Topological Analysis and Visualisation of Network Monitoring Data: Darknet case study

Marc Coudriau, Abdelkader Lahmadi, Jerome Francois

Topological Analysis and Visualisation of Network Monitoring Data: Darknet case study

Marc Coudriau*‡, Abdelkader Lahmadi†, Jérôme François‡
*ENS Ulm - Paris, France
Email: marc.coudriau@ens.fr
†LORIA - University of Lorraine, France
‡Inria Nancy Grand Est, Villèrs-les-Nancy, France
Email: {abdelkader.lahmadi,jerome.francois}@inria.fr

***Abstract*—Network monitoring is a primordial source of data in cyber-security since it may reveal abnormal behaviors of users or applications. Indeed, security analysts and tools like IDS (Intrusion Detection system) or SIEM (security information and event management) rely on them as a single source of information or combined with others. In this paper, we propose a visualisation method derived from the Mapper algorithm that has been developed in the field of Topological Data Analysis (TDA). The developed method and its associated tool are able to analyze a large number of IP packets in order to make malicious activities patterns easily observable by security analysts. We applied our method to darknet data, *i.e.* from an entire and supposed not used subnetwork in Internet and we have found that those observable patterns have been missed by Suricata, a widely used State-of-the-Art IDS.**

I. Introduction

Network monitoring has been extensively used for security, forensics and anomaly detection with a wide area of research and industrial developments [1] whose the main objective is to identify malicious activities based on traffic patterns and to trigger alerts. Those alerts are often processed by security experts who can rely on advanced log correlation engines or SIEM (Security information and event management) for incident respond purposes [2] with manual investigation or confirmation as a second step. It is worth noting that operational security will continue to rely on human experts rather than full automated methods because even few false positives (benign traffic dropped) is not acceptable and so needs a manual examination before final decision. One of the main challenges is the significance of the visualized data and the limited volume of data to deliver to the human analysts.

Instead of only relying on monitoring and analyzing events related to the defended system (*e.g.* an enterprise information system), we are considering some noisy traffic which is not targeted towards real services or hosts and known as Internet Background Radiation (IBR) [3]. It constitutes an important source of information for the prediction and the modeling of Internet malicious activities. IBR data collected from darknets (network telescopes), has been used to study worms, DDoS (Distributed Denial-of-Service) attacks and scanning activities [4]. Extracting and modeling these patterns to be used latter for predicting or better understanding major incidents and events in Internet is a challenging task. The generated and the collected IBR data has a considerable volume with a wide range of services and sources. Making these data useful requires the extraction of their structure and components to be able to use them as a source of information.

Therefore, we built a new method to extract and visualize the malicious components of IBR. The approach relies on Topological Data Analysis (TDA) [5] methods applied to the packet-level data to find persistent structures and patterns. Using this technique is mainly motivated by two reasons. First, its unsupervised pattern detection is appropriate for the analysis and classification of these data since we have no a priori knowledge about them. Second, its visualization capabilities makes the output understandable and interpretable by the human experts. We evaluated our method with a dataset collected from a darknet being a non used but accessible /20 network. Our results shows that our methodology can easily extract scanning activities and DDoS. Even if they seem rather basic malicious activities, a widely-deployed IDS, Suricata[1], was not always able to do so.

The remainder of the paper is organized as follows. In Section II we provide a description of the related work. Our method and implementation details are described in Section III. Experimental results are detailed in Section IV before drawing the final conclusions and future work.

II. Background and related work

Many studies have considered the analysis of IBR, mainly from a statistical perspective to characterise their sources, their originating networks and their services [3], [6]. The obtained results are then used to learn about the malicious activities including worms, viruses, denial of services attacks and their associated hosts and networks [7].

