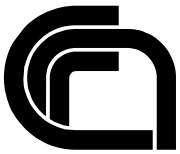

Consiglio Nazionale delle Ricerche

How long does it take before a new Internet node is contacted for the very first time?

F.Lauria

IIT TR-03/2018

Technical Report

Marzo 2018

Istituto di Informatica e Telematica

How long does it take before a new Internet node is contacted for the very first time?

Filippo Lauria - filippo.lauria@iit.cnr.it
Institute of Informatics and Telematics, Italian National Research Council
via G. Moruzzi, 1 - 56124 Pisa, Italy

Abstract
When connecting to the Internet a *new device* (e.g. a computer, a server, a consumer IoT device, etc.) that publicly exposes - i.e. uses a public IPv4 address - any service on any given TCP port (e.g. TELNET on port TCP/23, etc.), the new connected node could be remotely contacted by other network nodes that, both legitimately and maliciously, could attempt to remotely connect to the exposed service. To know if a remote connection attempt comes from a legitimate or a malicious node, it is possible to use *a honeypot:* a network node that acts as the *new device*, but actually works as a malicious nodes bait. The latter allows making the assumption that all the attempts, incoming to the honeypot, comes from malicious nodes. In this case, how long does it take before a malicious node attempts to remotely connect to the honeypot, for the very first time since it has been connected to the Internet? This article gives an answer to the latter question, describing *both network and software environments* used to get the appropriate measurements discussed within this document.

Keywords: malware, honeypots, network security, cybersecurity

Overview
A *new device*[1] connected to the Internet, that uses a public IPv4 address [1] and exposes a *network service*[2], can be remotely contacted by other network nodes that, both legitimately and maliciously, could attempt to remotely connect to the exposed service. Obviously, we are not interested in legitimate connection attempts, but only in the malicious ones. In order to determine if an attempt is legitimate or malicious, we have decided to use a *honeypot*. For the purpose of this document, a honeypot is a public Internet node that acts as the *new device* and therefore exposes a *fake network service* which is unknown and not meant to be used by any real user. For this reason, we can assume that *almost all*[3] these connection attempts come from malicious nodes. Under these hypotheses, how long does it take before a malicious node attempts to connect to the honeypot, for the very first time since it has been connected to the Internet?

[1] e.g. a computer, a server, a consumer IoT device, etc.
[2] e.g. TELNET on port TCP/23 [2], SSH on port TCP/22 [3], etc.
[3] third-party nodes performing TCP port scan, ICMP scan, etc. for diagnostic or statistical purposes cannot be considered malicious.

This document gives an answer to the latter question, describing *both network and software environments* used to get the appropriate measurements presented further below. The discussed measurements have been gathered by using an *ad hoc network environment* which has been set up in the larger network infrastructure of the *Italian National Research Council Research Area of Pisa* and chosen as case study network. We are aware that these measurements may include extra time due to both *network and computational overhead*, but, for the purpose of this article, *the amount of time due to these overheads has been considered negligible*.

We also recognize that the presented measurements are meaningful only under *specific settings*. For example, they could be meaningful in those network scenarios very similar to our case study network scenario. Last but not least, we are aware that these measurements strongly depend upon the variation of the *malicious activity rate* undergoing through the Internet, which could change very quickly over time. The last two assumptions and their related examples must be taken into account when referring to the findings discussed in this document.

Network and software environment

In Figure 1 it is shown how our case study network scenario has been set up. We expected connections incoming from the Internet to our honeypot, which has been configured to listen to connections on port TCP/23 (default for TELNET). We deliberately have chosen this TCP port, because various *consumer IoT devices* have been shipped by their related vendors with a TELNET daemon enabled which, in most of the cases, had weak factory default login credentials [4]. These two facts, combined with the lack of cybersecurity awareness among consumers, represent the main causes of IoT malware spreading.

