October 2019
Geoff Huston

Dark Traffic

Some time ago a number of Internet researchers started listening to the *background radiation* noise of the Internet.

> What do I mean by *background radiation*? I'm stealing the term from astronomy of course, where the term referred to the accidental discovery in 1964 of cosmic background radiation by American radio astronomers Arno Penzias and Robert Wilson. Their work confirmed earlier theoretical work concerning the remnants of the original "Big Bang" origin of the cosmos, which won them the 1978 Nobel Prize for Physics.

The Internet's dark traffic experiments are nowhere near so momentous, and refer to a simple experiment: if you were to announce some IP address prefix into the inter-domain routing space, and set up a host to record every packet that arrived destined to addresses within this announced address prefix what would you see? At no stage does the experiment's server respond to any incoming packet, so the collector is a *dark* collector that simply absorbs the traffic. At no stage are the announced addresses referred to as service addresses in the DNS, so the traffic is entirely unsolicited in any way. If absolutely nothing refers to these addresses, and no packets are emitted from these addresses, then would we expect to see any incoming packets?

To give away part of the story, the answer is that yes, incoming packets are a certainty in IPv4. "Why? is a good question at this stage. Is this just address scanning from tools such as zmap (https://zmap.io) and similar? Is this misconfiguration? Or is this the backscatter from various forms of *source spoofing attacks*. Or is this traffic just the remnants of an address-scanning viral attack, such as the *conficker virus* (https://en.wikipedia.org/wiki/Conficker)?

At APNIC Labs we first looked at this aspect of the Internet in 2010 when APNIC was assigned 1.0.0.0/8. At the time we were concerned that some of these addresses in this particular address prefix were just too "toxic" for normal use, in that they attracted so much unsolicited traffic that any other use of the IP address would be overwhelmed by the torrent of incoming garbage. We published some studies on what we saw in this and similar experiments in other IPv4 and IPv6 address prefixes (https://www.potaroo.net/studies/), and also presented on the results at various operational forums (such as https://www.potaroo.net/presentations/2011-01-28-ip-background-radiation.pdf).

Some ten years ago when we were undertaking this study, we saw at the time a very strong signal relating to the *conficker virus*. We also observed address scanners, misconfigured systems, game rendezvous traffic and source address spoofing, but by far the largest signature of unsolicited IPv4 traffic at the time was the result of the *conficker* address scanning protocol (this was evident due to an odd behaviour of *conficker* in that it would only scan the lower 'half' of each /8 address range).

IPv6 Dark Traffic?

We expected that IPv6 would be different. The far larger address span would mean that the IPv4 address scanner tools would just not work in IPv6, and this theory appeared to be supported by the evidence at the time (https://www.potaroo.net/ispcol/2010-07/dark6.html).

It is certainly true to state that what we see as dark traffic in IPv4 has no direct counterpart in IPv6. The nature of the massively sparse population of the low-end 64-bit interface identifier addresses in IPv6 makes address scanning pretty much impractical. There may be some small number of guess probes being directed to x::1 and x::2 addresses, but on the whole there is no evidence of any systematic scan of address space happening across all IPv6 addresses. So far there is no direct evidence of virus scanners probing into the dark address blocks in IPv6.

What we do see is some evidence of configuration errors in IPv6. The overwhelming volume of the traffic seen in this exercise is not truly dark packets, but leakage from private use contexts. Due to a failure in the local configuration, a sizeable amount of supposedly private network traffic is incorrectly sent out into the public IPv6 Internet. To a much lesser extent there is a small volume of dark traffic that is the result of transcription errors in editing DNS zone files with IPv6 addresses and local system configuration in manually setting up local IPv6 interface addresses.

Back to IPv4 dark traffic.

One Day in Japan

Let's zip forward to the present time, where Maztsuzaki Yoshinobu presented the results of a recent study on "The Background Noise of the Internet" at the APNIC 48 conference in September 2019 (https://conference.apnic.net/48/assets/files/APIC778/Background-noise-of-the-Internet.pdf).

This work was about trying to attribute a primary motivation to the received dark traffic. This could be due to malware propagation, address scanning and broken local configurations or reflection attacks of various forms. The data reflects a single day (10th January 2019) and the capture volume is some 600M packets. The breakdown observed was:

TCP	95%	577,340,492
UDP	4%	26,945,104
ICMP	1%	3,897,454
IP6	0%	2153

Most of the TCP traffic is the initial SYN of a TCP exchange, which is not unexpected in the context of dark traffic. Some 2% is a SYN+ACK which is either a response to an original SYN packet that used a spoofed source address or a scanning probe packet. All kinds of TCP flags values were seen including a close to "full house" where a packet contained every TCP flag in the same packet.

