Indian Institute of Technology
Kanpur

CS396A

Undergraduate Project Report

Monitoring darknets for detecting malicious activities

Author:
Devashish Kumar Yadav (13240)
Nikhil Vanjani (14429)

Mentor:
Prof. Sandeep K. Shukla
IIT Kanpur

April 28, 2017

Contents

1 Abstract

In an era where every device and service is going online, cyber attacks are on a rise. Attackers actively look for vulnerable devices and services running on the Internet. In this process, they scan the IPv4 address space to find addresses of hosts running vulnerable services. Dark IP space or darknet is defined as a pool of unallocated public IP addresses, which could belong to a certain organization. Monitoring incoming data on the darknet and its analysis could prove useful in detecting malicious activities in the wild. We setup a passive darknet data collection system for IIT Kanpur on a /24 IPv4 address space with some darknet and a few active IP addresses.
We analyze the collected data using our own scripts and some popular NIDSs like bro to ascertain the nature and distribution of the collected data and compare it with previous researches on the topic. We also try to find patterns among the darknet data and relate it with the current vulnerabilities in the wild.

2 Introduction

This report reiterates and extends our previous work[1] on darknet monitoring on IIT Kanpur. IIT Kanpur as an organization has different pools of /24 public IP addresses at its disposal. A part of the network addresses have been allocated to its active services e.g. institute web servers, servers run by different departments. Apart from these there are IP addresses unallocated to any service. In our work these have been referred to as the IP darkspace or darknets[2].

Ideally, there should be no data or very little data received on these darknet IP addresses for example data arising due to mistyped IP addresses. We set up a system to monitor darknet monitor on a /24 IP public IP address space of IIT Kanpur. We categorize and analyze the captured data to determine possible malicious activities directed at IIT Kanpur and in general potential vulnerabilities being exploited in the wild. This report is accompanied by a presentation which can be downloaded from the same source.

3 Background

There have been previous works on monitoring darknets focusing on different aspects of the darknet data to reach to some interesting conclusions. A comprehensive survey of such previous work can found referenced in [3]. Our works builds up on many them as we create a darknet monitoring system for IIT Kanpur.

3.1 Darknets Definition

Darknets are unallocated public IP addresses. For monitoring, normally a contiguous set of public IP addresses is chosen. Attackers keep scanning the Internet address space looking for vulnerable devices and services which can be exploited. Darknets can be used to detect such activities passively without interacting at all and hence not revealing any information about themselves.

3.2 Monitoring Systems

The idea of various types of monitoring systems is to set up online data collection/monitoring systems for detecting malicious activities. For this purpose we felt the need to list the categorization as done by [3] to highlight the subtle differences between these systems. We also note that our system is a mixture of more than one categories making it one of its kind.

- Passive monitoring systems
  - **Darknets** : Darknet monitoring systems similar to our setup collect data in a passive manner. In different versions of such monitoring systems there is zero or minimal interaction with the source.
- Active monitoring systems
  - **Honeypots** : These are interactive monitoring systems which disguise themselves as legitimate hosts in an attempt to trap attackers. These can be further classified into subcategories depending on the level of interaction with the incoming data.
  - **Greynets**[4] : These systems comprise of active addresses interluded with darknets. Our setup also contains a few active addresses but at the time of writing this report we didn't control responses on those addresses.

3.3 Nature of TCP Traffic

After separating TCP traffic from the total data we categorize TCP packets received on darknet as follows.

- **Scanning Traffic** : The purpose of scanning traffic(TCP SYN packets) can be to obtain reconnaissance of services running on different ports on a single host, or a particular service (like SSH) running on a port on multiple hosts. These have been generally referred to as vertical and horizontal scans respectively[5].

Vulnerable services on these hosts can be exploited to gain access to an internal network, spread malwares both inside a victim organization and outside. Hosts can also be compromised to become part of a botnet which could later be used to carry out Denial of Service attacks[6].

- **Backscatter Traffic** : TCP SYN+ACK, ACK, RST, RST+ACK packets. Usually backscatter traffic refers to responses to communications with spoofed source IP addresses. The backscatter traffic could be responses of attempts of denial-of-service attack[7] elsewhere. An example is a TCP SYN flood attack[8] where the attacker floods the victim with TCP SYN packets with possibly spoofed source IP.

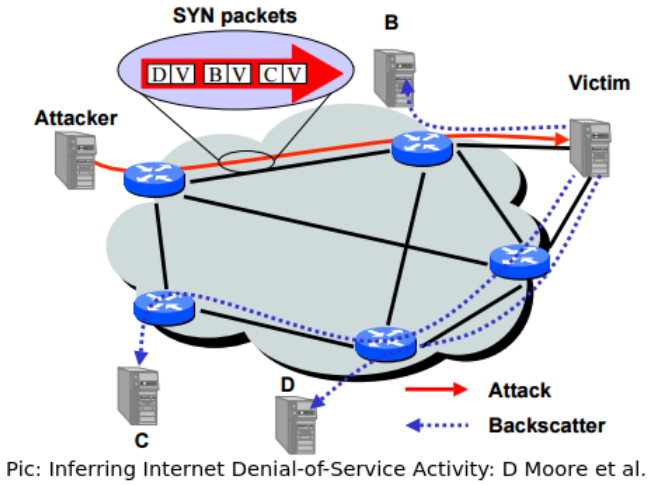


Figure 1: Example of origin of backscatter traffic[9]

- **Misconfigured Traffic** : Rest of the traffic; mostly due to some internal misconfigurations etc.

4 Setup

We collect data from a mirrored port in a IIT Kanpur /24 network with 162 Dark Space IP addresses and 14 active IP addresses. We write custom scripts that use *tcpdump* along with specially crafted filters to store data on our machine. Post capture we analyze the captured data using various scripts and NIDS bro. We also incorporate feeds from Collective Intelligence Framework(CIF)[10] along with Bro for better insights.

5 Darknet Profiling

We perform profiling based on various aspects like transport layer protocols, nature of TCP traffic, TCP and UDP ports distribution. We also compare our results with those obtained by Claude et al[2].