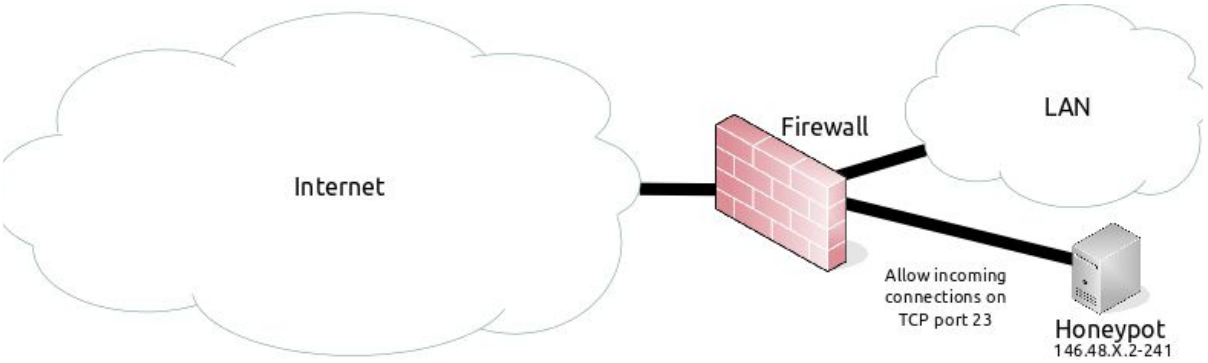

Figure 1: our case study network scenario

Of course, the incoming TCP connections, on the aforementioned TCP port, have been allowed through the firewall. Moreover, the honeypot implemented an *IP hopping mechanism*, which is the capability to automatically switch from an IP address to another. The possible IP addresses have been randomly chosen in the range *146.48.X.2-241,* where *X* has been used to mask the second least significant IPv4 octet.

The behavior of the honeypot is described in Diagram 1. After the honeypot has chosen a random IPv4 address in the aforementioned range, it starts listening to incoming TCP connections from the Internet. These two activities mimic the connection of a new device to the Internet.

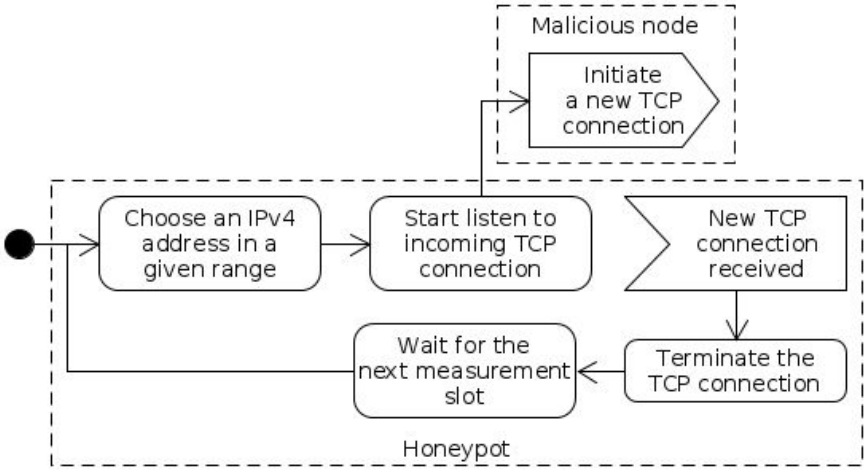

Diagram 1: diagram of the activities performed during the measurement process

When a malicious node initiates a new TCP connection with the honeypot, the only task performed by the honeypot is to immediately terminate that connection[4]. Finally, the honeypot waits for the next *measurement slot*, in order to repeat the process described previously.

In the above paragraph, we have introduced the concept of *measurement slot* (deepened in Figure 2). In our experiment, *we have subdivided a day into 48 measurement slots*, each of which lasted 30 minutes[5].

