Sensing the Noise: Uncovering Communities in Darknet Traffic

Sensing the Noise: Uncovering Communities in Darknet Traffic

Francesca Soro,
Mauro Allegretta, Marco Mellia
Politecnico di Torino
Torino, Italy
first.last@polito.it

Idilio Drago
Università degli Studi di Torino
Torino, Italy
idilio.drago@unito.it

Leandro M. Bertholdo
University of Twente
Enschede, The Netherlands
l.m.bertholdo@utwente.nl

***Abstract*—Darknets are ranges of IP addresses advertised without answering any traffic. Darknets help to uncover interesting network events, such as misconfigurations and network scans. Interpreting darknet traffic helps against cyber-attacks – e.g., malware often reaches darknets when scanning the Internet for vulnerable devices. The traffic reaching darknets is however voluminous and noisy, which calls for efficient ways to represent the data and highlight possibly important events. This paper evaluates a methodology to summarize packets reaching darknets. We represent the darknet activity as a graph, which captures remote hosts contacting the darknet nodes ports, as well as the frequency at which each port is reached. From these representations, we apply community detection algorithms in the search for patterns that could represent coordinated activity. By highlighting such activities we are able to group together, for example, groups of IP addresses that predominantly engage in contacting specific targets, or, vice versa, to identify targets which are frequently contacted together, for exploiting the vulnerabilities of a given service. The network analyst can recognize from the community detection results, for example, that a group of hosts has been infected by a botnet and it is currently scanning the network in search of vulnerable services (e.g., SSH and Telnet among the most commonly targeted). Such piece of information is impossible to obtain when analyzing the behavior of single sources, or packets one by one. All in all, our work is a first step towards a comprehensive aggregation methodology to automate the analysis of darknet traffic, a fundamental aspect for the recognition of coordinated and anomalous events.**

***Index Terms*—Security, graph analysis, community detection.**

I. Introduction

Darknets, also referred to as network telescopes, Internet sinks or Internet background radiation, are sets of regularly advertised IP addresses that do not host any device or service. Darknets are deployed with the purpose of collecting unsolicited packets reaching the unused ranges of addresses. Darknets have been proven central for the detection of events such as the spreading of new malware, network scans and misconfigurations [1], [2]. Even without hosting any production server, darknets receive a significant amount of traffic. In fact, different IPv4 darknets have been shown to constantly receive unsolicited traffic from thousands of sources [3]. This baseline noise is increasing throughout the years, and it is mixed with sudden peaks of unsolicited traffic that are seen whenever new malware activity or new large-scale scans start. Moreover, a significant percentage of traffic reaching darknets consists of backscattering, i.e., answers sent to the darknet by services receiving spoofed IP packets possibly during DDoS attacks.

Understanding the underlying phenomena behind darknet traffic is challenging given the large amount of packets and the heterogeneity of events generating them. The analysis of darknet traffic requires methods to separate occasional (and likely unimportant) events from large-scale coordinated actions, e.g., due to botnets. Several works propose methods to categorize darknet traffic [4], [5], [6], [7], [8]. These works usually rely on domain knowledge to create static categories of traffic (e.g., misconfiguration, backscattering etc). As such, they may miss events that diverge from expected signatures. More important, there is a lack for automatic ways to uncover correlations among different events, which could help reducing the manual work when interpreting darknet traffic.

When processing darknet traffic, we are interested in grouping packets from different sources that may suggest the presence of coordinated activities, due for example to the propagation of a botnet infection over different hosts. Such coordinated actions may result in temporal or spatial correlations, e.g., apparently unrelated sources cooperating to scan for a particular service, and they are invisible if hosts or packets are analyzed individually. Data should therefore be represented in an informative way, so to ease the inspection and uncovering of hidden patterns.

This paper investigates whether graph mining techniques can help to uncover such macroscopic coordinated events in darknet traffic. We represent the darknet traffic as a bipartite graph linking traffic sources to the contacted destination ports. We then run community detection algorithms over such graphs, in the search for devices performing similar activity against the darknets in the same time interval.

We evaluate the methodology using only TCP traffic, collected in two distinct darknets. We tune and test the methodology on three weeks of data captured from a darknet in Italy, composed by 3 $/24$ IPv4 networks. We then apply the methodology to one day of traffic captured in a $/19$ darknet deployed in Brazil. In both cases we found communities performing very homogeneous activity. We discuss the composition of the most relevant communities, their characteristics and peculiarities, showing some relevant behaviors that our algorithm is able to automatically detect. In particular, we find (i) communities composed by thousands of sources that focus on popular services; (ii) communities that focus on horizontal scans for vulnerable services.

Our work is a first step towards a methodology to automate

the analysis of darknet traffic. Our results show that darknets can be automatically characterized by using graph mining techniques to highlight interesting patterns over space and time. In the following, Section II summarizes related work. Section III describes the applied graph mining methodology. Section IV describes our datasets and provides a basic characterization of the data. Section V describes the output of the graph mining and community detection analysis. Section VI concludes the paper.

II. Related work

Darknets are extensively used for supporting cybersecurity. They have been applied, for example, in the investigation of DDoS phenomena [9], [5], [7], for the estimation of the IPv4 address space utilization [6] and for the analysis of Internet censorship [8]. Some papers characterize darknets in terms of deployment (i.e., centralized darknets vs sparse *greynets*), size and geolocation [10], [11], [12]. Instead of focusing on particular applications or darknet characteristics, we here introduce a generic methodology to ease the analysis of large volumes of darknet traffic.