There are some curiosities in the observed data. The *telnet* protocol has fallen into disuse these days, yet 73M TCP packets were addressed to port 23, the *telnet* port. Universal Plug and Play (UPnP) is the gift that keeps on giving. One interesting issue with UPnP (aside from the fact that it never ever should be exposed to the Internet, but often is), is the fact that it can be reached via various routes. One of the

lesser used routes is SOAP requests via TCP port 52869. Many hosts sent less than 10 packets, while a few sources, presumably scanners, send millions of packets.

Four Years in APNIC

This presentation has prompted me to take a look at a longer-term data collection that we've been assembling in APNIC. The address block is an IPv4 address prefix and it's been used as a dark traffic collector since March 2016, so the accumulated data set of received packets is considerable. The data collection is not continuous, as there have been interruptions to the collection in that period, but it does provide a long baseline data set that we can use to answer some questions about background radiation in today's IPv4 Internet.

The first is a look at the total volume of bytes generated by this dark traffic. This is shown in 5-minute averages across the entirety of the collection period in Figure 1.

There are a couple of notable aspects to this traffic profile. Firstly, over the four-year period the volume of this dark traffic is increasing. In 2016 we observed between 100Kbps to 300kbps of such traffic, while in September 20198 the average 5-minute incoming traffic volumes are between 300kbps to 700kbps. Secondly, the major change appears to lie in TCP packet volumes, while the UDP profile has not altered so significantly.

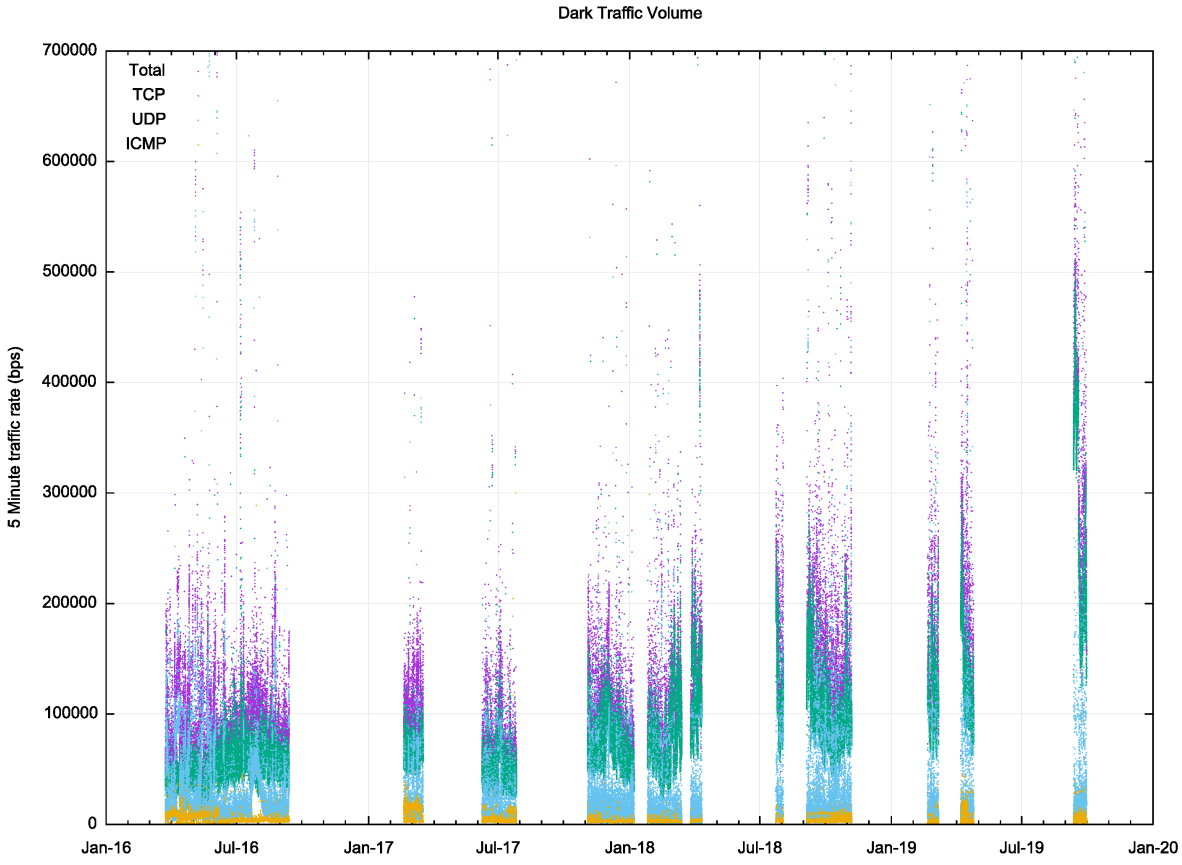

Figure 1 – Dark Traffic – 5-minute average traffic rate

This relative growth of TCP traffic over UDP and ICMP traffic is also evident when we look at the 5-minute average packet count over the same period, as shown in Figure 2, where the UDP rate is relatively steady while the TCP rate has increased.