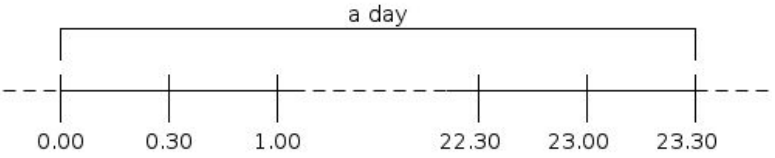

Figure 2: how a day has been divided in 48 measurement slots

In order to determine how long it takes before a malicious node attempts to connect to the honeypot for the very first time since it has been connected to the Internet, we need to introduce few definitions (which have been also illustrated in Figure 3):

- t_{SL}, the exact moment when the honeypot starts listen to incoming TCP connection;
- t_{CR}, the exact moment when a malicious network node initiates a new TCP connection with the honeypot.

[4] the typical low interaction honeypot behaviour

[5] we deliberately have chosen this high amount, in order to:

1. mimic the behaviour of the consumer who connects a new (IoT) device to the Internet;
2. roughly prevent false positive measurements due to TCP port scanning.

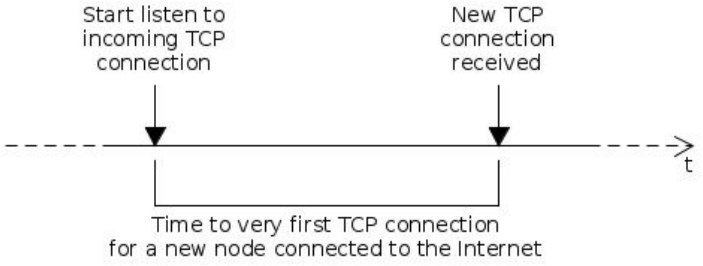

Figure 3: the graphical representation of t_{SL}, t_{CR} and T_{FC}

Having said that, to give an answer to the question introduced above in this document, we need to collect, for each of the IP addresses in the aforementioned range[6], the value of T_{FC}, which is given by the expression: $T_{FC} = t_{CR} - t_{SL}$.

Results

In this section, the gathered data has been reported and discussed. In particular, measurements, gathered by listening to incoming connections on TCP/23 port, during the experiment period going from the January, 16 (~14:30) to January, 21 (~17:02), have been shown in the following tables (Table 1-6). In each table, the gathered measurements have been grouped by day and each record contains the IP address[7] and t_{SL}, t_{CR}, T_{FC} values of that particular measurement.

Date	IP Address	t_{SL}	t_{CR}	T_{FC} (s)
2018-01-16	146.48.X.12	14:30:56	14:35:12	255.68
	146.48.X.65	15:00:12	15:01:43	90.91
	146.48.X.223	15:30:43	15:46:03	919.76
	146.48.X.4	16:00:03	16:01:27	84.67
	146.48.X.66	16:30:27	16:32:56	148.37
	146.48.X.28	17:00:56	17:07:09	373.59
	146.48.X.183	17:30:10	17:32:37	147.41
	146.48.X.96	18:00:37	18:09:09	511.46
	146.48.X.46	18:30:09	18:30:49	40.16
	146.48.X.169	19:00:49	19:04:54	245.39
	146.48.X.113	19:30:54	19:32:47	112.27
	146.48.X.230	20:00:47	20:04:03	196.17
	146.48.X.194	20:30:03	20:42:32	749.07
	146.48.X.77	21:00:32	21:14:34	841.64
	146.48.X.29	21:30:34	21:34:09	215.37
	146.48.X.52	22:00:09	22:25:54	1,544.57
	146.48.X.57	22:30:54	22:41:19	625.50
	146.48.X.181	23:00:20	23:16:21	961.94
	146.48.X.147	23:30:22	23:41:46	684.01

Table 1: data gathered during the first day of the experiment

[6] *146.48.X.2-241,* where *X* has been used to mask the second least significant IPv4 octet.

[7] IP addresses have been chosen randomly, by the honeypot.

