623, \ldots, 39810\}$. Among these, there are 23 $AN$-classes, 17 $A$-classes and 3 $R$-classes.

Contrarily to previous datasets, we observe three $AN$-classes in high degree classes: $C_{24} = \{399, \ldots, 502\}$,
We can see in Figure 13.d, that this is due to three nodes which have a constant degree fluctuating within each of these classes and, as a result, form the observed normal traffic.

In this dataset, several removals generated negative outliers. For instance, the removal of event \((C_{40}, T_{752})\) generated a negative outlier in class \(C_1\). The corresponding event was incorrectly identified by the set \(I(C_{40}, T_{752}) = \{(t, v) : t \in T_{752} \text{ and } d_t(v) \in C_{40}\}\). Indeed, this event corresponds to the spike of activity of node \(v_3\) from 755.3 to 756.5 (see Figure 13.d). Yet, the normal behaviour of \(v_3\) is to be linked to an average of 18,178 nodes of degree 1 over time. Thus, the removal of its activity during this time period leads to a negative outlier in \(C_1\), since, in addition to removing abnormal interactions of \(v_3\), it also removes its legitimate interactions. This shows that to identify event \((C_{40}, T_{752})\), we need to use a finer and more complex feature than the degree.

Finally, our method enabled us to detect 827 outliers and identify 796 of them (96%). To do so, we removed 1.2% of all the traffic. We see in Figure 13.c that, as with the two other datasets, removals lead to a traffic free of most degree-related anomalies. The number of abnormal seconds is equal to 19 before removals versus 0 after removals. Likewise, the average over time of the average degree per second goes from 9.311934 \cdot 10^{-5} to 9.198076 \cdot 10^{-5} after removals.

This dataset contains a MAWILab database to which we can compare our results [17]. It lists and labels anomalies in traffic from the MAWI archive by using a graph-based methodology that compares and combines the output of several independent anomaly detectors. On November 3rd, 2018, from 14:00 to 14:15, it indicates a total of 287 anomalous IP addresses. To each of these is associated a time period during which it is evaluated as abnormal and a label classifying its anomaly type among the following categories [32]:

- Point to point denial of service: a large number of packets are sent between two IP addresses;
- Distributed denial of service: a large number of packets are sent between multiple sources and one destination;
- Network scan: an IP address scans a network of several destination IP addresses;
- Port scan: an IP address scans several ports of one destination;
- Point multipoint: normal router traffic;
- Alpha flow: normal peer to peer traffic;
- Other: normal outage traffic;

Since we do not consider the port number, and given that the degree feature does not account for the number of exchanged packets, anomalies within the point to point denial of service, port scan and alpha flows categories cannot be detected by our method. Moreover, we do not consider events corresponding to legitimate traffic. This reduces the number of identified IP addresses to 77.

With our method, we find 33 anomalous IP addresses. Six of them are not listed by MAWILab. They correspond to node \(v_2\) in Figure 13.d and nodes \(v_4, v_5, v_6, v_7\) and \(v_8\) in Figure 13.e (left). Node \(v_2\) has been removed during its spike of activity from 798.18 to 799.87. Likewise, node \(v_7\) has been removed from 814.06 to 815.00 and node \(v_8\) on the whole time period during which it is active. Note that, as we can see in Figure 13.e (middle and right), nodes \(v_7\) and \(v_8\) activities are typical of nodes performing network scans which are usually detected by MAWILab. The remaining three nodes \(v_4, v_5\) and \(v_6\), have been removed on periods of respectively 0.0708s, 0.0677s and 0.181s because of their ephemeral activity within the \(A\)-class \(C_{22} = \{252, \ldots, 317\}\). This could be avoided by using larger classes (see section X-B).

In the network scan category, we identified 24 IP addresses among the 76 (32.6%) listed by MAWILab. All network scans involving more than 250 different destinations have been identified with our method. As mentioned above, we identified in addition two IP addresses that the MAWILab detectors missed (see Figure 13.e). Moreover, for the corresponding events, our temporal precision is much better than the one provided by MAWILab detectors. However, our method fails to identify IP addresses permanently linked to the network since they have constant degree profiles and therefore lead to \(AN\)-classes. More generally, we did not find IP addresses which scan networks involving less than 250 destinations since all classes below \(C_{22} = \{252, \ldots, 317\}\) are \(AN\)-classes, and since their activities are not linked to the ones of removed events. Nonetheless, time slices during which most of these network scans occur have been detected as outliers in their corresponding degree class. This inability to identify low degree classes events could be avoided by using a feature different from the degree, in which the corresponding malicious activities deviate more significantly.