In [13], authors suggest to represent network traffic as a bipartite graph linking IP sources to $/24$ destination networks. Other works suggest to represent generic network data by means of first order [4], [14] or second order [15] Markov Chains. We here follow a similar approach to represent the darknet traffic. We however go one step further and apply community detection to uncover coordination in patterns.

The detection of coordinated activity in network traffic has been targeted with other techniques. Authors of [16] propose a tool to easy the visualization of scanning activities, while [17] focus on topological analysis. Community detection is often used to evaluate social networks, but some works apply the algorithms to computer networks too. Authors of [18] separate legitimate and unsolicited email traffic. Authors of [19] identify patterns used by attackers contacting honeypots. We here verify whether community detection is suited when evaluating darknet traffic.

Authors of [20], [21] address a problem similar to ours. They however focus on Internet scans, characterized with an event-based graph defined as the sequences of ports contacted by scanners. We broaden the analysis to any activity on darknets, searching relationships on traffic reaching the darknet.

III. Methodology

We now describe out graph definition (Sect. III-A) and the used community detection algorithms (Sect. III-B).

A. Graph definition

We define our graph $G(V, E)$ with V being its set of nodes and E its set of edges. We want to represent the activity of remote sources sending traffic to darknets. After we take into account different aspects, we pick a definition for $G(V, E)$ that provides us not only good semantics, but also a manageable graph.

$G(V, E)$ is a weighted bipartite graph. Nodes in V represent, on the one hand, the sources sending traffic to darknets and, on the other hand, the destination *ports* contacted by the sources. We call S the set of nodes in V representing the traffic sources and P the set of nodes representing the contacted ports. There is a edge e between a node in S and a node in P if the source has sent packets to the given port. The weight w_e of e is the number of packets observed for the pair of nodes in the given time interval.

We focus on port numbers as they give us an indication of the services searched by sources. The decision on how to represent the sources is however harder. Creating a node for each single sender IP address may result in very large graphs. Grouping addresses according to networks is instead a more prominent alternative. To avoid setting a completely arbitrary aggregation, here we map the IP addresses to their respective Autonomous Systems (ASes).[1] As such, we report coordination among IP addresses that belong to different ASes.

B. Detecting communities

Community detection aims at finding subgroups of nodes that are densely connected – i.e., forming communities. In our specific case, communities would represent sets of ASes that contact similar sets of destination ports in a given time interval. For instance, a group of remote sources that behave similarly due to a botnet infection, or nodes under the control of the same attacker in distinct source ASes, that aim at finding vulnerabilities on similar targets.

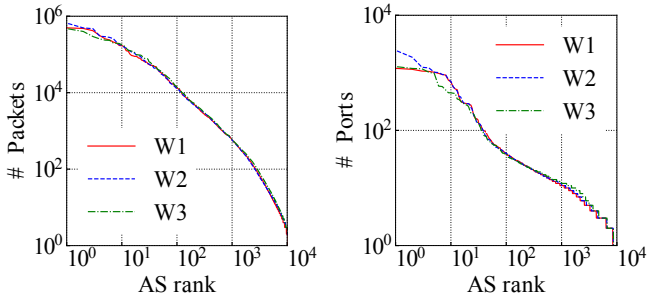
We here consider the Greedy Modularity Algorithm (GMA) [22].[2] This technique measures how strongly a graph can be separated into modules – i.e., groups of nodes that are strongly connected inside the group, while loosely connected with nodes belonging to other groups. The idea behind the algorithm is that a random graph is not expected to show cluster structures with condensed nodes and edges, while nodes having some sort correlation will form a modular structure. To recognize modular structures, the GMA exploits the concept of modularity [24], defined as:

$$Q = \frac{1}{2m} \sum_{i,j} [w_{ij} - \frac{k_i k_j}{2m}] \delta(c_i, c_j) \quad (1)$$

where w_{ij} represents the weight of the edge between node i and node j, k_i is the sum of the weights of the edges on i, c_i is the community to which i is assigned, $\delta(u, v)$ is 1 if $u = v$ and 0 otherwise, and $m = \frac{1}{2} \sum_{i,j} w_{ij}$. The algorithm starts by initializing a community per node; at the first iteration it hence yields $|V|$ communities. The second iteration proceeds by calculating the modularity between each nodes and its neighbours, performing the merge between the pair of nodes that have the highest modularity gain. At each iteration the algorithm merges together the neighbouring communities yielding positive gains in modularity. It stops when it is

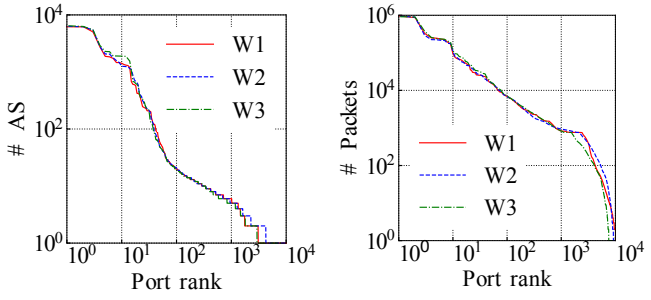
[1] https://pypi.org/project/pyasn/

[2] We have also considered the Label Propagation Algorithm [23], which however produces less meaningful results, e.g., putting almost all nodes in a single community.