Date	IP Address	t_{SL}	t_{CR}	T_{FC} (s)
2018-01-17	146.48.X.217	00:00:46	00:03:28	161.90
	146.48.X.98	00:30:28	00:33:36	188.40
	146.48.X.167	01:00:36	01:32:54	1,937.70
	146.48.X.26	02:00:54	02:00:59	5.25
	146.48.X.137	02:30:59	02:53:12	1,332.87
	146.48.X.87	03:00:12	03:03:33	200.42
	146.48.X.229	03:30:33	03:30:36	3.36
	146.48.X.149	04:00:36	04:09:18	521.26
	146.48.X.27	04:30:18	04:32:23	125.42
	146.48.X.101	05:00:23	05:13:58	815.30
	146.48.X.205	05:30:59	05:35:26	267.62
	146.48.X.2	06:00:26	06:01:53	87.07
	146.48.X.122	06:30:53	06:32:34	100.60
	146.48.X.157	07:00:34	07:07:51	436.74
	146.48.X.90	07:30:51	07:33:16	145.22
	146.48.X.160	08:00:16	08:11:29	672.57
	146.48.X.36	08:30:29	08:50:44	1,215.36
	146.48.X.64	09:00:44	09:04:15	210.36
	146.48.X.74	09:30:15	09:32:57	162.42
	146.48.X.202	10:00:57	10:08:44	466.93
	146.48.X.196	10:30:44	10:31:17	32.54
	146.48.X.131	11:00:17	11:02:55	157.46
	146.48.X.54	11:30:55	11:37:32	397.44
	146.48.X.179	12:00:32	12:02:38	125.77
	146.48.X.7	12:30:38	12:33:02	144.14
	146.48.X.175	13:00:02	13:00:52	49.27
	146.48.X.86	13:30:52	13:53:29	1,356.91
	146.48.X.102	14:00:29	14:02:57	148.34
	146.48.X.236	14:30:57	14:31:48	51.10
	146.48.X.118	15:00:48	15:02:13	84.68
	146.48.X.97	15:30:13	15:43:55	822.09
	146.48.X.177	16:00:55	16:01:50	54.63
	146.48.X.70	16:30:50	16:34:26	216.39
	146.48.X.156	17:00:26	17:06:28	361.53
	146.48.X.119	17:30:28	17:35:09	280.98
	146.48.X.123	18:00:09	18:00:11	2.26
	146.48.X.211	18:30:12	18:42:53	761.87
	146.48.X.184	19:00:54	19:02:45	111.73
	146.48.X.210	19:30:45	19:45:20	874.36
	146.48.X.237	20:00:20	20:22:08	1,308.16
	146.48.X.148	20:30:08	20:30:23	14.93
	146.48.X.151	21:00:23	21:09:51	568.14
	146.48.X.171	21:30:51	21:31:20	28.34
	146.48.X.100	22:00:20	22:02:58	158.37
	146.48.X.192	22:30:58	22:45:49	891.37
	146.48.X.152	23:00:50	23:03:37	167.89
	146.48.X.142	23:30:38	23:34:22	224.79