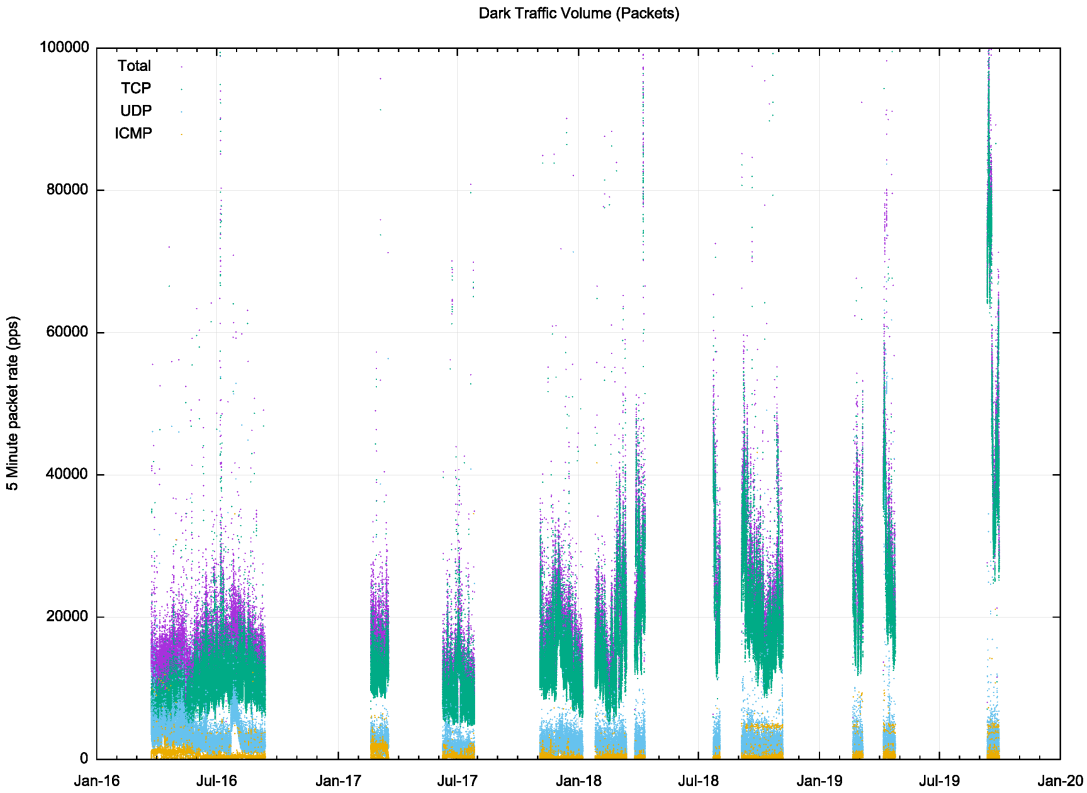

Figure 2 – Dark Traffic – 5 minute average packet rate

The monthly average profile is shown in Figures 3, 4 and 5.