In the distributed denial of service category, only one anomaly is identified by MAWILab. The corresponding node have a maximum degree of 53. Hence, we do not find it with our method for the same reasons as above.

The six remaining nodes we identified fall in the point to multipoint category that we do not consider as it constitutes normal router traffic.

Finally, Figure 13.c (right) shows the average degree per second after removing events identified with our method as well as the ones identified in the MAWILab dataset. We see that, with MAWILab, the average over time is affected by the removals. This is mostly due to the poor time precision used by MAWILab to describe anomalies. Indeed, 63% of IP addresses are identified as abnormal on the whole trace, including IP addresses which have a global constant degree with only a few spikes. With our method, instead, when nodes
In this section, we perform a series of experiments on the first obtained in the first experiment, and other sets obtained by I. Variation of Time Slice Sizes are not removed and normal traffic stays unchanged. Due to aggregation over a larger period, it is expected that the larger \( \tau \) one time slice to another. Let \( \tau \) be the set of identified outliers using time slices of size \( \tau \). In order to evaluate the impact of \( \tau \), we measure the Jaccard similarity coefficient between \( \mathcal{I}_{2,0} \), obtained in the first experiment, and other sets obtained by varying \( \tau \):

\[
J(\mathcal{I}_{2,0}, \mathcal{I}_\tau) = \frac{|\mathcal{I}_{2,0} \cap \mathcal{I}_\tau|}{|\mathcal{I}_{2,0} \cup \mathcal{I}_\tau|}.
\]

Results are depicted in Figure 14.a. We see that identified sets \( \mathcal{I}_\tau \) are identical from \( \tau = 0.2 \) up to \( \tau = 20.0 \). This shows that our method is stable with respect to this parameter. Below this range, we are able to identify slightly more outliers. On the contrary, when the size increases, we identify less and less outliers until no more is identified after \( \tau = 175.0 \). This is explained by the number of \( AN, A \) and \( R \)-classes according to \( \tau \) in Figure 14.b: the more \( \tau \) increases, the higher the number of \( R \)-classes and the lower the number of \( A \)-classes in which we are able to identify events. When we reach \( \tau = 175.0 \), all classes are rejected, hence no outlier is detected. This increase in the number of rejected classes is provoked by the very small number of time slices when \( \tau \) gets larger. Indeed, time slices are insufficiently numerous to establish a normal behaviour and, for all classes \( C \), fits between fractions \( f_1(C) \) and a normal distribution are more likely to be rejected.

We identify more abnormal couples \((t, v)\) when \( \tau \) is small. However, this result should be taken with caution. As we can see in Figure 14.b, when \( \tau \) decreases, the number of rejected classes increases and the number of \( AN \)-classes decreases. Indeed, the smaller the time slice, the less the behaviour between time slices is similar, which leads to a rejection of normal behaviour. \( A \)-classes are not affected: in most time slices, there is no node that reaches a degree within the class, whatever the time slice size. Moreover, their number increases. This is explained in Figure 15. We see that for \( \tau = 0.25 \), there is 83\% of \( T_i \) for which the fraction \( f_1(C_{34}) \) is zero, against 67\% for \( \tau = 2.0 \). When \( \tau \) decreases, the proportion of time slices without traffic in the class compared to the ones that contain traffic is much higher than in experiments with a larger \( \tau \). If the increase of \( A \)-classes makes it possible to identify more outliers, the decrease of \( AN \)-classes, on the other hand, prevents us from determining if a removal is bad or not by the appearance of a negative