(a) Packets per AS (b) Contacted ports per AS

Fig. 1. Per-AS breakdown in the Italian darknet. Notice the log-log scales.



(a) Source ASes per Port (b) Packets per Port

Fig. 2. Per-port breakdown in the Italian darknet. Notice the log-log scales.

unable to merge any new elements to any community. As $G(V, E)$ is a bipartite graph, our communities will contain both nodes belonging to S (sources, i.e., ASes) and P (ports). Every packet in our dataset, which originate the edges in $G(V, E)$, is labeled with two communities: The community of its source AS and the community of its destination port. When quantifying activity of communities in terms of packets, we therefore split results into *AS communities* and *port communities*. In other words, an AS s_i will belong to a single AS community c_j (also containing other ASes that have performed activity similar to s_i), but its activity will be reflected into more than one port community, implying that not all ports contacted by the s_i are also part of c_j. The same concept applies when tackling the problem from the point of view of the destination ports.

IV. Datasets

A. Darknets

We rely on packets captured from two darknets, in Italy and Brazil, respectively. The darknet in Italy is hosted at the Politecnico di Torino campus network and is composed by three $/24$ IPv4 networks. The darknet in Brazil is hosted by a research network operator and is composed by a $/19$ IPv4 network. We use three weeks of packet traces collected in January 2020. More details on the traffic composition can be found in [10]. We focus on TCP traffic, which sums up to 87% of the packets in the traces. We tune and evaluate our methodology using the Italian traces. Afterwards we use the Brazilian traces to confirm our findings. Since both darknets are physically located in different continents and logically located in far away IPv4 ranges,[3] we can verify whether the main events observed in a darknet are also observable in another network.

B. Popularity of ASes

We start by reporting some basic characteristics of the darknet traffic. Figure 1 breaks down the data according to source ASes. We show only results for the Italian darknet. Figures for the Brazilian darknet are qualitatively similar, even if it receives much more packets, since it aggregates more addresses than the Italian one. Different lines represent each of the three weeks in the data, we analyse every week independently from the others.

Figure 1(a) shows the number of packets received from each AS. ASes are ranked in the x-axis according to the number of packets they send. Results are almost coincident for the three weeks. Most ASes are populated by sources which send only few packets to the darknet in a week (notice the log-log scale). Yet, a small group hosts some *heavy hitters* which send thousands of packets. Notice how the top $1\,000$ ASes are origin of more than $1\,000$ packets each, some of them of hundreds of thousands of packets (leftmost part of the plot).

More interesting, Figure 1(b) shows the ranking of ASes according to the number of destination ports their hosts contact. We see that only a small number of ASes targets 100 ports or more (leftmost points). Hosts targeting lots of destination ports are likely performing Internet scans. The remaining ASes target a small number of destination ports, suggesting either targeted activity or random behaviour – e.g., hosts in different ASes collaborating for distributed port scans.

C. Popularity of ports

Figure 2 depicts a breakdown according to destination ports. As for the previous section, we show only results for the Italian darknet. Figure 2(a) depicts the number of unique ASes that contact each port. The leftmost part of the plot shows that only a minor number of ports is contacted by a large number of ASes. Only the most contacted ports receive packets from 30 or more ASes.

Figure 2(b) depicts the number of packets received by each port. Again only a small subset of ports see a significant number of packets. We see that most of the $65\,536$ available TCP ports do not receive any packets in a week, whereas most of the targeted ports receive less than 100 packets (notice the log scales).

The ports receiving lots of packet are the ones corresponding to services often targeted by remote attacks. In Figure 3 we highlight the most contacted ports in the Italian darknet, along with the percentage of packets received by each of them. We see that ports hosting terminal services (e.g., port 23, 22 and 3389), web servers (e.g., ports 80 and 8080) and databases (e.g., port 1433) dominate the list, attracting large percentage of traffic.

[3]Privacy requirements imposed by the network operators prevent us from disclosing the IPv4 prefixes hosting the darknets.

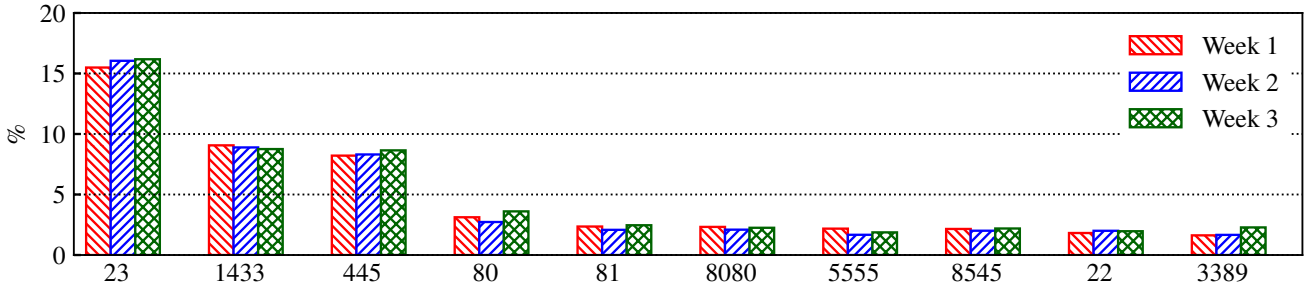


Fig. 3. Percentage of packets directed to the top-10 destination ports.