A main technique for passive monitoring of IBR is darknets. A darknet [4] is a whole subnetwork, which is announced over Internet such that packets sent to the IP addresses are properly routed over. However, this subnetwork does not host any services and so no legitimate traffic is supposed to reach it. The entity hosting the darknet is then silently collecting all incoming packets, *i.e.* without replying to any of them. Such an infrastructure is mainly characterized by the size of

[1]https://suricata-ids.org/

the subnetwork defined by the prefix length. Although such an approach seems rather limited than more active techniques like honeypots emulating real services, they have been shown to be complementary [8], [9].

There have been several papers focusing on the characterisation of darknet data [4], [10]. Rather than analyzing full darknet data, examining selected traffic is helpful to address a particular threat, as for example DNS queries to identify DrDOS (Distributed Reflection Denial of Service) [11]. Fachkha et al [4], have provided a study of darknet data and the attack activities that could be investigated throughout its data. They characterised the main activities of a darknet which are scanning and worms propagation. They also summarized the major techniques, including clustering and visualisation, that have been used to analyze darknet data and to identify the profiles of the observed threats.

Visualisation of darknet data has been addressed in several works [12]. InetVis [13] plots darknet data on a 3D scatter plot and highlights visual patterns. Although there is no pre-processing or analysis of the data beforehand, this tools demonstrate the ability of highlighting malicious activity against other IDS like Bro or Snort [14]. In [15], authors have developed a 3D visualisation tool to monitor darknet traffic in real time. The tool is dedicated to the visualisation of raw data at the packet level in a 3 dimensions space to show the alerts with a sphere representing the Internet and multiple rings to represent the monitored darknets. Similarly to [13], their engine is only limited to the visualisation without classification or clustering of darknet monitoring data. Our developed method is not focusing on the visualization of raw data, but instead we are only visualising extracted topological features from darknet data. Existing visualisation techniques of darknet data could be improved by using our method to better visualise persistent patterns of observed attacks.

In this work, we mainly rely on Topological Data Analysis (TDA) which is an emerging field of mathematics that have been applied in several applications to analyse complex high dimensional data, including 3D shape matching, structure of materials and biology [16]. To the best of our knowledge this is the first work using this technique on security monitoring events such as packets received by a darknet. We applied TDA, in particular its mapper method, to extract invariant patterns of multi dimensional monitoring data collected from a darknet and to visualise them in 3D environments. The goal is to help administrators and security analysts to better identify persistent threats.

III. Methodology

A. Objective

We are tackling the problem of extracting activities from noisy monitoring data collected by a darknet. The goal is to identify main perceived activities such as scanning and DDoS attacks while continuously receiving packets, for instance from the /20 subnetwork of the darknet used in our experiments. The available data is collected in May 2016 with a rate of 3 millions packets per day and we considered all the received packets without filtering and no a priori knowledge.

A technique based on grouping used by an IDS, such as Suricata, is able to identify some patterns sharing one or multiple attributes. However, it is not robust to noise and it is unable to remove that noise to clearly identify all the attack patterns. We can also rely on clustering and ranking to detect the patterns by training a model and defining a set of thresholds to select candidates. However, such methods require to repeatedly train the models when facing new data.

DDoS or scanning activities are usually observed as lines on scatter plots where axis represents the ports or IP addresses like in [13]. For example, in Figure 1, vertical lines can be assumed as horizontal scans where, for each of them, a single IP address try to reach all IP addresses of our darknet. Whereas it is a simple example for illustration purposes, real data is plotted in Figure 3(a) where both source and destination IP addresses and ports are plotted. In that case, human eyes cannot easily distinguished scanned activities whereas they become finally visible after our analysis in Figure 4 where, for example, a single source doing horizontal is represented as a triangle (or slice).

B. Method overview

Detecting activities from darknet events requires the updating of the detection mechanism output according to the new arrived events. The events of a darknet are mainly received packets that we need to classify to detect relevant activities. We propose an unsupervised method for detecting activities from monitoring data, in particular collected from a darknet by using Topological Data Analysis (TDA). This approach consists in generating representations from complex high dimensional data and extract their invariant topological features to discover relationships and patterns in data. The TDA technique has the advantage to be coordinate and deformation invariant, and it is able to deal with compressed representation. It does not depend on a coordinate system so it is able to analyse monitoring data collected from different format. It is also less sensitive to noise and it is able to handle approximate data. These features guide our choice in using this approach since a darknet collect all data without distinction a priori, *e.g.* multiple attacks or packets due to misconfigurations are mixed together, can only observe a partially an activity wider than the size of the considered subnetwork, *e.g.* a scan at the Internet scale.