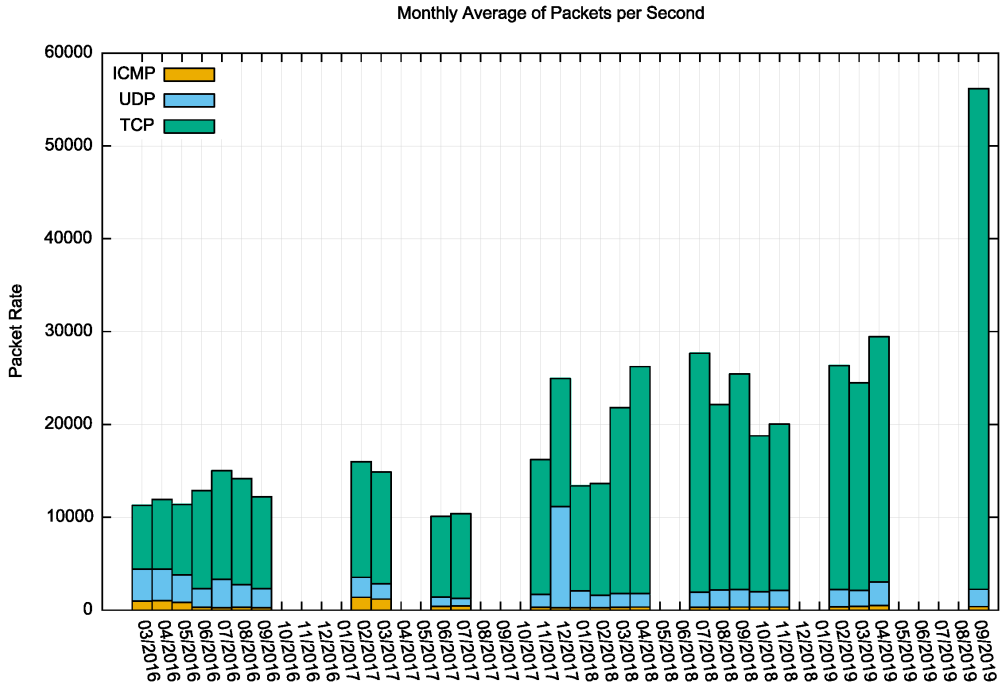

Figure 3 – Dark Traffic – monthly average packet rate per second

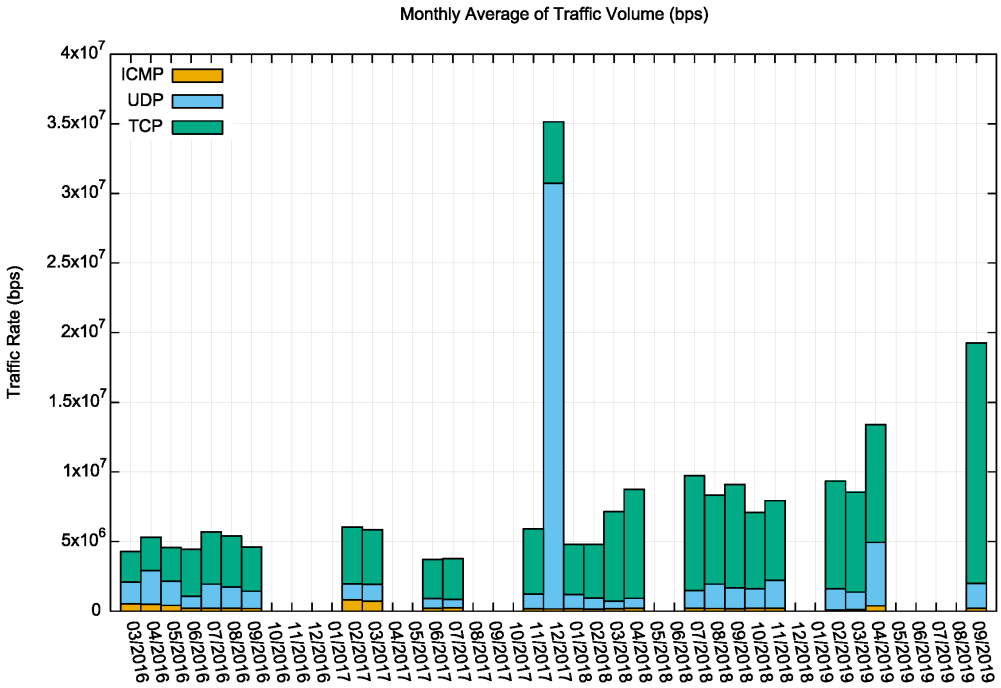

Figure 4 – Dark Traffic – monthly average traffic volume per second

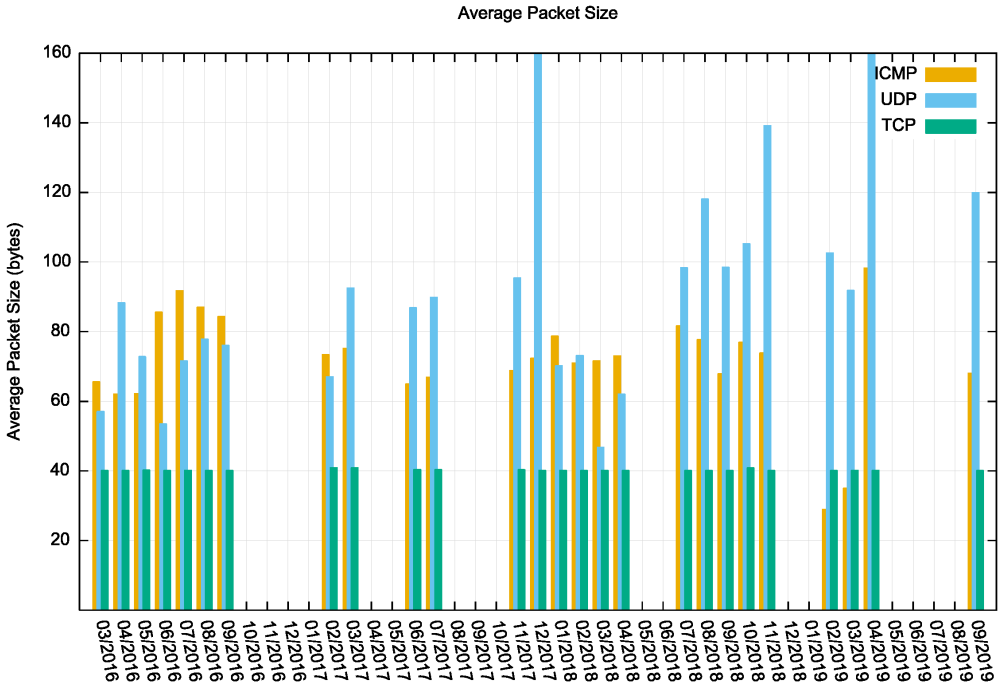

Figure 5 – Dark Traffic – monthly average packet size

The TCP packet rate has increased substantially over this period, with the most recent monthly TCP packet rate being four times the initial rates observed in early 2016. The average TCP packet size remains at 40 bytes, indicating that most of the observed TCP packets contain no payload.