All in all, we see heavy-tails in the popularity of both ASes and ports. For building the graphs in the coming sections, we filter out traffic going to *unpopular ports*. This step is relevant, since some graph mining algorithms do not scale well for large graphs. We set the threshold at 100 packets based on Figure 2(b). We argue that most packets going to those ports are due to misconfigurations and backscattering [10]. The latter, in particular, are packets sent by *victims* of spoofed IP attacks, which therefore are less likely to show interesting coordination. This threshold allows us to filter out (on average) 93% of the ports, while still retaining 98.7% of the observed packets in the traces. Given that our objective is to identify the spread of wider phenomena and possible coordinated actions among ASes, we choose instead not to set a threshold to filter out the sources.

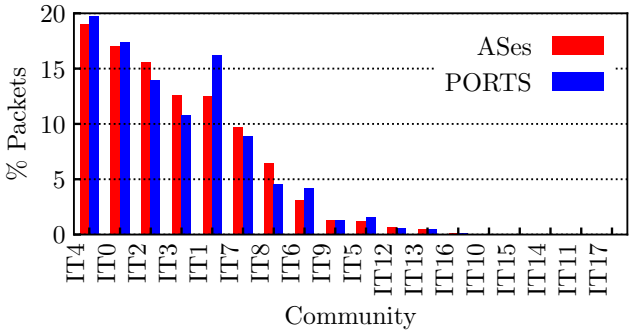
V. Darknet communities

In this section we summarize the communities found in both darknets. The resulting graphs are composed by an average of around 16 000 nodes and 67 000 edges per week (Italy), and around 32 000 nodes and 142 000 edges (Brazil).

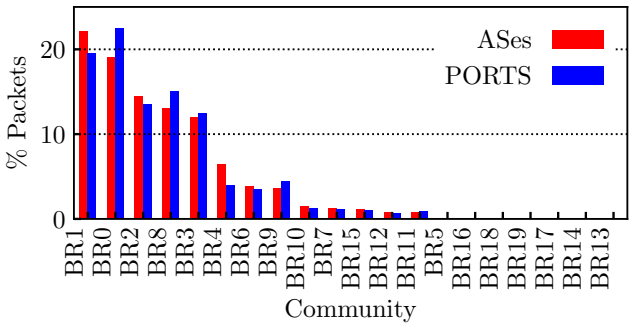
A. Community popularity

Figure 4 provides a high-level view of the communities found for the first analyzed week (day in the BR case). The community detection algorithm has found 18 (20) communities of variable sizes for Italy (Brazil). The figure shows the percentage of packets per community. Recall that each packet may be associated with 2 communities, one for its AS and one for its destination port. The total number of packets per community is however very similar regardless on whether ports or ASes are considered for this association.

The top three communities are involved in more than half of the total weekly traffic in both darknets. Intuitively, these communities may include sources searching for common Internet services, e.g., scans for the popular ports seen in Figure 3. We will investigate this hypothesis later. Observe that only the top 12 (13) communities in Italy (Brazil) have a significant number of packets. Communities with negligible number of packets are mostly formed by a few ASes that target a single port. We ignore these negligible communities in the analysis that follows. Tables I and II report a more detailed breakdown of each community. The communities are sorted as in Figure 3, according to the number of packets they include. As we can see, most of the communities target all the available addresses in the darknets. Some of the largest communities prove to collect either more sources (as in IT1 or BR0), or more destinations (as in IT3 and BR3). Only by having a look at these data we are able to identify more specific and less evident events as possible net scans on a single port (e.g., as in IT10, IT17 and BR17), or neglect what really may seem isolated phenomena, as the ones in BR13.



(a) Italy - Week 1



(b) Brazil - Day 1

Fig. 4. Distribution of packets per community.

B. Community structure

Figure 5 summarizes the structure of popular communities. It depicts a scatter plot that compares the number of ports and ASes in the communities. Every dot represents a community, and the dot sizes represent the average number of IP addresses per AS belonging to it. It helps us to understand how pervasive the activities are in the ASes, e.g., whether only

TABLE I
BASIC STATISTICS PER COMMUNITY - IT

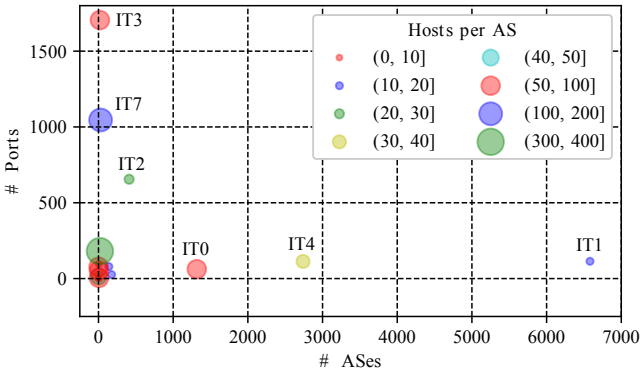
	# Ports	# ASes	# server IPs	# client IPs
IT4	895	2739	760	92122
IT0	222	1316	760	88864
IT2	1426	410	760	8449
IT3	2717	19	760	1072
IT1	439	6581	760	118269
IT7	1519	28	760	3073
IT8	401	19	760	6490
IT6	149	176	760	2921
IT9	105	6	760	359
IT5	89	139	760	2353
IT12	121	6	760	366
IT13	109	2	760	11
IT16	30	2	760	92
IT10	1	1	384	1
IT15	17	1	760	100
IT14	6	1	760	32
IT11	6	1	6	3
IT17	1	1	102	1