In this work, we have mainly applied the Mapper algorithm [17] from TDA coupled with DBSCAN clustering algorithm [18] to reveal patterns of activities observed from a darknet. The Mapper technique is combining data projection over overlapping hyper cubes and partial clustering within each of them. It is more robust to noise and it is able to extract in better way patterns from darknet data than using grouping by attributes. The output of the algorithm is a graph where each node represents a cluster of packets. Edges are created between nodes if they share a common packet. The Mapper algorithm mainly requires two parameters to control the resolution of

the obtained graph which are the number of intervals over the range of the filter function values and the percentage of overlap between successive intervals.

C. Detailed technique

As a pre-processing step, we extract six features from each packet: the timestamp, the source and destination IP addresses and ports, and the protocol (TCP, UDP or ICMP). Each packet is then mapped to a vector in $\mathbb{R}^6$ by converting the selected 6 features to numerical format. IP addresses are represented by their respective integer values between 0 and 2^{32}. We used the integer values of the ports and we encoded the name of the protocols to numerical values. We have to note that in this work, we have chosen from a darknet packet 6 features, however the TDA technique itself could be applied on high dimensional data since it takes as input vectors of size n without making a restriction on this size.

Then, the unsupervised mapper algorithm of TDA as detailed in [17] is applied on the obtained vectors to transform the packets feature space into a topological space. The mapper algorithm extracts geometric features of the data to represent them as simplicial complexes or data shapes of relevant activities in a darknet. It proceeds in the following steps.

First, a filter function is applied to provide a value to each data point. In the general use of TDA, interesting dimensions are selected at this step, using knowledge about the shape of the data. In our case, a priori knowledge is weak and this selection has been made at the time of the vectorisation. The filter is therefore $\mathbb{R}^6$ identity most of the time. The only exception is the use of the projection $\mathbb{R}^6 \mapsto \mathbb{R}^5$ which removes the time component in order to associate packets that repeat in time (see DDoS section IV-C). The dataset is then divided into smaller subsets by dividing the filter's range into a set of smaller overlapping intervals. Hence the original 6-dimensional hypercube containing all data is divided into multiple and overlapping 6-dimensional hypercubes. This step relies on two parameters: the first parameter is the resolution which represents the length of the interval over the filter function range and the second parameter is the overlap parameter which represents the percentage of overlap between successive intervals.

The second step consists into applying the clustering algorithm, DBSCAN, within each individual hypercube. It takes two parameters, ϵ and $minpts$ along with the data. It will mark as being in the same cluster, the points having at least $minpts$ at a distance lower than ϵ (neighbours) and will propagate the search to those neighbours. A point is considered as noise if it has less than $minpts$ neighbours. To compute the similarity between the packets when using the DBSCAN clustering algorithm, we have used the following metrics. For the timestamp attribute and IP destination addresses, the difference between their respective values is suitable metric since the former is a measurable quantity and the latter are within the same /20 subnetwork. For the source IP addresses, we also used the distance between their respective values as a metric, that could be problematic if the IP addresses are not in the same subnetwork. For instance, with this metric the address 2.2.2.2 will be more close to 3.3.3.3 than 11.11.11.11. For the protocol names and ports, we have used the equality metric between their respective values to output 0 or 1 since a distance between their values does not make sense. The proposed metrics to measure the similarity between packets remains simplistic since they could cluster packets from different subnetworks with a close difference, but our results show that they remain good heuristics for clustering scanning and DDoS activities. We have made this choice since to the best of our knowledge, there are no well known distance metrics for measuring the similarities for network packet attributes, mainly regarding IP addresses, protocol names and ports.