Table 2: data gathered during the second day of the experiment

Date	IP Address	t_{SL}	t_{CR}	T_{FC} (s)
2018-01-18	146.48.X.121	00:00:22	00:06:47	384.93
	146.48.X.155	00:30:48	00:58:12	1,644.01
	146.48.X.117	01:00:12	01:07:27	435.26
	146.48.X.199	01:30:27	01:30:38	10.93
	146.48.X.111	02:00:38	02:01:27	48.92
	146.48.X.68	02:30:27	02:33:49	201.51
	146.48.X.59	03:00:49	03:03:43	173.98
	146.48.X.75	03:30:43	03:34:19	215.95
	146.48.X.32	04:00:19	04:14:11	831.88
	146.48.X.240	04:30:11	04:32:20	129.81
	146.48.X.39	05:00:21	05:01:04	42.92
	146.48.X.9	05:30:04	05:30:53	49.14
	146.48.X.176	06:00:53	06:13:33	759.69
	146.48.X.33	06:30:33	06:54:36	1,443.76
	146.48.X.165	07:00:36	07:03:49	192.68
	146.48.X.239	07:30:49	07:41:44	655.33
	146.48.X.21	08:00:45	08:02:42	117.74
	146.48.X.145	08:30:42	08:37:24	401.83
	146.48.X.6	09:00:24	09:00:48	23.11
	146.48.X.136	09:30:48	09:58:08	1,640.02
	146.48.X.206	10:00:08	10:13:07	778.95
	146.48.X.25	10:30:07	10:42:35	748.47
	146.48.X.139	11:00:35	11:21:42	1,266.56
	146.48.X.158	11:30:42	11:30:45	2.85
	146.48.X.103	12:00:45	12:03:21	155.71
	146.48.X.186	12:30:21	12:30:50	28.82
	146.48.X.94	13:00:50	13:02:11	81.76
	146.48.X.140	13:30:11	13:39:55	583.02
	146.48.X.5	14:00:55	14:05:17	261.96
	146.48.X.120	14:30:17	14:30:54	37.83
	146.48.X.201	15:00:55	15:05:09	254.84
	146.48.X.193	15:30:09	15:35:19	309.27
	146.48.X.92	16:00:19	16:03:12	172.97
	146.48.X.225	16:30:12	16:32:51	159.19
	146.48.X.212	17:00:51	17:23:57	1,385.64
	146.48.X.144	17:30:57	17:34:20	203.12
	146.48.X.107	18:00:20	18:07:10	409.47
	146.48.X.170	18:30:10	18:34:43	272.80
	146.48.X.159	19:00:43	19:32:49	1,926.71
	146.48.X.164	20:00:49	20:12:04	674.91
	146.48.X.17	20:30:04	20:34:09	244.41
	146.48.X.174	21:00:09	21:07:46	456.82
	146.48.X.15	21:30:46	21:49:44	1,137.93
	146.48.X.60	22:00:44	22:00:56	11.91
	146.48.X.84	22:30:56	22:38:50	473.73
	146.48.X.141	23:00:50	23:07:07	377.84
	146.48.X.105	23:30:08	23:34:57	289.82

Table 3: data gathered during the third day of the experiment

Date	IP Address	t_{SL}	t_{CR}	T_{FC} (s)
2018-01-19	146.48.X.208	00:00:57	00:06:29	331.26
	146.48.X.154	00:30:29	00:38:09	459.88
	146.48.X.207	01:00:09	01:03:32	203.73
	146.48.X.241	01:30:32	01:39:12	519.75
	146.48.X.219	02:00:12	02:01:27	74.62
	146.48.X.89	02:30:27	02:32:28	120.63
	146.48.X.106	03:00:28	03:36:23	2,155.70
	146.48.X.67	04:00:24	04:08:36	492.39
	146.48.X.85	04:30:36	04:57:54	1,638.43
	146.48.X.180	05:00:54	05:27:56	1,621.10
	146.48.X.49	05:30:56	05:37:12	376.82
	146.48.X.19	06:00:13	06:20:04	1,191.38
	146.48.X.182	06:30:04	06:43:20	796.28
	146.48.X.220	07:00:20	07:05:43	322.44
	146.48.X.47	07:30:43	07:32:28	105.24
	146.48.X.61	08:00:28	08:07:05	396.78
	146.48.X.37	08:30:05	08:30:37	31.84
	146.48.X.135	09:00:37	09:04:07	209.82
	146.48.X.11	09:30:07	09:34:20	253.26
	146.48.X.110	10:00:20	10:01:13	52.41
	146.48.X.23	10:30:13	10:35:28	314.85
	146.48.X.35	11:00:28	11:23:42	1,394.44
	146.48.X.50	11:30:42	11:30:48	6.17
	146.48.X.198	12:00:48	12:11:08	619.80
	146.48.X.127	12:30:08	12:39:08	539.93
	146.48.X.78	13:00:08	13:29:46	1,777.37
	146.48.X.63	13:30:46	13:32:15	89.67
	146.48.X.18	14:00:15	14:17:17	1,021.69
	146.48.X.232	14:30:17	14:34:46	268.65
	146.48.X.213	15:00:46	15:07:56	430.06
	146.48.X.116	15:30:56	15:37:10	373.93
	146.48.X.41	16:00:10	16:00:20	9.67
	146.48.X.72	16:30:20	16:35:02	282.04
	146.48.X.200	17:00:02	17:00:57	55.03
	146.48.X.38	17:30:57	17:35:48	291.28
	146.48.X.228	18:00:49	18:06:15	326.89
	146.48.X.234	18:30:16	18:39:15	539.32
	146.48.X.14	19:00:15	19:13:13	778.28
	146.48.X.73	19:30:13	19:32:12	118.46
	146.48.X.13	20:00:12	20:12:26	733.86
	146.48.X.69	20:30:26	20:33:05	159.71
	146.48.X.215	21:00:06	21:00:36	30.57
	146.48.X.168	21:30:36	21:31:30	54.06
	146.48.X.88	22:00:30	22:08:20	469.99
	146.48.X.130	22:30:20	22:32:57	156.61
	146.48.X.20	23:00:57	23:10:58	600.71
	146.48.X.195	23:30:58	23:32:18	80.48