TABLE II
BASIC STATISTICS PER COMMUNITY - BR

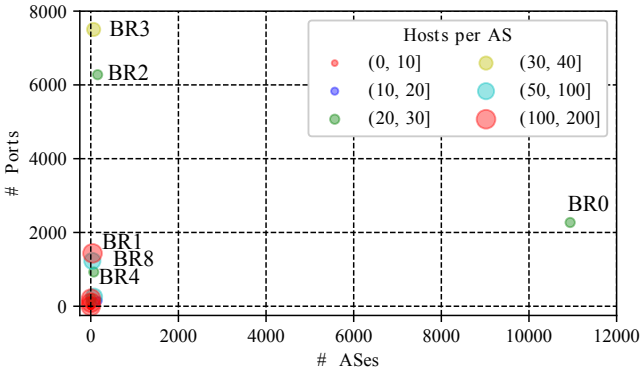
	# Ports	# ASes	# server IPs	# client IPs
BR1	4283	156	8192	3123
BR0	2632	10940	8192	267880
BR2	13637	68	8192	2035
BR8	12513	40	8192	5841
BR3	18053	65	8192	2549
BR4	4628	36	8192	1873
BR6	417	170	8192	2424
BR9	796	75	8192	5240
BR10	450	7	8192	150
BR7	341	5	8192	21
BR15	454	6	8192	1051
BR12	1721	6	8192	633
BR11	488	9	8192	196
BR5	21	1	2638	176
BR16	4	2	8192	13
BR18	12	1	8192	4
BR19	4	1	6168	4
BR17	1	1	8190	1
BR14	2	1	1876	2
BR13	1	1	1	1

some few clients per AS participate in suspicious activities, or whether lots of hosts in the ASes join them.

Focusing on the top figure (Italy), see how the communities that group more ASes (i.e., IT1, IT4 and IT0) have a low number of hosts. Moreover, these ASes send packets to a small number of destination ports. These communities are also among the largest ones uncovered by the modularity algorithm. In a nutshell, the results suggest the existence of sparse sources, distributed in a large number of ASes, targeting some specific (popular) ports. This signature matches scans towards specific services, carried on by very distributed sources (e.g., botnets).



(a) Italy



(b) Brazil

Fig. 5. Structure of communities in the Week 1.

Focus now on communities IT3 and IT7 in the same figure. Those communities are formed by a low number of ASes. However, these ASes present a large number of IP addresses. More interesting, the communities target a large number of destination ports. In other words, we see a small number of ASes in which a large number of IP addresses contribute to scan thousands of destination ports. These characteristics are typical of *horizontal scans*, such as those performed by tools like zmap or nmap, likely run on servers hosted in specifc ASes.

The remaining communities match patterns where a small number of ASes targets specific ports. Those cases include port numbers usually not used by very popular services and ports not exploited in common attacks. This signature hints to sources performing sporadic scans, misconfigurations and other rare anomalies.

Interesting, the bottom plot in Figure 5 confirms results with data from the Brazilian darknet. In this case, however, the number of communities in some sectors of the plot is reduced. Manual inspection shows that some of the communities seen in the Italian darknet may be merged by the modularity algorithm when applied to the (larger and noisier) Brazilian dataset. Other communities are completely diverse. Some source ASes are observed only in one of the datasets and target completely different sets of destination ports. These results confirm previous works [10], [3] that show dissimilarities in traffic reaching darknets deployed at diverse IP ranges.

Fig. 6. Example output of the community detection algorithm.

C. Ports per community

We next dig into the ports each community targets as a way to understand the services these communities are probing. First, we visually explore the communities in search for patterns. One of such visualizations is provided in Figure 6, which reports a sample of the communities for the Italian darknet. Node sizes are proportional to their degrees. The figure shows several behaviors. The community in lilac is an example of targeted action from distributed ASes. It gets represented as a strong concentration around two destination ports, namely ports 23 and 2323. Hundreds of ASes are connected to those nodes. These are ports usually hosting terminal services such as Telnet, pointing to the interest of some sources by those services. A similar consideration holds for the community in green, for which ports 1 433 and 445 are the most common targets. The community in blue, on the other hand, may indicated an horizontal action carried over by groups of ASes, likely scanning hundreds of ports.

The heatmaps in Figure 7 extend the analysis. They show with colors how the traffic for the top-50 ports ($x - axis$) is distributed according to the several communities ($y - axis$). The rightmost column of the plot quantifies, for each community, the percentage of traffic that is *not* related to the top-50 ports. Communities and ports are sorted according to their popularity.

Focusing on the most popular community (IT4) at the top plot (Italy), observe how 92% is targeting several ports in the top-50. The community in particular contacts five ports that often host services vulnerable to remote attacks, such as MSSQL (port tcp/1433) and Microsoft Active Directory (port tcp/445). The second most popular community (IT0) is also strongly concentrated on the top-50 ports, differently from IT4, (e.g., port tcp/22), with only 3.9% of its traffic goes to the remaining ports.