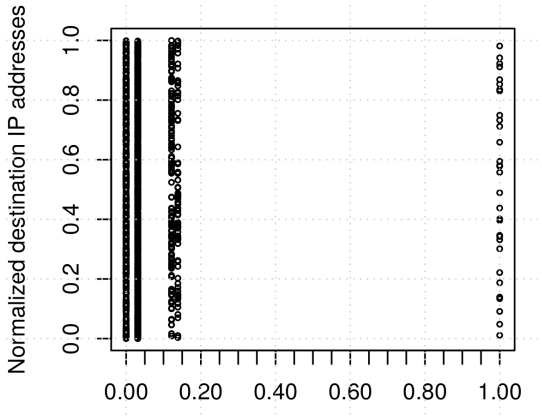


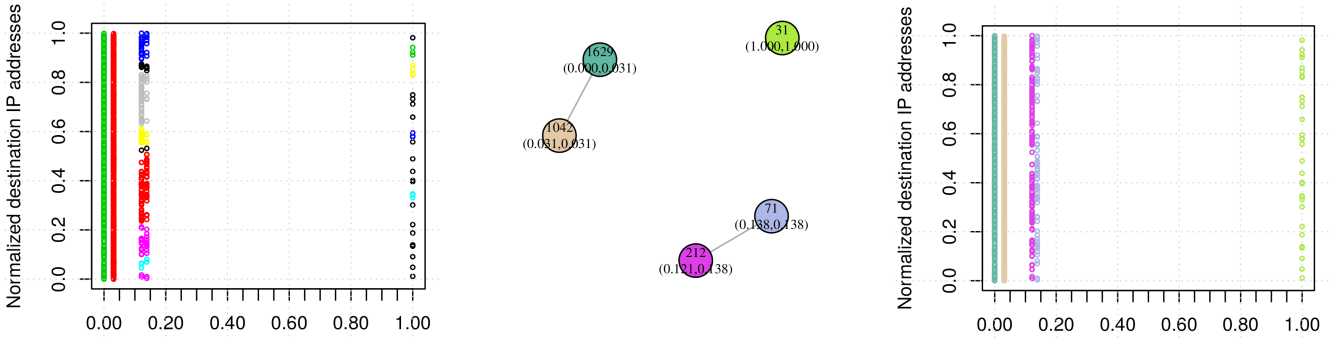
Fig. 1. Example of a 2 dimensions dataset of a scanning activity observed by a darknet (original data).

The final output of the algorithm is a topological graph where each node is a cluster of points within one hypercube. If two clusters (from different hypercubes) have one or more shared points with respect to the specified overlapping interval, the nodes are linked together.

D. Example of a scanning activity

To illustrate the mapper algorithm, we applied the TDAmapper R package [19] on a single dataset of a scanning activity manually extracted from the darknet. From this dataset we considered only 2 dimensions which are the source and the destination IP addresses of each received scanning packet. We have firstly projected the dataset in its two dimensions space as depicted in Figure 1. We mainly observe 5 vertical lines at the different scanning sources over the destination IPs of the darknet. Addresses have been normalised using a min-max scaler. Then, for demonstration purposes, we applied on the dataset the DBSCAN clustering algorithm with $\epsilon = 0.2$ and $minpts = 3$. The algorithm finds 14 clusters that are represented with different colours on the original data as depicted in Figure 2(a). DBSCAN is able to identify the first two vertical lines as distinct clusters since they are dense but not the other vertical lines identified as multiple clusters.

Then, we applied TDA on the same dataset with a 30 intervals resolution and 50% of overlap (which implies many DBSCAN executions). The result is depicted in Figure 2(b)



(a) The dataset clustered using the DBSCAN algorithm with $\epsilon = 0.2$ and $minpts = 3$.

(b) The mapper graph of the dataset using 30 intervals with 50% overlap.

(c) The dataset clustered with the nodes of the mapper graph.