Table 4: data gathered during the fourth day of the experiment

Date	IP Address	t_{SL}	t_{CR}	T_{FC} (s)
2018-01-20	146.48.X.91	00:00:19	00:05:41	322.97
	146.48.X.172	00:30:42	00:34:20	218.25
	146.48.X.153	01:00:20	01:06:24	364.59
	146.48.X.218	01:30:24	01:40:23	598.97
	146.48.X.185	02:00:24	02:20:27	1,203.73
	146.48.X.124	02:30:27	02:30:34	6.77
	146.48.X.178	03:00:34	03:21:19	1,245.16
	146.48.X.99	03:30:20	03:41:47	687.24
	146.48.X.128	04:00:47	04:03:05	137.95
	146.48.X.214	04:30:05	04:48:42	1,116.97
	146.48.X.222	05:00:42	05:02:37	114.88
	146.48.X.233	05:30:37	05:32:30	113.20
	146.48.X.203	06:00:30	06:18:16	1,065.91
	146.48.X.163	06:30:16	06:39:32	555.51
	146.48.X.138	07:00:32	07:19:22	1,130.22
	146.48.X.45	07:30:22	07:37:23	420.56
	146.48.X.162	08:00:23	08:11:19	656.65
	146.48.X.129	08:30:19	08:42:16	716.40
	146.48.X.227	09:00:16	09:04:19	243.37
	146.48.X.216	09:30:19	09:39:01	522.17
	146.48.X.104	10:00:02	10:03:06	184.40
	146.48.X.197	10:30:06	10:40:19	613.24
	146.48.X.235	11:00:19	11:01:07	47.15
	146.48.X.79	11:30:07	11:42:22	735.47
	146.48.X.108	12:00:22	12:15:07	885.03
	146.48.X.143	12:30:07	12:35:44	336.93
	146.48.X.24	13:00:44	13:05:43	298.16
	146.48.X.238	13:30:43	13:31:59	76.00
	146.48.X.76	14:00:59	14:04:35	216.53
	146.48.X.40	14:30:35	14:33:18	162.37
	146.48.X.231	15:00:18	15:07:07	409.07
	146.48.X.114	15:30:07	15:33:24	197.08
	146.48.X.93	16:00:24	16:01:00	36.20
	146.48.X.16	16:30:00	16:40:00	599.32
	146.48.X.3	17:00:00	17:01:53	113.10
	146.48.X.166	17:30:53	17:31:09	16.09
	146.48.X.53	18:00:09	18:07:25	436.28
	146.48.X.42	18:30:26	18:38:53	507.15
	146.48.X.81	19:00:53	19:13:00	727.44
	146.48.X.191	19:30:00	19:35:51	351.05
	146.48.X.126	20:00:52	20:18:01	1,029.71
	146.48.X.58	20:30:01	20:46:14	972.25
	146.48.X.30	21:00:14	21:07:03	409.46
	146.48.X.224	21:30:03	21:34:53	289.99
	146.48.X.56	22:00:53	22:07:42	408.81
	146.48.X.188	22:30:42	22:34:03	200.89
	146.48.X.161	23:00:03	23:07:56	472.53
	146.48.X.82	23:30:56	23:44:07	791.51