The community IT2 is the clearest example of sources performing large-scale horizontal scans for popular services. It dominates many of the top-50 ports, including services such as Telnet, HTTP, HTTPs, among others. Yet, notice how half of the traffic from this AS community is directed to ports not in the top-50 set. Finally, notice community IT9, which focuses mostly on port 8 291, related to a vulnerability on MikroTik RouterOS Winbox. This targeted behaviour is also visible in other communities that do not dominate any port in top-50 (e.g., IT3, IT5 and IT12).

The bottom plot (Brazil) confirms the general division of communities in (i) vertical scans for popular services – e.g., BR0; (ii) horizontal scans on multiple ports – e.g., BR2 and BR6; and (iii) targeted activity – e.g., BR11. Yet, here we clearly see major differences in formation of communities when compared to the Italian dataset. Notice, for example, how the port tcp/23 (telnet) is dominated by a single community (BR11), whereas it is dominated by sources performing horizontal scans in the Italian case.

D. Temporal behaviour

Finally, we investigate how sources in different communities behave over time. Figure 8 reports timeseries of the number of packets per hour for different ports in a community. We take four arbitrary communities and report activity for their five most active ports. Figure 8(a) and Figure 8(b) show results for IT4 e IT5 covering 1 week. Figure 8(c) and Figure 8(d) show results for BR0 and BR16 covering 1 day to improve visualization.

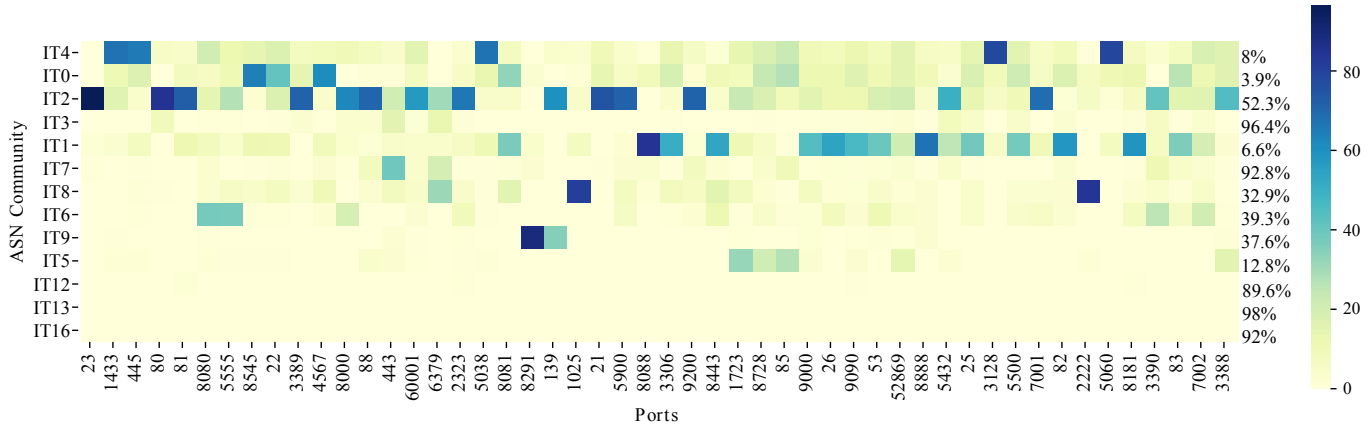
Recall from previous sections that IT4 is the most active community and vertically focuses on popular services. Figure 8(a) confirms, among the top-5 most targeted ports, the prevalence of ports 1 433 and 445, with a lower amount of traffic to ports 22, 23 and 8 291. We see that sources in this community produce a constant amount of noise. A similar level of traffic on each port reaches the darknet without any apparent daily or weekly patterns.

Figure 8(b), on the other hand, shows a targeted community. Here the traffic volume is much less prominent, with only few dozens of packets reaching the darknet in each hour. The ports in the $[1\,540 - 1\,545]$ range are often used by a cluster management framework (RDS services). Notice how the few sources participating in this community have an orchestrated behaviour, alternating the contacted ports after some days of activity, which may suggest the presence of a coordinated, low-rate action. Similar results hold for the other two analyzed weeks, not shown for brevity.

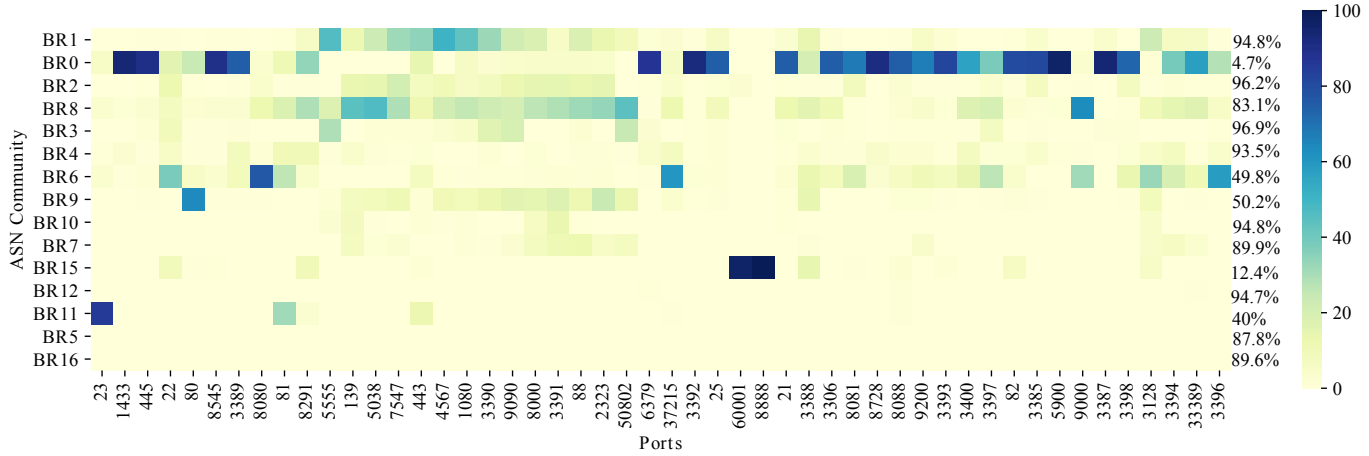
Figures in the Brazilian darknet follows very close patterns. In Figure 8(c) we observe the hourly pattern for the 5 most contacted port in BR0 – an example of community engaged in vertical scans. Figure 8(d) shows instead BR16, an example of targeted community. Similar patterns as for the Italian case emerge in both cases. Notice however that sources in BR16 alternate ports with much higher frequency as before.