Fig. 2. Analysis applied to the two-dimensional example of Figure 1

where 5 nodes appear, divided in a single node and 2 groups of 2 connected nodes. Each node is annotated with the number of contained points and the minimum and maximum value of its associated interval. In Figure 2(c), we have coloured the original data with the colours of the nodes of the mapper graph. We observe that each group of 2 connected nodes represents two close vertical scanning lines and the isolated node of the graph represents the last vertical scanning line at the right side of the figure. In this example, we have shown that the mapper algorithm with partial DBSCAN clustering is clearly able to identify the patterns of a scanning activity. Then, our goal is to apply the mapper technique to process a noisy dataset collected from a darknet as depicted in Figure 3(a) to extract such patterns.

E. Implementation details

We have implemented our own tool of the mapper algorithm to extract topological graphs by processing darknet packets and also their visualisation with 3D coordinate system with coloured links between two planes. As already introduced, one plane represents the source IP address and port of the packet and the second plane represents the destination IP address and the port of the packets. This 3D visualisation is similar to the technique proposed by Nunnally et al [20]. Before any processing, the result is similar to Figure 3(a).

Our mapper tool relies on an extended code from KeplerMapper (github.com/MLWave/kepler-mapper), ported to Python3 and not requiring Sklearn anymore. The simplicial complex created by the Mapper is rendered as a 2D graph using physical-like forces (Figure 3(b)). The size of each point is proportional to the logarithm of the number of packets it contains. Furthermore, our approach relies on DBSCAN as described before. Due to efficiency reason, it has been implemented in C. To quickly find the ϵ neighbors, various methods exist but we preferred to use a naive search since the number of dimensions is rather high and a pre-calculation using space partitioning is therefore more costly than considering each packet on typical inputs ($n = 1000$). On large inputs ($n = 100000$) it has nearly the same performance than Scikit-Learn implementation which was first used, whereas it is more efficient on small ones.

The output of the mapper is a simplicial complex, represented by a flat graph as shown in Figure 3b. The colors arbitrarily associated to each connected component can then be used to color every packet of each component on the previous three-dimensional visualisation, giving representations as shown in Figure 3(c).

IV. Experimental Results

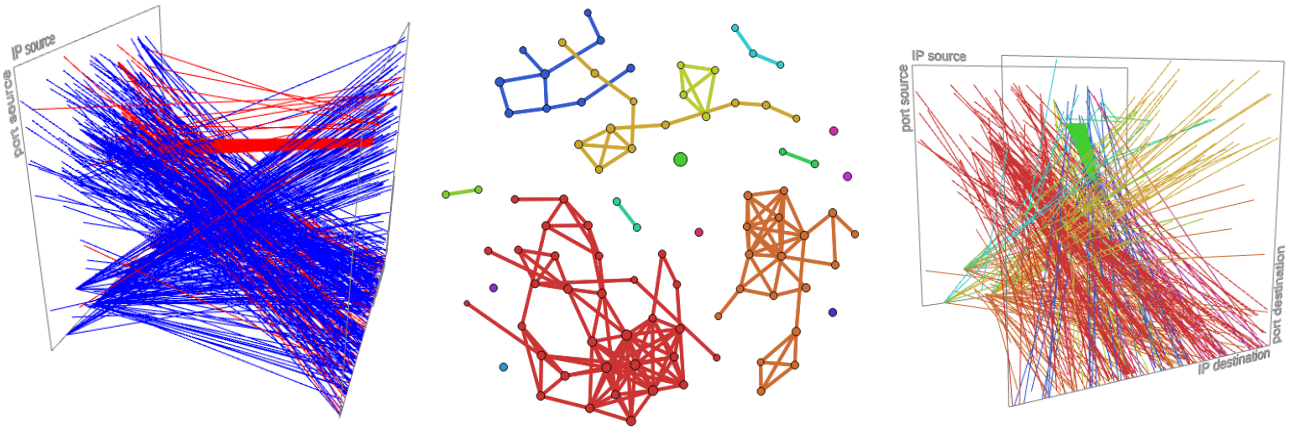
A. Separating patterns

A first run on the data depicted in Figure 3(a) leads to the graph obtained in Figure 3(b). After manual investigation, we found that the large green dot is a scanning activity trying to exploit a known router vulnerability on port 53413. The red component denotes a set of packets looking for Telnet or SSH accesses. The orange one is a sparse scan on those ports but originated by a single address. The yellow component gathered two randomized scans and some noise. They may be discriminated by tuning the algorithm parameters. Using the topological analysis technique, we are able to separate efficiently different malicious activity patterns while they are originally highly mixed as shown in Figure 3(c), where each cluster color is reported on its individual packet.