For darknet profiling, we use different filters for tcpdump and wireshark to refine and categorize the data into active and darknet and then further separate various subcategories. We write python scripts to distinguish data by nature of TCP traffic, TCP/UDP port distribution and perform some advance statistical analysis. We use Matlab for visualizations of the distributions and create the time series plots.

5.1 Protocols distribution and Nature of TCP ports

	tcp	udp	icmp	other
packets	43,548,519	4,476,557	1,473,156	140,487
percent(#packet)	87.73%	9.01%	2.97%	0.28%
bytes	2698459844	1017078171	142532007	37544388

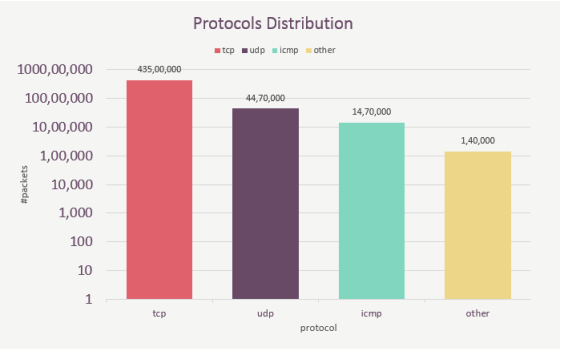
Table 1: Distribution of data by transport layer protocols

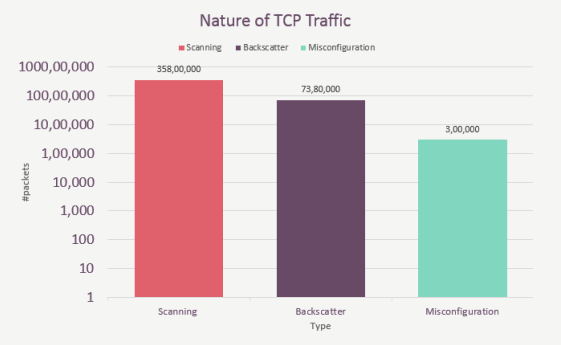
	Scanning Traffic	Backscatter	Misconfiguration
packets	35,862,917	7,385,003	300,599
percent(# Packets)	82.35%	16.95%	0.69%
bytes	2184807886	493046909	20605049

Table 2: Distribution by nature of TCP traffic

The percentage distribution in Table 1 is similar to that obtained by Claude et al[2].

The nature of TCP traffic results in Table 2 reveal that scanning or network probing activities constitute the majority of darknet traffic on the monitored IIT Kanpur. Such traffic could be interpreted as an indication of port scanning and/or vulnerability probing. Such attacks, in general, are preliminary triggered before launching a targeted attack towards a specific system. It is interesting to note that the contribution of backscatter traffic is quite significant as compared to the results obtained by Claude et al[2] (scanning - 68%, backscatter - 2%, misconfiguration- 30%)





5.2 TCP and UDP destination ports distribution

For TCP ports, we observe around 44% traffic on port 23. This is way more in percentage as compared to the profiling done by Claude et al.[2]. As seen from the tables and figures, as expected services running on common ports like telnet, SSH and database servers are most attractive points of scan. Apart from this, there are a few ports highlighted which received an unusually high traffic. We take a closer look at them to find their connection with the scans by Mirai botnet for vulnerable IoT devices and other recent events.

For UDP ports, we observe highest traffic on ports 5060 which is used for Session Initiation Protocol(SIP) and 53413. SIP might be expected as at any time there are scanning activities looking for SIP servers. The scans on UDP port 53413 can be attributed to attackers looking for the exploitation of backdoor in the Netis routers.

TCP Services Destination Ports		
Port	Packets	Percent
23 (Telnet)	19,398,733	44.54
1433 (ms-sql-s)	1,850,730	4.24
22(SSH)	1,698,446	3.90
2323	1,505,190	3.45
7547	1,272,297	2.92
5358(wsdapi-s)	1,203,563	2.76
3306(mysql)	1,126,883	2.58
23231	925,685	2.12
3389(RDP)	742864	1.70
6789	589161	1.35
80(HTTP)	409729	0.94
8080(http-alt)	324896	0.74

UDP Services Destination Ports		
Port	Packets	Percent
5060(SIP)	750502	29.87
53413(Netcore Routers)	718045	28.58
1900(UPnP)	409829	16.31
123(NTP)	180668	7.19
33434(traceroute)	110926	4.41
53(DNS)	102722	4.08
161(SNMP)	78955	3.14
19(Chargen)	42894	1.70
1701(vpn)	32721	1.30
17(qotd)	28718	1.14
111(SunRPC)	28692	1.14
27244	27239	1.08