With the exception of two months (December 2017 and April 2019) the UDP traffic rate has been relatively steady, but the average UDP packet size has increased substantially, from an average of 60 to 80 bytes per UDP packet to the most recent monthly average of 120 bytes per packet.

The profile of TCP port numbers has changed over time as well. In 2016 TCP Port 23 (telnet) accounted for more than a third of all TCP packets, whereas by 2019 this port accounted for just 6% of all packets. In 2016 the heaviest used 25 TCP ports accounted for 84% of all TCP packets, yet by 2019 this also dropped to 21% (Figure 6). All TCP port numbers are seen in this traffic collection.

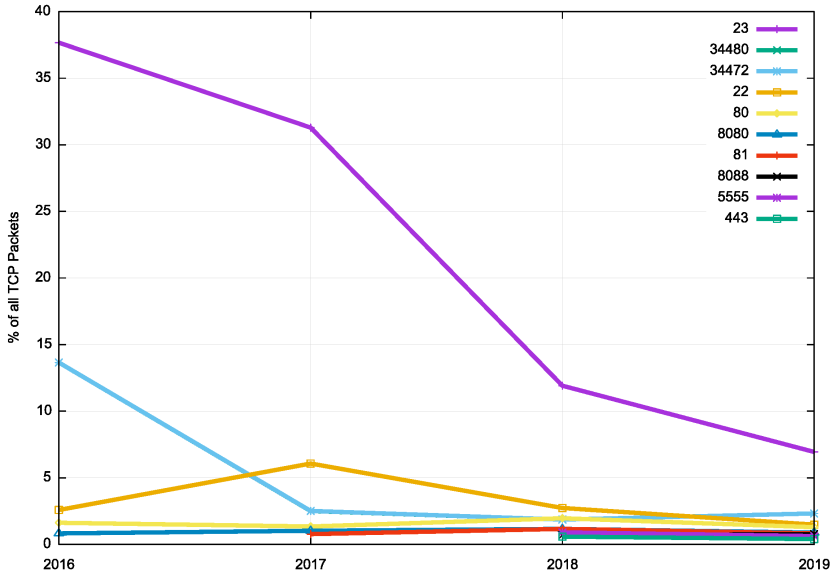

Figure 6 – TCP Port use in a year-by-year basis

UDP packets have a different profile, as the packets are not necessarily part of an protocol-level handshake as we observe with the 3-way TCP handshake. In 2016 port 53413 accounted for one third of all UDP packets. This has declined and in 2019 the most commonly seen port number is port 34480 (Figure 7). There is a visible level of probing for open UDP ports, namely on port 179 (chargen), port 111 (RPC) and 389 (LDAP) which can be used for DDOS amplification attacks if there is a promiscuous server attached to those UDP ports.