B. Topologies of scanning activities

The spatial continuity of sweeping scans makes them really easy to identify out of noise, even with overlapping schemes. A small ϵ coupled with a important $minpts$ value will drive the clustering to produce accurate slices of scan while a small overlap value allows the mapper to group those slices coherently. With our implementation applied on the 6 normalized dimensions, $\epsilon = 0.05$, minpts $= 20$, overlap $= 5\%$, we obtained good results. We have to note that the parameter values are set manually using a trial and error method until a pattern is detected and then they remain valid for similar patterns.

Figure 4 is the result of this scan analysis with more data than Figure 3(a). Since the packets marked as noise by the



(a) Original data: The source plane is on the left and the darknet on the right. Packets shown in blue are of TCP type and the red ones are UDP.

(b) Clustered Mapper graph

(c) Original data coloured by cluster

Fig. 3. Topological analysis of 1000 packets captured on the 13 April 2016 with $\epsilon = 0.5$, minpts $= 3$, overlap $= 10\%$.

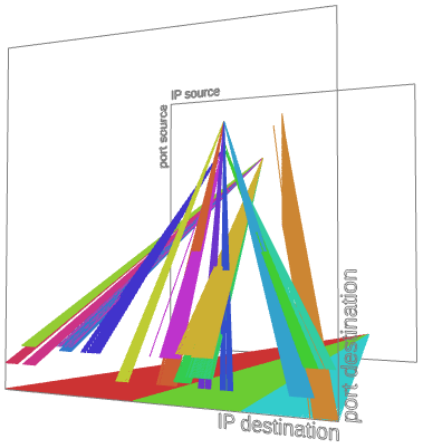


Fig. 4. Scan detection using the mapper on 8000 consecutive packets. Scans are distinct and countable.

clustering are not drawn, the result is much more understandable than a raw visualisation as proposed by Nunnally and al [20] and depicted in Figure 3(a). Slices of scans are clearly visible where multiple IP addresses of our darknet are being targeted. The only variable that have been chosen to obtain a well-looking render is the starting point of the packet fetch in order to have a range of scanning behaviours. As continuous scans are really similar, those parameters are suitable for the whole data we have.

We have also applied the official set of rules of Suricata version 3.0 on the same dataset to detect inside them scanning activities. We found that Suricata was able to detect only four of them because its rules mainly rely on grouping packets within a time window to detect such activities.

C. Topologies of DDoS activities

Running the mapper on another day, we found a large cluster on the graph from 10 am to the end of the day. On the 3D view however, it appears a discrete line. The explanation is that the same packet has been sent at a high rate. This UDP packet with 53 as source port is a DNS response to a supposed request (*i.e.* one of our darknet IP address has been spoofed). This is available for the first 1203 packets while the rest of the packets (318710) are DNS failure messages to the requested domain. With more than 310 000 packets sent over 16h (a darknet IP usually receives about 3000 packets daily), no deep analysis is necessary (statistical approach would have been suitable). However, the interesting fact is that topological analysis would have worked as well – *i.e.* present a dot in the graph – on a much shorter time, where statistical tools may not be triggered.

For DDoS, the spatial connection between packets is reduced and therefore they are more difficult to identify. The visualisation shown in Figure 5 is isolating one DDoS by considering only UDP packets and only clusters with at least 10 elements. Such filtering is legitimate as a DDoS is usually targeted towards a unique protocol and should involve many packets, and so bigger clusters. As the available data is less than in the previous cases, parameters should have been exaggerated ($\epsilon = 0.03$, minpts $= 100$, overlap $= 0.01$) to regroup them in a single component.