VI. Conclusions

In this paper we presented a community detection-based methodology aimed at easing the analysis of large amounts of darknet traffic. Thanks to it, we were able to detect



(a) Italy



(b) Brazil

Fig. 7. Percentage of packets per Port to ASN communities.

coordinated events, such as network scans due to botnets. Our methodology has automatically identified and isolated sources engaging in three major categories of events: vertical, horizontal and targeted scans.

Our work is a preliminary study of the application of community detection algorithms to darknet traffic analysis. Many promising directions emerge. We plan to deepen the investigation on sources behaving similarly. In the same line, we plan to characterize the ASes hosting sources sending packets to darknets. Finally, we observe that some activities in darknet are rather constant, thus becoming less relevant. We plan to apply advanced complex network approaches to filter out the *expected* noisy traffic, so to highlight events that are rare and potentially more interesting for cyber-security applications.

References

[1] 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, 2015.

[2] K. Benson, A. Dainotti, k. claffy, A. C. Snoeren, and M. Kallitsis, "Leveraging internet background radiation for opportunistic network analysis," in *Proceedings of the 2015 Internet Measurement Conference*, 2015, pp. 423–436.

[3] P. Richter and A. Berger, "Scanning the scanners: Sensing the internet from a massively distributed network telescope," in *Proceedings of the Internet Measurement Conference*, ser. IMC '19, 2019, p. 144–157. [Online]. Available: https://doi.org/10.1145/3355369.3355595

[4] S. Abraham and S. Nair, "A predictive framework for cyber security analytics using attack graphs," *arXiv preprint arXiv:1502.01240*, 2015.

[5] 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.

[6] A. Dainotti, K. Benson, A. King, B. Huffaker, E. Glatz, X. Dimitropoulos, P. Richter, A. Finamore, and A. C. Snoeren, "Lost in space: improving inference of ipv4 address space utilization," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 6, pp. 1862–1876, 2016.

[7] M. Jonker, A. King, J. Krupp, C. Rossow, A. Sperotto, and A. Dainotti, "Millions of targets under attack: a macroscopic characterization of the DoS ecosystem," in *Proceedings of the 2017 Internet Measurement Conference*, 2017, pp. 100–113.

[8] A. Dainotti, C. Squarcella, E. Aben, K. C. Claffy, M. Chiesa, M. Russo, and A. Pescapé, "Analysis of country-wide internet outages caused by censorship," in *Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference*, 2011, pp. 1–18.