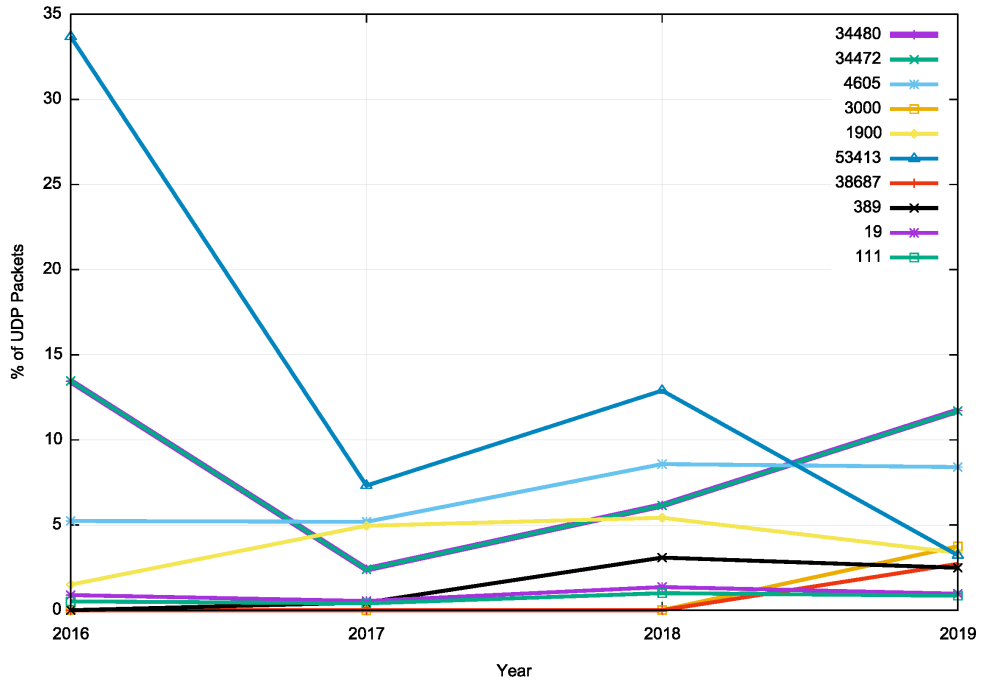

Figure 7 – UDP Port use in a year-by-year basis

The UDP traffic profile is somewhat different and incoming UDP port 3000 traffic accounted for one fifth of all incoming UDP traffic in 2019 (Figure 8)

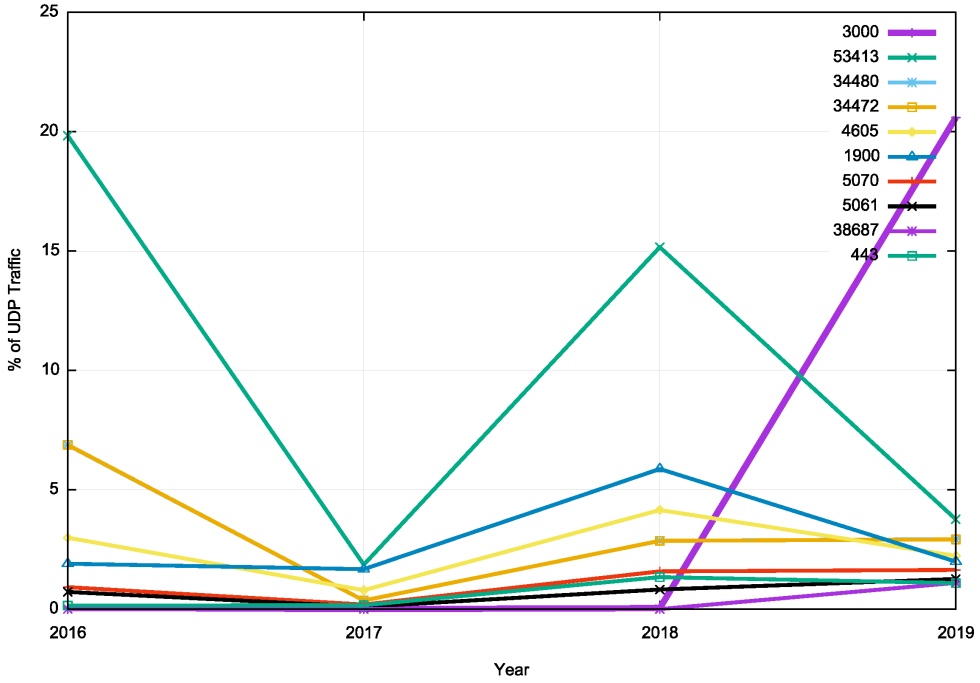

Figure 8 – UDP Port use by volume in a year-by-year basis

Which IP addresses are sending traffic to this announced IPv4 address prefix?

Table 1 shows the top 10 source addresses that sent traffic into this dark prefix across a 6-day period in September 2019:

Rank	IP Address	% of packets	Origin AS
1	45.136.109.82	9.27%	AS49505, SelectTEL, RU
2	81.22.45.115	9.19%	AS49505, SelectTEL, RU
3	81.22.45.49	9.17%	AS49505, SelectTEL, RU
4	81.22.45.48	9.16%	AS49505, SelectTEL, RU
5	81.22.45.51	9.14%	AS49505, SelectTEL, RU
6	93.174.93.195	7.99%	AS202425, INT-NETWORK, SC
7	80.82.78.104	3.57%	AS202425, INT-NETWORK, SC
8	185.40.4.165	3.35%	AS50113, SuperServersDatacentre, RU
9	81.22.45.253	2.25%	AS49505, SelectTEL, RU
10	81.22.45.250	2.12%	AS49505, SelectTEL, RU

Table 1 – Source IP addresses of incoming dark traffic – Top 10 Sources

We can group these source addresses by their origin AS. The result, shown in Table 2 is the top 10 networks that originate the most traffic to this dark prefix.