We have to note that despite the apparent sparseness due to the large number of sources, each source is sending maximum number of packets to the destination. It creates several clusters that are even more dense than scans. Parameters sufficient to eliminate scan will keep them. After the clustering, the Mapper algorithm is gathering them.

D. Performance analysis

The machine on which tests have been run is equipped with a Quad Core CPU cadenced at 2.83GHz, a 15GB RAM and works with Linux Mint. A not parallelised analysis including clustering and mapping of 1024 packets takes between $0.4s$ and $0.9s$, depending on the packets themselves. Using the four cores and a 20% overlap of datasets, our approach provides

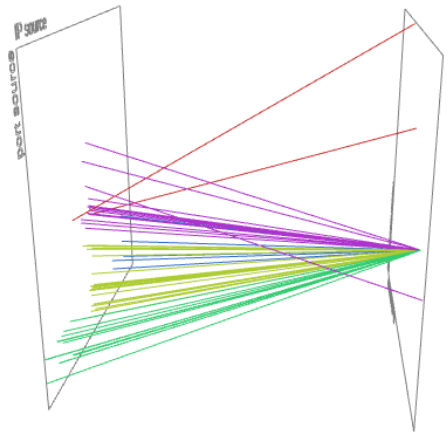


Fig. 5. Visualisation of a DDoS using Mapper. Many different sources are suddenly sending UDP packets to a single port on a single address. The simple lines that still appear are in fact constituted of a heavily repeated packet that could be either a non-distributed DoS or a misconfigured device.

acceptable performance since an approximate number of 3 millions packets per day are analyzed in 11 minutes. The clustering is indeed quadratic in the worst case but it is applied on hypercubes that represent only a fraction of the data. Besides, the apparent complexity is below quadratic in all our practical experiments. More computing power could be used to increase the number of hypercubes. Three divisions over the six dimensions, resulting in 729 hypercubes have been used in this work. However, finer analysis would be obtained with a more important number of hybercubes, coupled with a algorithm deciding which dimensions are the more useful to slice precisely.

For 32768 packets simultaneously, which already represent more than a quarter hour and is sufficient with most requests (for example what has this known attacker sent today?), two minutes are required. Hence, a human analyst can use our tool in an near real-time analysis by clustering the data within a time or space window (number of packets) and updating the visualization. Making the tool operating in real-time requires mainly more computing resources and the parallelization of the tool.

V. Conclusion

In this paper, we proposed a methodology relying on topological analysis applied on network packets. It is able to cope with large volume of traffic in order to extract groups of packets belonging to the same activity, especially malicious ones, even if they are all heavily mixed. Our experimental validation demonstrates this ability on real data collected from a darknet and shows that our method discovered efficiently more scans than the well-used IDS Suricata. Our developed tool is currently for internal use, however it could be provided upon request. In future work, the approach and its associated tool will be extended to consider other types of abnormal traffic as well as integrating new dimensions from packet headers and contents.

Acknowledgment

The authors thank Frederic Beck for his help in the collection of the data used in this paper. This work was partially funded by HuMa, a project funded by Bpifrance and Region Lorraine under the FUI 19 framework. It is also supported by the High Security Lab hosted at Inria Nancy Grand Est.

References

[1] M. H. Bhuyan, D. K. Bhattacharyya, and J. K. Kalita, "Network anomaly detection: Methods, systems and tools," *IEEE Communications Surveys Tutorials*, vol. 16, no. 1, pp. 303–336, First 2014.

[2] S. Bhatt, P. K. Manadhata, and L. Zomlot, "The operational role of security information and event management systems," *IEEE Security Privacy*, vol. 12, no. 5, pp. 35–41, Sept 2014.