Table 5: data gathered during the fifth day of the experiment

Date	IP Address	t_{SL}	t_{CR}	T_{FC} (s)
2018-01-21	146.48.X.43	00:00:07	00:01:19	71.55
	146.48.X.132	00:30:19	00:36:19	359.64
	146.48.X.71	01:00:19	01:27:45	1,646.89
	146.48.X.55	01:30:46	01:39:40	534.23
	146.48.X.150	02:00:40	02:03:09	149.35
	146.48.X.51	02:30:09	02:42:12	722.49
	146.48.X.95	03:00:12	03:00:21	8.96
	146.48.X.80	03:30:21	03:34:41	260.00
	146.48.X.133	04:00:41	04:08:45	483.78
	146.48.X.115	04:30:45	04:31:55	70.32
	146.48.X.109	05:00:55	05:33:05	1,929.97
	146.48.X.34	06:00:06	06:38:26	2,300.59
	146.48.X.22	07:00:26	07:05:20	293.79
	146.48.X.112	07:30:20	07:30:23	2.75
	146.48.X.187	08:00:23	08:01:36	73.47
	146.48.X.204	08:30:36	08:33:45	188.75
	146.48.X.134	09:00:45	09:14:50	844.50
	146.48.X.48	09:30:50	09:32:57	127.03
	146.48.X.190	10:00:57	10:06:54	356.51
	146.48.X.8	10:30:54	10:33:38	164.45
	146.48.X.83	11:00:38	11:06:19	340.55
	146.48.X.173	11:30:19	11:36:42	383.25
	146.48.X.10	12:00:42	12:00:54	11.70
	146.48.X.62	12:30:54	12:38:05	430.98
	146.48.X.146	13:00:05	13:06:00	355.38
	146.48.X.125	13:30:01	13:44:59	897.97
	146.48.X.189	14:00:59	14:20:30	1,171.28
	146.48.X.209	14:30:30	14:41:23	653.07
	146.48.X.226	15:00:23	15:03:55	211.50
	146.48.X.31	15:30:55	15:33:17	142.79
	146.48.X.44	16:00:18	16:06:02	344.58
	146.48.X.221	16:30:02	17:02:04	1,922.04

Table 6: data gathered during the sixth day of the experiment

As it could be noticed in the above tables (Table 1-6), in some cases, it has been necessary to use two measurement slots per IP address, in order to gather a single incoming connection (i.e. there are some records where $T_{FC} > 1800\ seconds$, where 1800 seconds is the total length of a measurement slot).

Conclusion

In this section, we want to explain and summarize the results presented in the previous section. In particular, Chart 1 shows the number of connections detected per *slices*, where a *slice* is obtained by splitting a single measurement slot into many smaller time slices, each of which lasts 5 minutes.