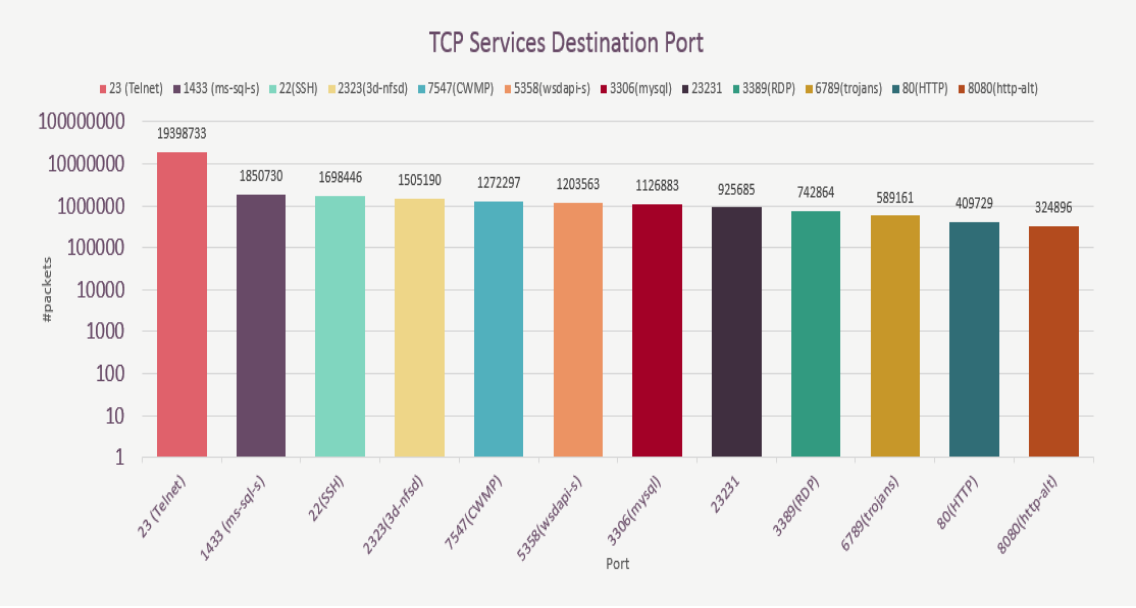


Figure 2: distribution of tcp service destination port

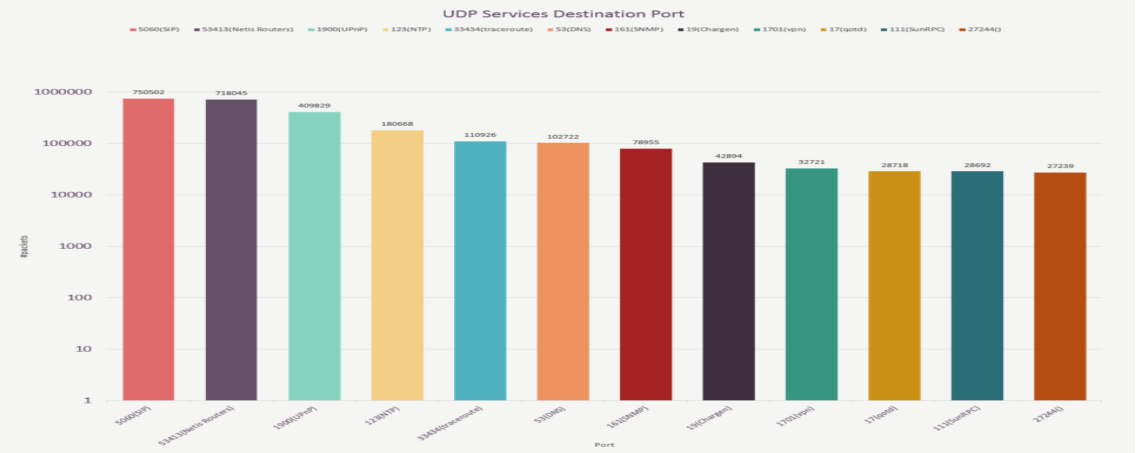


Figure 3: distribution of udp service destination port

5.3 Time Series

Time series analysis could be a provide some useful insights. We used it to obtain the change in trends of darknet data over some ports which were receiving unusually high traffic and associate them with vulnerabilities in the wild. Time series analysis can also be used to find the distribution of incoming data throughout a day which could be helpful for a relative comparison of number of packets for different protocols/ports. It has been used in the past for building forecasting models for DoS attempts[11].

Here we visualize the time series of various protocols and nature of TCP traffic.