[3] R. Pang, V. Yegneswaran, P. Barford, V. Paxson, and L. Peterson, "Characteristics of internet background radiation," in *Proceedings of the 4th ACM SIGCOMM Conference on Internet Measurement (IMC)*, 2004, pp. 27–40.

[4] C. Fachkha and M. Debbabi, "Darknet as a source of cyber intelligence: Survey, taxonomy, and characterization," *IEEE Communications Surveys Tutorials*, vol. 18, no. 2, pp. 1197–1227, Secondquarter 2016.

[5] G. Carlsson, "Topology and data," *Bulletin of the American Mathematical Society*, vol. 46, no. 2, pp. 255–308, 2009.

[6] E. Wustrow, M. Karir, M. Bailey, F. Jahanian, and G. Huston, "Internet background radiation revisited," in *Proceedings of the 10th ACM SIGCOMM Conference on Internet Measurement (IMC)*, 2010, pp. 62–74.

[7] D. Moore, C. Shannon, D. J. Brown, G. M. Voelker, and S. Savage, "Inferring internet denial-of-service activity," *ACM Trans. Comput. Syst.*, vol. 24, no. 2, May 2006.

[8] J. François, R. State, and O. Festor, "Activity Monitoring for large honeynets and network telescopes," *International Journal On Advances in Systems and Measurements*, pp. 1–13, 2008.

[9] P. Chatziadam, I. G. Askoxylakis, and A. Fragkiadakis, *A Network Telescope for Early Warning Intrusion Detection*. Springer, 2014.

[10] E. Balkanli, J. Alves, and A. N. Zincir-Heywood, "Supervised learning to detect ddos attacks," in *Computational Intelligence in Cyber Security (CICS), 2014 IEEE Symposium on*, Dec 2014, pp. 1–8.

[11] C. Fachkha, E. Bou-Harb, and M. Debbabi, "Inferring distributed reflection denial of service attacks from darknet," *Computer Communications*, vol. 62, pp. 59 – 71, 2015.

[12] V. T. Guimarães, C. M. D. S. Freitas, R. Sadre, L. M. R. Tarouco, and L. Z. Granville, "A survey on information visualization for network and service management," *IEEE Communications Surveys Tutorials*, vol. 18, no. 1, pp. 285–323, Firstquarter 2016.

[13] J.-P. van Riel and B. Irwin, "Inetvis, a visual tool for network telescope traffic analysis," in *Proceedings of the 4th International Conference on Computer Graphics*. ACM, 2006.

[14] B. Irwin and J. P. van Riel, "Using inetvis to evaluate snort and bro scan detection on a network telescope," in *Proceedings of the Workshop on Visualization for Computer Security*. Springer, 2008.

[15] D. Inoue, M. Eto, K. Suzuki, M. Suzuki, and K. Nakao, "Daedalus-viz: Novel real-time 3d visualization for darknet monitoring-based alert system," in *Proceedings of the Ninth International Symposium on Visualization for Cyber Security*, ser. VizSec '12, 2012, pp. 72–79.

[16] M. Offroy and L. Duponchel, "Topological data analysis: A promising big data exploration tool in biology, analytical chemistry and physical chemistry," *Analytica Chimica Acta*, vol. 910, pp. 1 – 11, 2016.

[17] G. Singh, F. Memoli, and G. Carlsson, "Topological Methods for the Analysis of High Dimensional Data Sets and 3D Object Recognition," in *Eurographics Symposium on Point-Based Graphics*. The Eurographics Association, 2007.

[18] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise." AAAI Press, 1996, pp. 226–231.

[19] P. Pearson, D. Muellner, and G. Singh, *TDAmapper: Analyze High-Dimensional Data Using Discrete Morse Theory*, 2015, r package version 1.0. [Online]. Available: https://github.com/paultpearson/TDAmapper/

[20] T. Nunnally, P. Chi, K. Abdullah, A. S. Uluagac, J. A. Copeland, and R. Beyah, "P3d: A parallel 3d coordinate visualization for advanced network scans," in *2013 IEEE International Conference on Communications (ICC)*, June 2013, pp. 2052–2057.