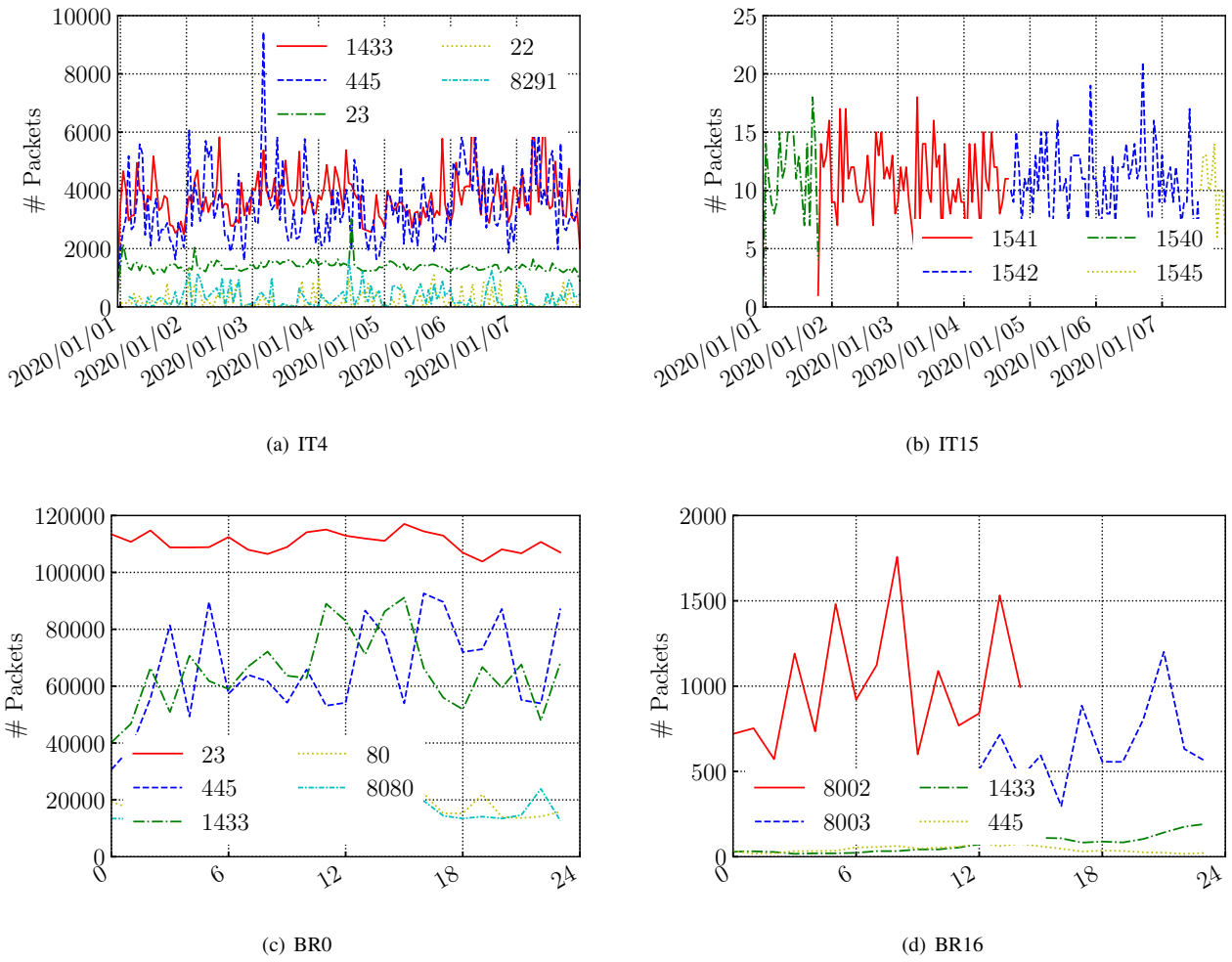
[9] D. Moore, C. Shannon, D. J. Brown, G. M. Voelker, and S. Savage, "Inferring internet denial-of-service activity," *ACM Transactions on Computer Systems (TOCS)*, vol. 24, no. 2, pp. 115–139, 2006.

[10] F. Soro, I. Drago, M. Trevisan, M. Mellia, J. Ceron, and J. J. Santanna, "Are darknets all the same? on darknet visibility for security monitoring," in *2019 IEEE International Symposium on Local and Metropolitan Area Networks (LANMAN)*. IEEE, 2019, pp. 1–6.

[11] W. Harrop and G. Armitage, "Defining and evaluating greynets (sparse darknets)," in *The IEEE Conference on Local Computer Networks 30th Anniversary (LCN'05) l.* IEEE, 2005, pp. 344–350.

[12] 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*, 2010, pp. 62–74.

[13] S. Noel, E. Harley, K. H. Tam, M. Limiero, and M. Share, "Cygraph: graph-based analytics and visualization for cybersecurity," in *Handbook of Statistics*. Elsevier, 2016, vol. 35, pp. 117–167.



(a) IT4

(b) IT15

(c) BR0

(d) BR16

Fig. 8. Time patterns of the top-5 most active ports in selected communities.

[14] C. Phillips and L. P. Swiler, "A graph-based system for network-vulnerability analysis," in *Proceedings of the 1998 workshop on New security paradigms*, 1998, pp. 71–79.

[15] M. Rosvall, A. V. Esquivel, A. Lancichinetti, J. D. West, and R. Lambiotte, "Memory in network flows and its effects on spreading dynamics and community detection," *Nature communications*, vol. 5, no. 1, pp. 1–13, 2014.

[16] 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, virtual reality, visualisation and interaction in Africa*, 2006, pp. 85–89.

[17] M. Coudriau, A. Lahmadi, and J. Francois, "Topological analysis and visualisation of network monitoring data: Darknet case study," in *2016 IEEE International Workshop on Information Forensics and Security (WIFS)*. IEEE, 2016, pp. 1–6.

[18] F. Moradi, T. Olovsson, and P. Tsigas, "An evaluation of community detection algorithms on large-scale email traffic," in *International symposium on experimental algorithms*. Springer, 2012, pp. 283–294.

[19] A. Bar, B. Shapira, L. Rokach, and M. Unger, "Identifying attack propagation patterns in honeypots using markov chains modeling and complex networks analysis," in *2016 IEEE international conference on software science, technology and engineering (SWSTE)*. IEEE, 2016, pp. 28–36.

[20] S. Lagraa and J. François, "Knowledge discovery of port scans from darknet," in *2017 IFIP/IEEE Symposium on Integrated Network and Service Management (IM)*. IEEE, 2017, pp. 935–940.

[21] S. Lagraa, Y. Chen, and J. François, "Deep mining port scans from darknet," *International Journal of Network Management*, vol. 29, no. 3, p. e2065, 2019.

[22] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," *Journal of statistical mechanics: theory and experiment*, vol. 2008, no. 10, p. P10008, 2008.

[23] U. N. Raghavan, R. Albert, and S. Kumara, "Near linear time algorithm to detect community structures in large-scale networks," *Physical review E*, vol. 76, no. 3, p. 036106, 2007.

[24] M. E. Newman, "Modularity and community structure in networks," *Proceedings of the national academy of sciences*, vol. 103, no. 23, pp. 8577–8582, 2006.