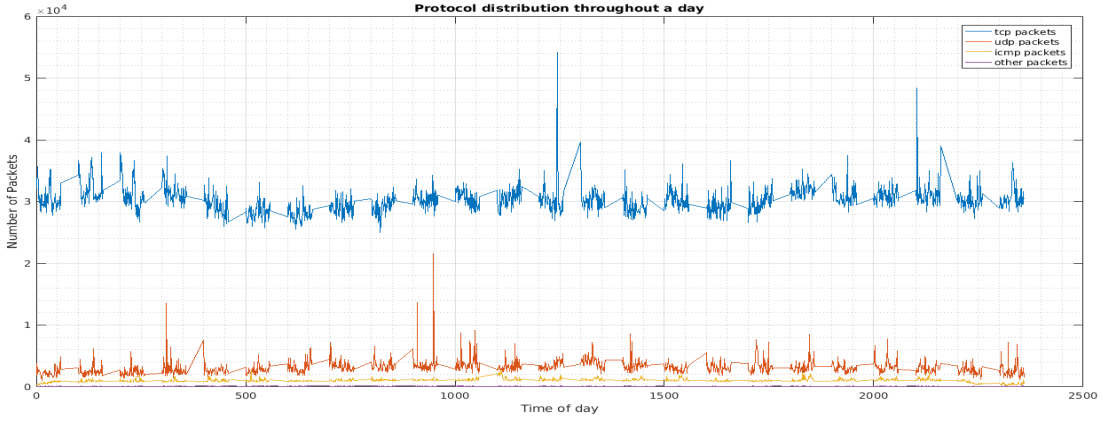


Figure 4: Average Protocol distribution throughout the day

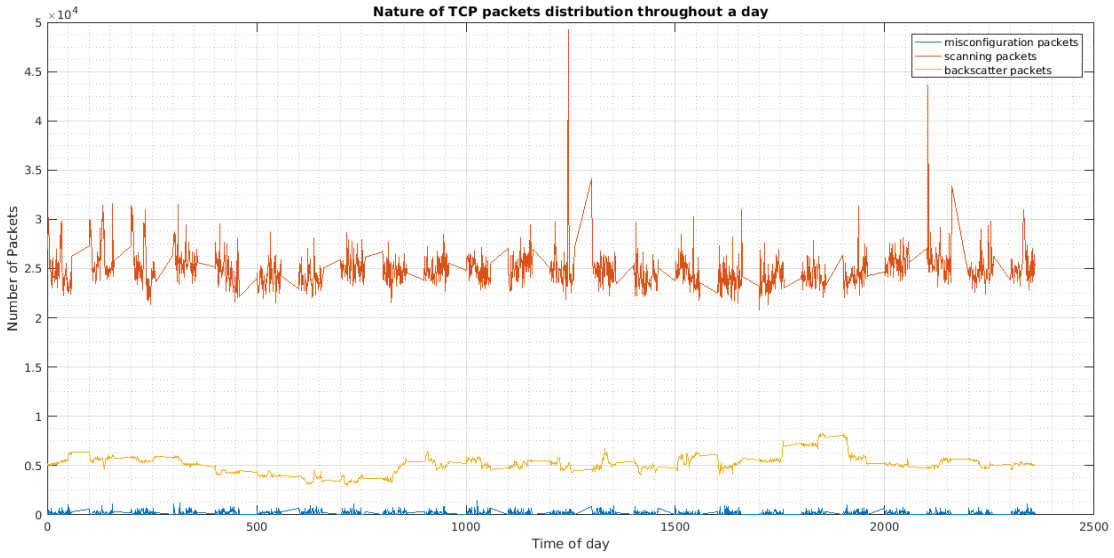


Figure 5: Average nature of TCP traffic throughout the day

6 Analysis

6.1 Time between packets

Scanning activities generally scan a large area either horizontally, vertically(number of ports) or both. Most scans deploy automated scripts to carry out the activity. Too fast a scan can alert the organization being scanned due to sudden influx of data. Therefore, attackers generally minutely vary their time of sending subsequent packets or slow it down considerably to avoid raising flag. We make a distribution of how the average time between packets differ for different scanning attempts from different source addresses.

Since most scan attempts cover a large area, they often deploy automated scripts. We filter out a pool of continuous darknet IP addresses for our analysis (30 in our case). We write a sophisticated python script to create a distribution of these scan attempts. We run Bro with varying parameters to find possible scanning attempts. We map these attempts categorized by different source IP addresses, same source IP address - repeated attempts using python. We refine the data based on certain thresholds(5 minute threshold to receive the next packet) and take the average time used between different scan attempts. We note that the histogram plotted for the time then resembles a normal distribution. The mean and standard deviation of the

following distribution are 14 seconds and 4.214 seconds. Apart from this there are packets that are received at a time difference very close to zero. We conclude that these could be due to scanners sending multiple packets in parallel.

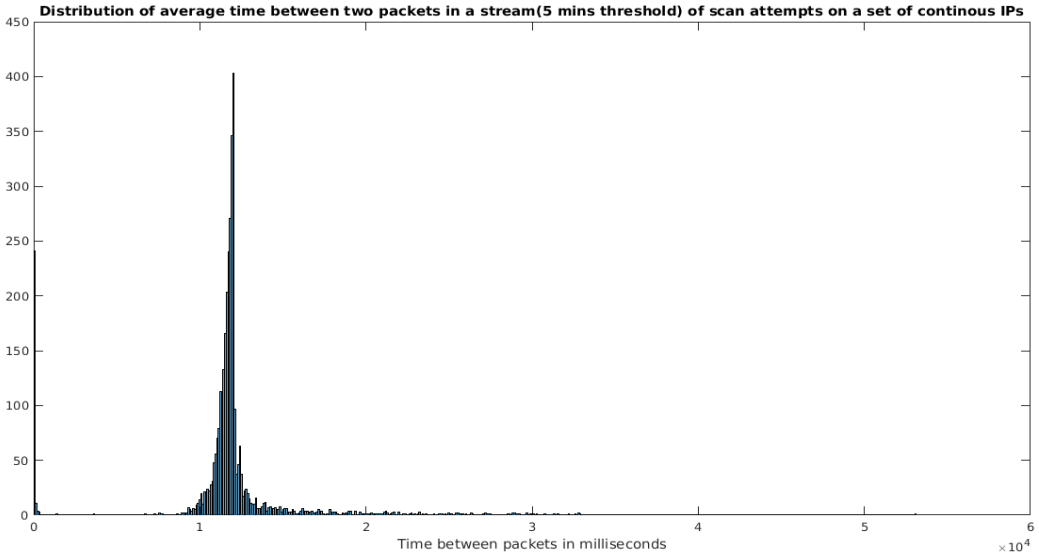


Figure 6: Average time between subsequent packets in a scan attempt

6.2 Time Series Analysis

Here, we provide some observations about the ports observed to be having unusually high scan activity. We show that the zero day vulnerabilities discovered during this time period were also observed in IIT Kanpur's network. The emphasis of this study is that because IIT Kanpur also has a big Internet backbone and the observations made globally resonate with those here, so by investing resources in continuous monitoring of scanning activities on all ports, we can contribute our bit to detecting zero day vulnerabilities.

6.2.1 Mirai Botnet

Mirai Botnet is a program designed to harness insecure IoT devices to run massive denial of service attacks. It is designed to run on IP cameras and other Internet-connected devices. It tries various hard-coded root passwords, infects the device, and then sends out traffic to a preset target. It was first found in August 2016.

Hackers used the botnet to send a 620 Gbps DDoS to KrebsOnSecurity, a popular security blog by Brian Krebs.(Source [12])

In our data there were significant activities on ports being normally scanned by mirai. Unusual traffic detected on port 2323 on starting time around Dec 16, 2016; on port 6789 around Dec 19, 2016; on port 23231 around Dec 23, 2016. Our observations for detection are also in accordance with the global observation as can be seen in the following figure 7.