Rank	AS	% of Packets	Packet Count	AS Name
1	AS49505	62.63%	1,814,509,070	SelectTEL, RU
2	AS202425	13.05%	378,150,214	INT-NETWORK, SC
3	AS50113	3.35%	97,130,636	SuperServersDatacentre, RU
4	AS4134	2.12%	61,325,258	ChinaNET, CN
5	AS14061	1.41%	40,969,795	DigitalOcean, US
6	AS38814	1.10%	31,918,420	Asiamax, VPN SP, HK
7	AS4837	1.04%	30,232,124	China UNICOM, CN
8	AS35582	0.96%	27,866,505	Chistyakov, RU
9	AS3462	0.73%	21,067,910	HINET, TW
10	AS135905	0.64%	18,561,224	VIETNAM PT, VN

Table 2 – Source AS addresses of incoming dark traffic – Top 10 Sources

It is a matter of some concern that almost two thirds of all this unsolicited traffic seen by this dark collector was originated by a collection of IP addresses that are located within a single Russian network. However, it should be noted that because these are incoming packets and the dark traffic collector does not respond in any way, we have no way of knowing if these are real or spoofed source addresses. Nevertheless, it is reasonable to conclude that this intense level of scanning is not an innocent exercise, and the outcome of this scanning can only result in a comprehensive inventory of IPv4 visible end points and the TCP and UDP ports where they are observed to respond.

Are these sources sending traffic to all addresses in the announced prefix, or are some addresses being preferred within this address prefix? We can compare the relative count of incoming packets per address to a model of even traffic distribution to derive a relative intensity index. Addresses with a high intensity value are receiving more traffic than would be the case were the traffic to be evenly distributed across all addresses. Five addresses were observed to receive a disproportionately high level of incoming packets, as shown in Table 3.

The high rate of packets addressed to the address x.x.168.192 appears to be the outcome of an error in converting between host and network byte order, as it is likely that this is leakage of packets addressed to the private network prefix 192.168.x.x, and somehow the IP destination address in these packets have had their byte order transposed such that the packet was addressed to x.x.168.192. A similar story may be behind the relatively high use of the destination address x.x.0.127, being a byte order transposition of the address 127.0.x.x.

Rank	Address	Packet Share	Intensity
1	x.x.159.177	0.41%	271
2	x.x.159.193	0.41%	270
3	x.x.160.49	0.41%	269
4	x.x.160.65	0.41%	269
5	x.x.48.234	0.27%	175
6	x.x.0.0	0.05%	32
7	x.x.0.222	0.04%	23
8	x.x.0.17	0.03%	23
9	x.x.0.18	0.03%	23
10	x.x.0.2	0.03%	23
11	x.x.0.19	0.03%	23
12	x.x.168.192	0.03%	21
13	x.x.2.43	0.03%	20
14	x.x.0.127	0.02%	12
15	x.x.0.122	0.02%	12
16	x.x.32.0	0.02%	10
17	x.x.0.5	0.01%	8
18	x.x.148.111	0.01%	7
19	x.x.5.71	0.01%	6
20	x.x.20.24	0.01%	5

Table 3 – Destination addresses of incoming dark traffic – Top 20 addresses

The packet volumes of the top 80 individual addresses are shown in Figure 9. It is evident that five addresses have received a significant volume of traffic, and a further twenty addresses have received more than the average traffic share. After that the traffic per address appears to even out (Figure 10).