![Fig. 13: Fifteen minutes from November the 3rd 2018 - (a - b) The heterogeneity of the global degree distribution and the similarity of local degree distributions are verified. (c) Results are validated by the average degree per second after removals. Contrary to our method, we observe a decrease of the average over time when removing events identified by MAWILab. (d) AN-classes are observed in high degree classes: \( v_1 \) is responsible for the normal traffic observed in \( C_{24} = \{399, \ldots, 502\} \); \( v_2 \) for the one in \( C_{27} = \{795, \ldots, 1001\} \) and \( v_3 \) for the one in \( C_{40} = \{15849, \ldots, 19953\} \). (e) Nodes \( v_2, v_4, v_5, v_6, v_7 \) and \( v_8 \) are not detected by MAWILab. However, we see that the removal of the abnormal activity of \( v_2 \) is responsible for the disappearance of the spike around \( t = 800 \) in (c) and that \( v_7 \) and \( v_8 \) have a suspicious activity which is usually detected by MAWILab. Note that high degree node \( v_3 \) in (d) has been removed from the calculation in the average degree per second in (c) to have a greater clarity.](image_url)
To identify more outliers (observation 1), the decrease of $\tau$ increases (see Figure 16.a). For $r = 10^{-5}$, there are 23,983 classes; for $r > 1.6$, the total number of classes is smaller than 4 and reaches 1 for $r \geq 4.4$. Consequently, the smaller $r$, the higher the computation time.

When we look at the size and similarity of identified sets, we observe several phenomena (see Figure 16.b):
1) the number of identified outliers increases for $r < 0.2$;
2) we do not identify outliers for $r \in [2.2, 2.3]$ and $r > 4.4$;
3) the Jaccard index is higher than 0.8 for $r \in [10^{-5}, 1.6]$ and $r \in [2.4, 3.3]$;
4) the Jaccard index fluctuates between 0.8 and 1 for $r \in [0.2, 1.6]$;
5) the number of identified outliers drops from $r = 1.7$, increases from $r = 2.4$, then drops again from $r = 3.4$. Once again, these observations are linked to the proportions of the three classes types. We explain them in detail in the following.

When the degree class size is too small, classes do not integrate too much traffic. As a consequence, we identify more outliers (observation 1), the decrease of $\tau$ increases (see Figure 16.a). For $r = 10^{-5}$, there are 23,983 classes; for $r > 1.6$, the total number of classes is smaller than 4 and reaches 1 for $r \geq 4.4$. Consequently, the smaller $r$, the higher the computation time.

When we look at the size and similarity of identified sets, we observe several phenomena (see Figure 16.b):
1) the number of identified outliers increases for $r < 0.2$;
2) we do not identify outliers for $r \in [2.2, 2.3]$ and $r > 4.4$;
3) the Jaccard index is higher than 0.8 for $r \in [10^{-5}, 1.6]$ and $r \in [2.4, 3.3]$;
4) the Jaccard index fluctuates between 0.8 and 1 for $r \in [0.2, 1.6]$;
5) the number of identified outliers drops from $r = 1.7$, increases from $r = 2.4$, then drops again from $r = 3.4$. Once again, these observations are linked to the proportions of the three classes types. We explain them in detail in the following.

When the degree class size is too small, classes do not integrate too much traffic. As a consequence, we identify more outliers (observation 1), the decrease of $\tau$ increases (see Figure 16.a). For $r = 10^{-5}$, there are 23,983 classes; for $r > 1.6$, the total number of classes is smaller than 4 and reaches 1 for $r \geq 4.4$. Consequently, the smaller $r$, the higher the computation time.

When we look at the size and similarity of identified sets, we observe several phenomena (see Figure 16.b):
1) the number of identified outliers increases for $r < 0.2$;
2) we do not identify outliers for $r \in [2.2, 2.3]$ and $r > 4.4$;
3) the Jaccard index is higher than 0.8 for $r \in [10^{-5}, 1.6]$ and $r \in [2.4, 3.3]$;
4) the Jaccard index fluctuates between 0.8 and 1 for $r \in [0.2, 1.6]$;
5) the number of identified outliers drops from $r = 1.7$, increases from $r = 2.4$, then drops again from $r = 3.4$. Once again, these observations are linked to the proportions of the three classes types. We explain them in detail in the following.
The smaller discretization effects. In the one-hour traffic trace, classes are outliers and incorrect removals could be accepted which see in Figure 19, high degree classes are a degree larger than 1. We did not plotted the proportion of classes for 1 degree resulting distribution looks the same as when it only contains that when class AN 1-class. Moreover, we see in Figure 18 that when class R-classes and to a decrease of the number of A-classes. For r > 0.2, proportions of A and AN-classes are similar and the proportion of R-classes is low. Note that we did not plotted the proportion of classes for r > 1.4 because of fluctuations due to the small number of classes.