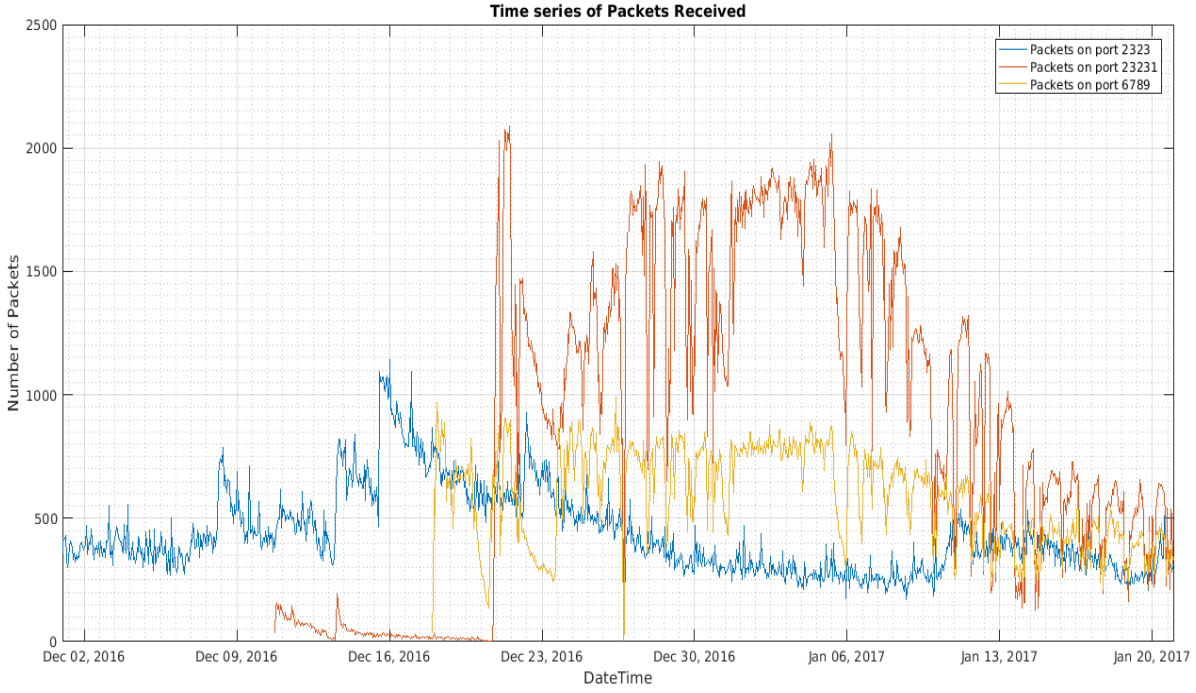


Figure 7: Scan attempts on port 2323, 232321, 6789 for 50 day period

6.2.2 Scans on telnet port 23

This is the most attractive port in our darknet data as can also be seen in section 5.2 . The most common service running on it is Telnet, the most popular program for remote access to unix machines. It has numerous security vulnerbilities like stack-based buffer overflow, buffer overflow in remote command servers, many trojans use this port.[13]

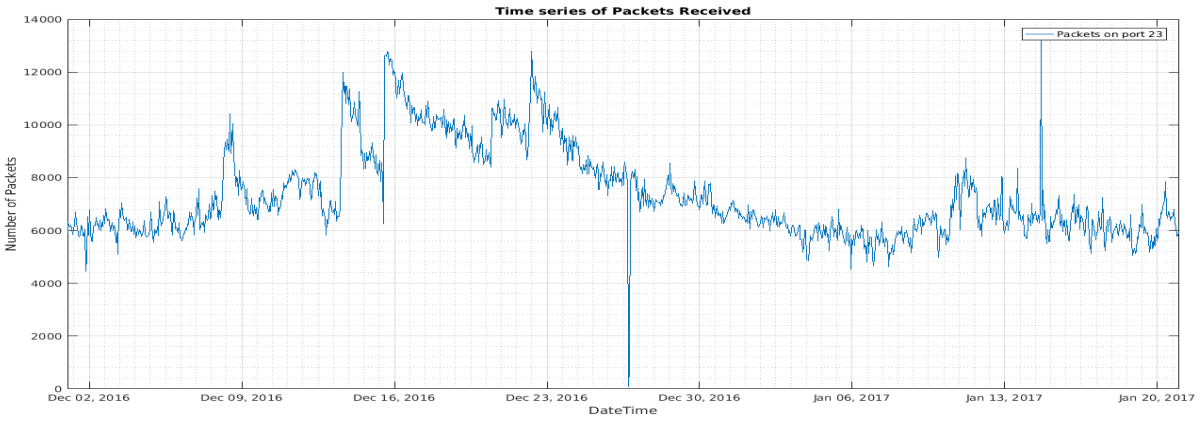


Figure 8: Scan attempts on port 23 for 50 day period

6.2.3 Scans on port 7547

Mirai botnet's source code was made public in September and later on hackers modified the code to increase the number of IoT devices they could compromise. An occurrence of such attempts came in news in late November when traffic on port 7547 increased to the extent that it was 2nd most scanned port after port 23. It was discovered that some hackers were exploiting some modems produced by Deutsche Telekom on which port 7547 was open.[14] The attack can be used to execute arbitrary commands. From our data we observer that the scanning activity on this port was high around first week of December'16 matching the time of the attack(Last week of November'16).

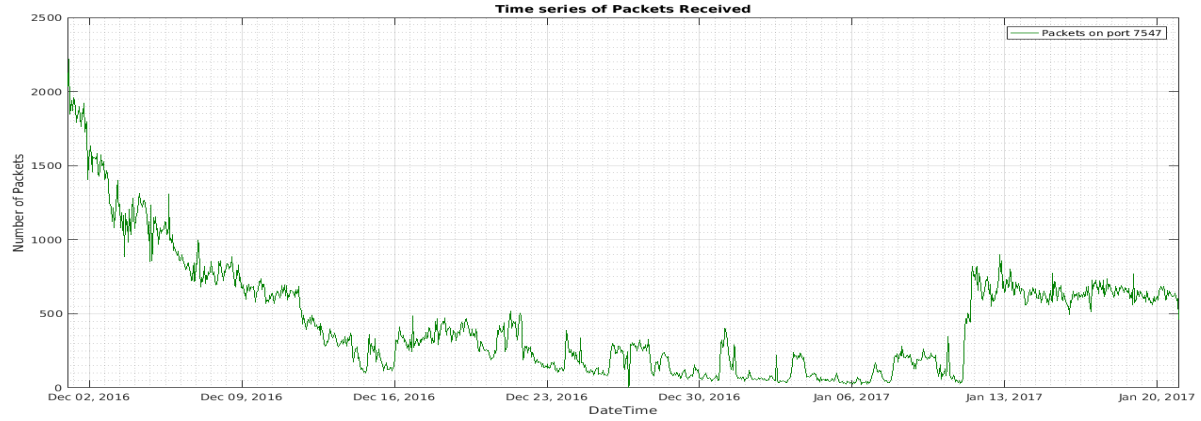


Figure 9: Scan attempts on port 7547 for 50 day period

6.3 Collective Intelligence Framework (CIF)

CIF is used as a cyber threat intelligence management system. It constantly keeps updating information and provides feeds about known botnets, malwares, scanners, spams, hijacked IPs from many sources and also associates some confidence value with each of them.[10]

We have data of darknet as well as few active IP addresses which are distributed in between the dark space. Motive behind port and address scan can be established by analyzing data from same source IP addresses in darknet and active data. Our system differs from a honeypot in the manner that the few active IP addresses present are not fake machines but real services running on the network. And moreover we don't control them.