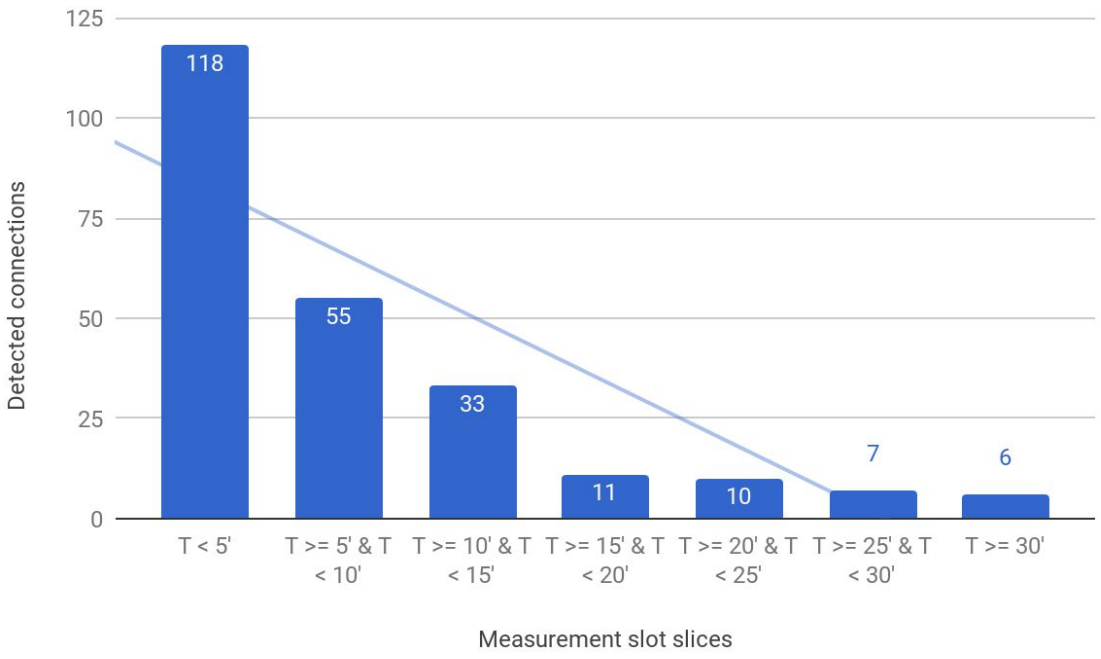

Chart 1: detected connections per measurement slot slices

Furthermore, the smallest bar of Chart 1 considers those connection attempts, where T_{FC} has exceeded the total length of a single measurement slot. Another fact that could be noticed is that most of the measured T_{FC} values, i.e. 206 out of 240 (~86%), have been detected in the first 15 minutes of a measurement slot. For this reason, it has been considered appropriate to compute the average $\overline{T_{FC}}$, considering only the aforementioned 206 measurements, and it turned out to be:

$$\overline{T_{FC}} \simeq 5' 11''.$$

In conclusion, considering all the simplifying assumptions introduced in the previous sections, we can assume that, for example, a new consumer IoT device placed for the very first time in network scenarios similar to the one described above, connected to the Internet using a public IPv4 address, listening to incoming remote connections on port TCP/23, on average[8] could be contacted by malicious nodes in little more than 5 minutes.

[8] as mentioned in the previous paragraph, the average value has been calculated on 206 T_{FC} values out of 240

References

[1] R. Housley, J. Curran, G. Huston, D. Conrad (August 2013). RFC 7020 - The Internet Numbers Registry System. www.ietf.org/rfc/rfc7020.txt

[2] J. Postel, J. Reynolds (May 1983). RFC 854 - Telnet Protocol Specification. www.ietf.org/rfc/rfc854.txt

[3] T. Ylonen, C. Lonvick (January 2006). RFC 4253 - The Secure Shell (SSH) Transport Layer Protocol. www.ietf.org/rfc/rfc4253.txt

[4] B. Krebs (October 2016). - Who Makes the IoT Things Under Attack? www.krebsonsecurity.com/2016/10/who-makes-the-iot-things-under-attack

[5] Types of Honeypots. wikipedia.org/wiki/Honeypot_(computing)#Types