for r = 2.2, there are two AN-classes; for r = 2.3; there are one AN-class and one R-class; and finally, for r ≥ 4.4, there is only one AN-class. Moreover, we see in Figure 18 that when class C1 contains several values of degrees, the resulting distribution looks the same as when it only contains degree 1. Hence, the large proportion of couples (t, v) having degree 1 obstructs the traffic of couples (t, v) which have a degree larger than 1. As a consequence, we detect less outliers and incorrect removals could be accepted which could also lead to a disruption of normal traffic.

Finally, observations 3, 4 and 5 are explained by discretization effects. In the one-hour traffic trace, classes are arranged such that low degree classes are AN-classes and high degree classes are A-classes. Let kr id be the smallest degree from which we are able to identify events. As we can see in Figure 19, kr id is different depending on the chosen r. The smaller kr id, the larger the number of detected outliers. Until r = 0.2, kr id is lower than 250 and the number of identified outliers is maximal. Then, kr id fluctuates between 250 and 2,512, which explain observation 4. A lot of outliers are located within this range. Hence, if kr id ∈ [250, 2512], these outliers are identified, otherwise they are not, which explains observation 3. Regarding observation 5, this is what happens: for r ∈ [1.7, 2.1], the number of classes is equal to three. There are two AN-classes and one A-class. The smallest degree of identification kr id increases with r which causes the drop in the number of identified outliers. For r ∈ [2.2, 4.3], the total number of classes is two. The number of detected outliers depends on classes proportions: there are either two AN-classes (r = 2.2), or one AN-class and one R-class (r = 2.3), or one AN-class and one A-class (r ∈ [2.4, 4.3]). Finally, we observe the same phenomenon, for r ∈ [2.4, 4.3], kr id increases with r which causes the drop of the number of identified outliers until only one class remains for r > 4.3.
The method instability with respect to \( r \) is only observed for a number of classes lower than 4 \((r > 1.6)\). For a number of classes ranging from 24,000 \((r = 10^{-5})\) to 4, the method is stable and exhibits very similar results. Therefore, \( r \) must be chosen based on data range, by keeping a reasonably high number of classes.

XI. CONCLUSION

When dealing with IP traffic, we are faced with IP addresses having very different behaviours. In this context, one question arises: how to differentiate normal behaviours from abnormal ones. In this paper, we proposed a solution to this issue. We introduced a method that detects outliers in IP traffic modelled as a link stream by studying the degree of node over time. We applied our method on three datasets from the MAWI archive: one-hour long IP traffic trace from June 2013, one-day long IP traffic trace from June 2013 and a fifteen-minute long IP traffic trace from November 2018. Likewise, we performed series of experiments by varying the parameters used. In all these situations, we obtained stable results pointing interesting anomalous activities in IP traffic, independently of the degree order of magnitude. Moreover, we surgically removed anomalous traffic, which allowed us to validate our identification, identify more subtle outliers and recover a normal traffic with respect to the degree feature.

This work however is only a first step towards anomaly detection in link streams and may be improved on several aspects. We could extend our method with more complex features than the degree in order to find more complex anomalies and to identify the remaining events unidentified with the degree. This task would be simplified by the fact that largest anomalies have already been removed from the remaining traffic, allowing for a more detailed and finer analysis. At broader scale, our work could be useful in the field of IP traffic modelling as we would be able to generate normal traffic according to a specific feature. Likewise, thanks to their individual extraction, anomalies could also be studied separately for a better characterization.

ACKNOWLEDGEMENT

This work is funded in part by the European Commission H2020 FETPROACT 2016-2017 program under grant 732942 (ODYCCEUS), by the ANR (French National Agency of Research) under grants ANR-15- E38-0001 (AlgoDiv), by the Ile-de-France Region and its program FUI21 under grant 16010629 (iTRAC).

REFERENCES