We used CSIRT Gadgets' massive-octo-spice[10] project for installing CIF and obtained the feeds of IPv4 addresses on botnets, hijacked, malware, scanner, spam for a certain threshold confidence(above 0.5) and then compared these IP lists with lists of IPs scanning our darknet. We also incorporated these feeds with NIDS bro and ran it on our data from active IP addresses to generate the intel.log files which identified the IP addresses which tried to make contact with our network.

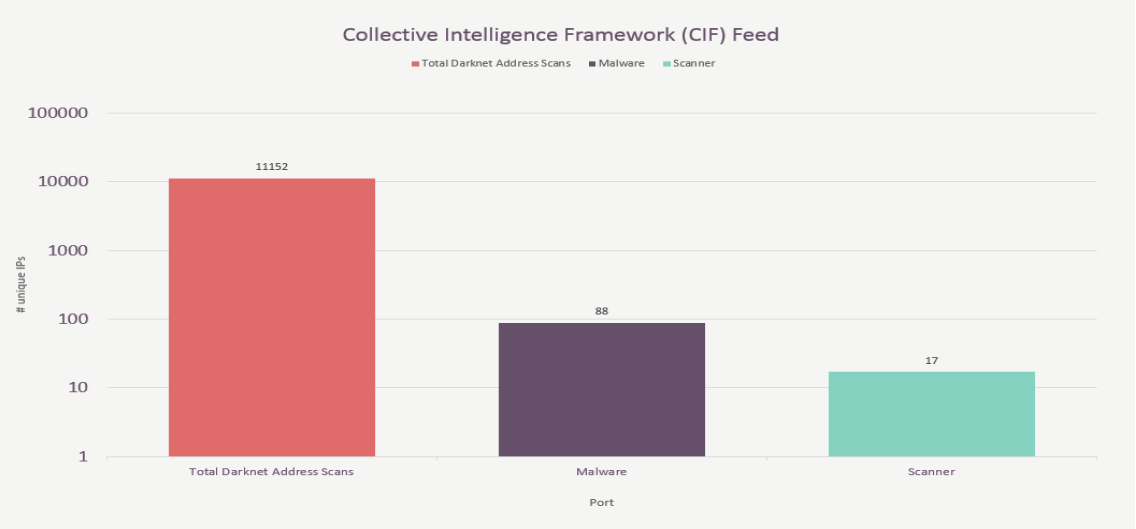


Figure 10: CIF analysis I

Total unique IPs performing darknet address scans = 11152
Number of IPs among these classified as malware sources (confidence > 60%) = 88
Number of IPs among these classified as scanners (confidence > 75%) = 17

Total unique IPs performing darknet address scans = $|setA|$ = 11152

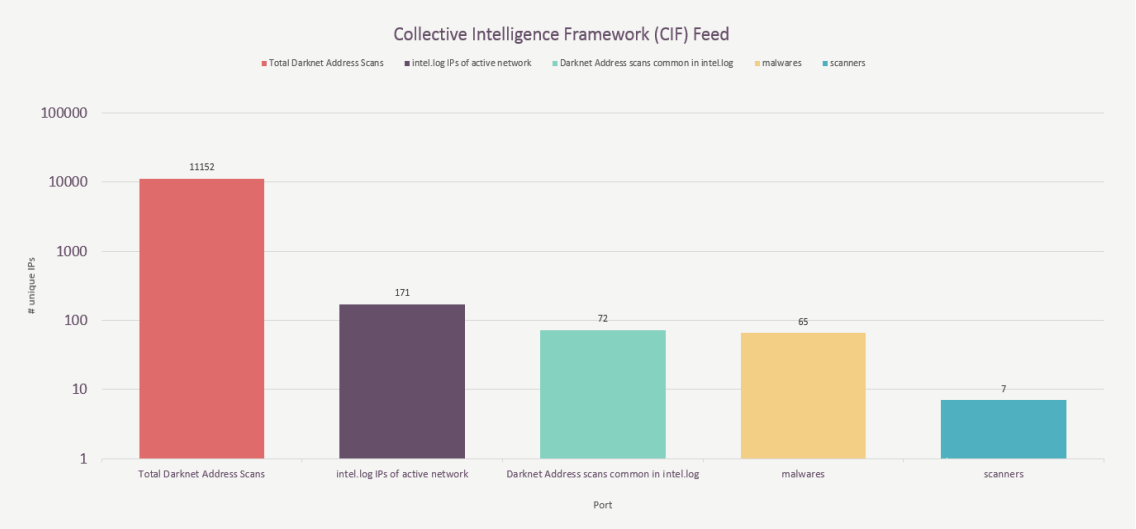


Figure 11: CIF analysis II

Total unique IPs collected in intel.log for active network = $|setB| = 171$
Number of IPs common in set A and B = $|setC| = 72$
Number of IPs among these classified as malware sources (confidence $> 60\%$) = 75
Number of IPs among these classified as scanners (confidence $> 75\%$) = 7

6.4 Geographical Distribution

Geographical analysis of source IP address and scan attempts can reveal specific attacks or malwares originating from a specific country or region. These can be sometimes misleading as many attacks happen through remotely controlled machines or VPN services. Yet this provides an overall view of the origination of packets of certain types based on the geographical location.

We create a python script which utilizes both online and offline resources to geolocate a given list of IP addresses. The script fetches longitudes and latitudes of the provided list of IP addresses addresses and creates HTML pages with corresponding data to plot them on Google Maps. We have included a few maps highlighting the geographical distribution of packets received on IIT Kanpur darknet data collection system.