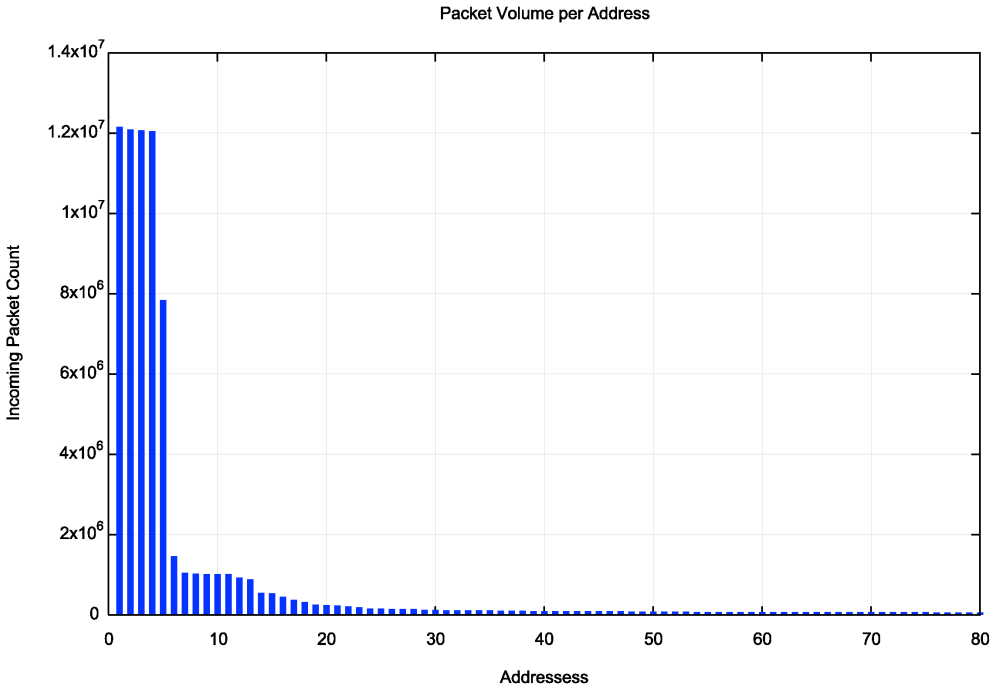

Figure 9 – Incoming Packets per Destination Address- Top 80 addresses

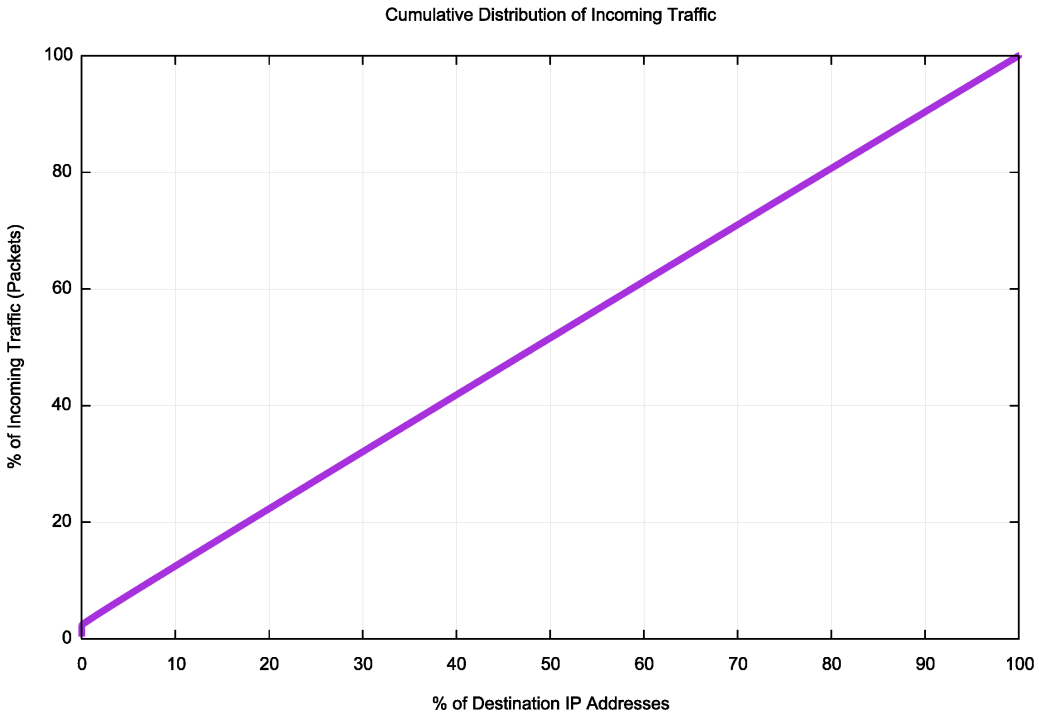

Figure 10 – Incoming Packets per Destination Address

Conclusions

What can we say about our observations of dark traffic in across this period?

When we first looked at dark traffic a decade ago it appeared that the traffic profile was dominated by malware. Infected hosts scanned the IPv4 address space looking for similar vulnerabilities. The scan was dominated by a single TCP port., providing a clear signature of the malware in question. These days it

appears that the traffic profile is now dominated by deliberate scanning for open TCP ports performed by a small number of scanners.

Today's scanning is thorough, in that all TCP ports appear to be tested over time, and it appears that all IPv4 addresses are tested over time. This scanning is not widespread, however. A small number of sources from just a couple of networks appear to account for three-quarters of this scanning activity.

The dark traffic rate has escalated in recent times, and the traffic levels observed in 2019 appear to be some four times greater than what was observed in 2016.

If the Internet was ever a benign place, it is certainly not so today. Any and every device that is exposed to the Internet will be continuously and comprehensively scanned. Any known vulnerability in an exposed host will be inevitably exposed through this concentrated scanning.

Dark traffic is not going away in the IPv4 Internet. The 32-bit address space and the 16-bit port numbers is just too small to drown out the scanners.

There is no comparable evidence of large-scale scanning in IPv6, as the 128-bit the number space is just too large to allow the same form of comprehensive scanning. That does not mean that IPv6 hosts are immune from various forms of exploitation. But it does imply that the means of discovery of vulnerable hosts in IPv6 will necessarily differ from what we observe in IPv4.

Disclaimer

The above views do not necessarily represent the views or positions of the Asia Pacific Network Information Centre.

Author

Geoff Huston B.Sc., M.Sc., is the Chief Scientist at APNIC, the Regional Internet Registry serving the Asia Pacific region.

www.potaroo.net