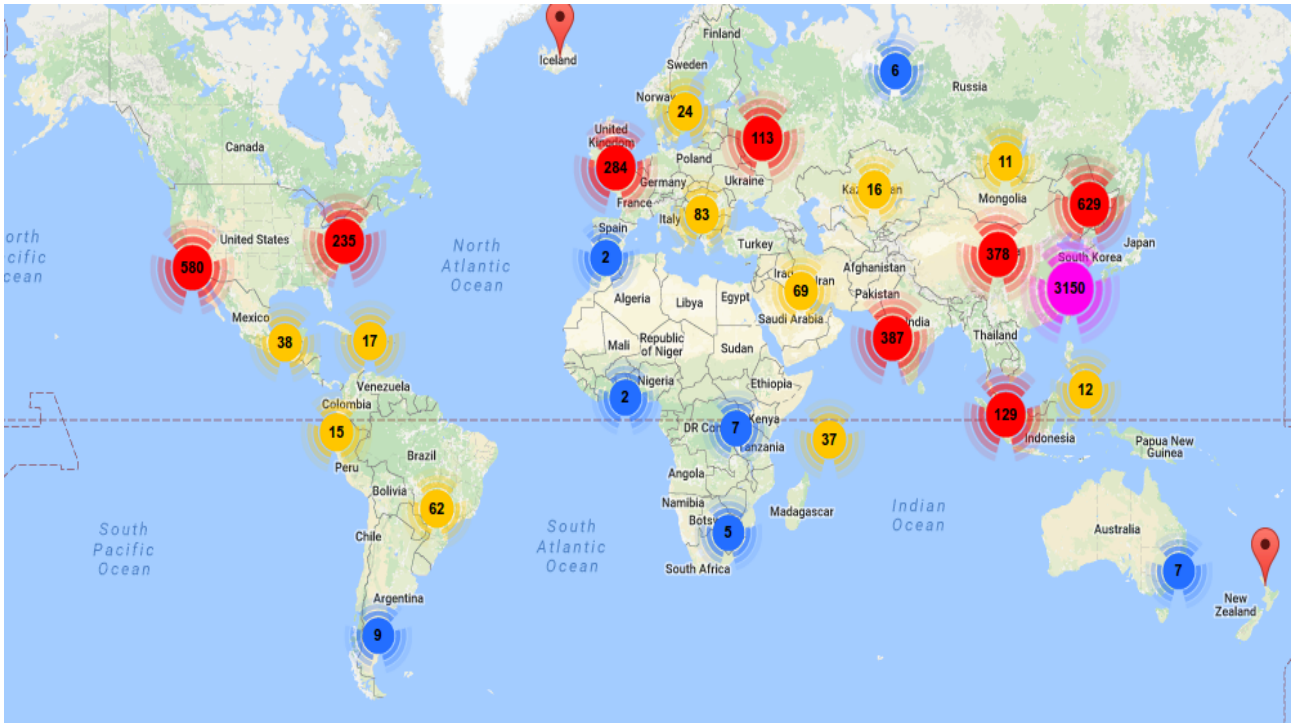


Figure 12: Total scan attempts from unique IP addresses on IIT Kanpur

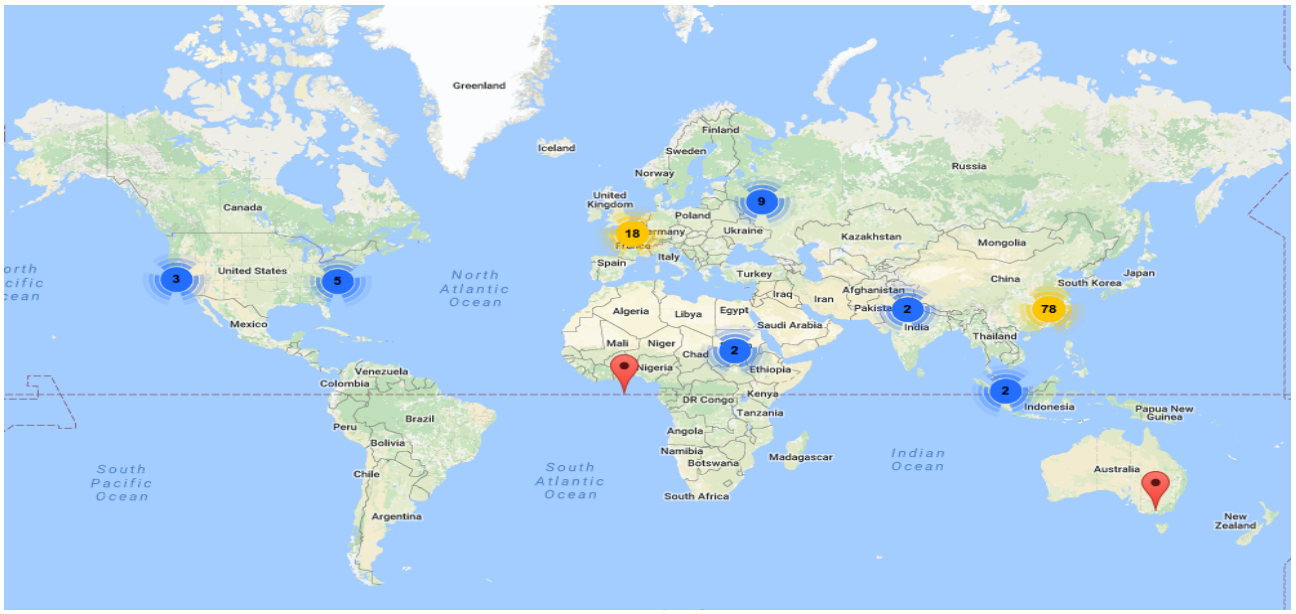


Figure 13: Total password guessing attempts on port 22(SSH) identified by bro

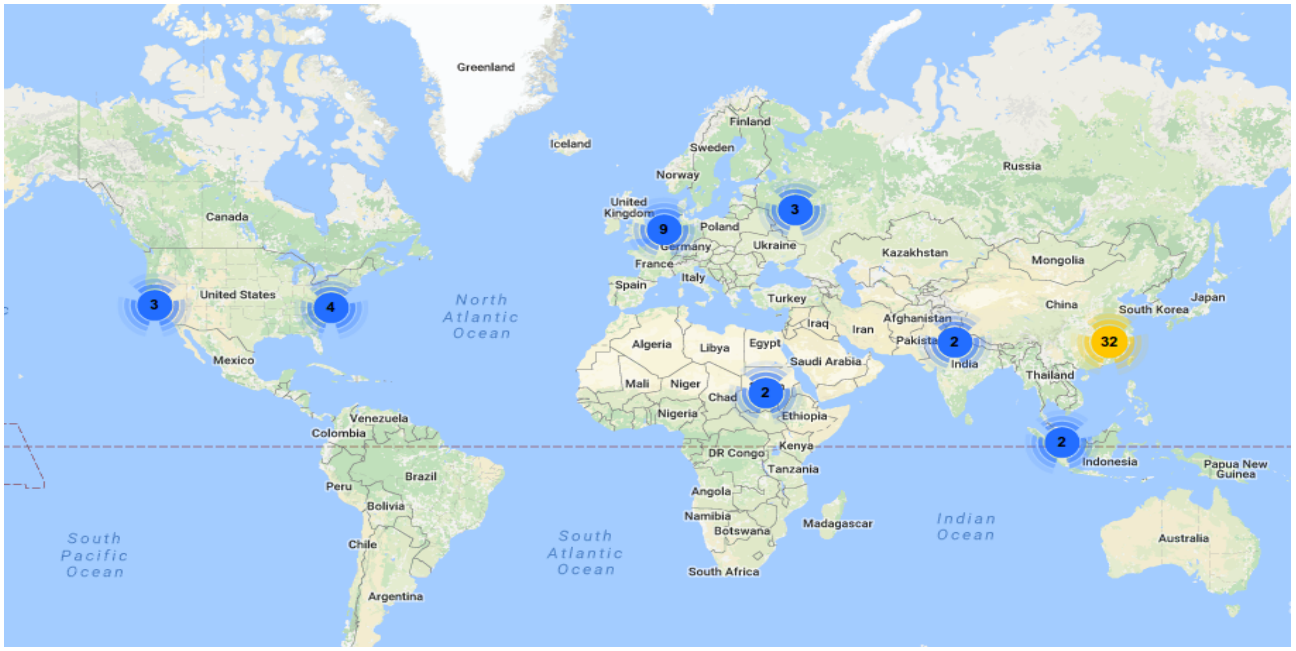


Figure 14: 22(SSH) password guessing IP addresses which also tried scanning our network during the length of our study

7 Conclusion

- As also mentioned in previous report[1] the major challenge with passive monitoring on darknet IP space is that since there is no interaction with the attacker majority packets on the darknet IP space are only possible scanning attempts. Since the actual attack can't happen on a darknet monitoring system, we can at best speculate the motive of these scans.
- For a better analysis we require a monitoring system with at least a daemon running preforms basic tasks like completing a TCP handshake and acknowledging data. This is safer and less interactive than active monitoring systems such as honeypots as this is very much restricted in its replies.
- An attacker actively scanning a network is unlikely to try again once they find an active service.
- Active IP addresses form an essential part in analysis of darknet data. Greynets can thus provide a deeper insight than simple darknets.

References

[1] Cs498a (2016-17 sem 1) monitoring darknets for detecting malicious activities - http://web.cse.iitk.ac.in/users/dugc/ugpreports2016-17/devyadav.pdf.

[2] Claude Fachkha, Elias Bou-Harb, Amine Boukhtouta, Son Dinh, Farkhund Iqbal, and Mourad Debbabi. Investigating the dark cyberspace: Profiling, threat-based analysis and correlation. In *7th International Conference on Risks and Security of Internet and Systems, CRiSIS 2012, Cork, Ireland, October 10-12, 2012*, pages 1–8, 2012.

[3] C. Fachkha and M. Debbabi. Darknet as a source of cyber intelligence: Survey, taxonomy, and characterization. *IEEE Communications Surveys Tutorials*, 18(2):1197–1227, Secondquarter 2016.

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

[5] Cynthia Bailey Lee, Chris Roedel, and Elena Silenok. Detection and characterization of port scan attacks. *Univeristy of California, Department of Computer Science and Engineering*, 2003.

[6] Esraa Alomari, Selvakumar Manickam, B. B. Gupta, Shankar Karuppayah, and Rafeef Alfaris. Botnet-based distributed denial of service (ddos) attacks on web servers: Classification and art. *CoRR*, abs/1208.0403, 2012.

[7] D. Moore, C. Shannon, D. Brown, G. Voelker, and S. Savage. Inferring Internet Denial-of-Service Activity. *ACM Transactions on Computer Systems*, 24(2):115–139, May 2006.

[8] Haining Wang, Danlu Zhang, and Kang G. Shin. Detecting syn flooding attacks. In *Proceedings.Twenty-First Annual Joint Conference of the IEEE Computer and Communications Societies*, volume 3, pages 1530–1539, June 2002.

[9] David Moore, Colleen Shannon, Douglas J. Brown, Geoffrey M. Voelker, and Stefan Savage. Inferring internet denial-of-service activity. *ACM Trans. Comput. Syst.*, 24(2):115–139, May 2006.

[10] Collective intelligence framework(cif) : https://github.com/csirtgadgets/massive-octo-spice.

[11] C. Fachkha, E. Bou-Harb, and M. Debbabi. Towards a forecasting model for distributed denial of service activities. In *2013 IEEE 12th International Symposium on Network Computing and Applications*, pages 110–117, Aug 2013.

[12] Mirai botnet - techcrunch : https://techcrunch.com/2016/10/10/hackers-release-source-code-for-a-powerful-ddos-app-called-mirai/.

[13] Port 23 details : http://www.speedguide.net/port.php?port=23.

[14] Mirai evolving: New attack reveals use of port 7547 : https://securityintelligence.com/mirai-evolving-new-attack-reveals-use-of-port-7547/.