Introduction

In this study we strive to assess the issues and the economics of IOT devices security. With recourse to the data of IOT devices that are compromised and used for attacks we try to extract some useful metrics for decision makers. Dataset that we are working on is put together using the collected data from a honeypot that mimics IOT devices by running a telnet server over port 23. In addition to the chain of commands, a reverse scan of the attackers has been performed and the result exists in a separate dataset.

IOT devices are more under blind attacks rather than targeted attacks. After attacking a number of IOT devices, the attacker creates a network of these victims. These networks are known as botnets and our analysis shows that the attacks are mainly from a botnet called WOPBOT. Our judgement is based on the type of credentials used for connection and the malware installed after infection. The credentials suggest an attack involving and mainly targeting CCTV H.264 DVR. Apart from attacks coming from aforementioned botnet, we found other attacks that just simply reports a message like "Attacker IP", "Attacker Port".

We aim to offer valuable metrics that can help security decision makers when they are deciding on the budget for security. Our analysis in this study is mainly valuable for IOT device vendors. Our methodology in this study is identification of the main security issue, presentation of ideal security metrics including but not limited to existing metrics and a definition of metrics that can be extracted from the dataset. Finally this document presents an early result of dataset analysis based on the chosen metrics.

What security issue does the data speak to?

In 2008 the number of connected things to the internet exceeded the number of people on the earth. A majority of these things are Internet Of Things (IOT) devices. The idea behind IOT is to let devices talk to people, applications and each other. Interestingly enough the ability of IOT devices to talk to each other makes them interesting for attackers. Recently the largest Distributed Denial Of Service (DDOS) attack has been launched involving 152000 IOT devices[1]. The number of IOT devices on the one hand and their lack of security on the other hand are the main motivations of the attackers to target IOT devices.

IOT devices from the technology perspective can be divided into two categories. A vast number of IOT devices are built on existing technologies. The main economic drive for such choice is time to market and cost. Existing technologies never ensure security even when they have security controls unless customized for their specific purpose. For instance a secure version of Linux operating system for Desktop users does not guarantee security for IOT devices since IOT does not need most of the desktop user capabilities. Provided this, IOT devices built with existing technologies are not hardened for their specific environment by default. The second category contain devices that are built using proprietary technologies. The main driver behind such choices is monopoly and the absence of standards. Proprietary technologies also have their own problem from a security perspective. First vendors tend to choose security through obscurity although it has been proven that unknowingness of the vendor technology does not make the device secure. Second a proprietary technology is under less levels of scrutiny by end users and it is more probable that flaws remain unseen.

Regardless of the device development, the number of compromised devices suggest lack of built-in security and corporate defense in the devices. In other words, the devices would have been secure if either there were enough built-in security controls or there were defensive technologies like firewalls. From the device scan data a total of 63 unique IPs were for firewall devices, all the same product, called SonicWall. The SonicWall is a product from Dell and looking at the banner response from the device all the messages sent on port 80 for HTTP were let through, since it gave a “200 OK” response. This indicates that even with a strong firewall like SonicWall, the IOT devices were compromised with a simple password bruteforce. In other words without strong security policies and practices, further security controls will not be effective. Figure 2 shows that a tangible number of devices are running secure services like SSL or SSH. Nevertheless if the default password is not changed or there is a hardcoded password, SSH or SSL cannot ensure security. Although defense in depth is a well-known security practice among security professionals, IOT devices do not follow it. As a result any failure in protecting the device or any vulnerability leads to the maximum damage. For instance, in most of the cases the default user has the highest privileges. This means if the user does not change the default password the attacker has full access and can run any commands. In addition, the underlying operating system of IOT devices is not hardened and they support all the commands and functionalities a normal computer does. As an instance we found commands to install Mirai malware, which is an ELF trojan backdoor especially designed for embedded Linux operating systems e.g. busybox, on the IOT device that runs an email spammer. The commands have a format like this:

```
‘busybox tftp‘ -r [MalwareFile] -g [IPsource]

‘busybox tftp‘ -g -l ‘dvrHelper’ -r [MalwareFile] [IPsource]
```

With a defense in depth practice in place, tftp command usage could have been limited just to the DNS address of the vendor. In that case, even after the compromise of the device the attacker cannot install malicious malware on the host.

More elaborate analysis of the dataset shows that the nature of the attacks is a dictionary attack. The attacker (and bots in the infection stage) uses a list of default, hardcoded and simple passwords for telnet. If the IOT device is vulnerable, the attacker tries delivering the shellcode that is usually installing a binary that makes the IOT device a bot. The success of this attack is the result of the following security weaknesses:

- The authentication protocol is weak
- The IOT device does not have a strong password
- The IOT device is not protected by a firewall
- The IOT device user access control is poorly configured
- The operating system is not hardened

In order to avoid the aforementioned security issues vendors should follow a security by design practice. We believe in order to follow such practice, vendors must first address default and hardcoded passwords as the number one issue. Shodan[2] service as a scanner of vulnerable IOT devices, exploiting default password, issue has been around for years. Yet vendors still rely on the same vulnerable concept of embedding default password. Default password vulnerability needs user awareness and action in order to be fixed. As IOT devices are becoming more widespread, relying on user intelligence and skill in this regard seems to be a pitfall. In the rest of this paper we focus on weak password policy issue.

What would be the ideal metrics for security decision makers?

The ideal metrics must shed light on the perfect amount of economical investment for vendors required to solve the aforementioned security issue. For instance regarding the default password problem, the vendor is not directly harmed by the problem. Nevertheless indirectly, the IOT infection may lead to reduction of IOT speed, maintenance problems, and finally to customer dissatisfaction. A combination of metrics including percentage of total vendor device infections because of password problem, average percentage of performance reduction after the infection and percentage of dissatisfied customers because of such infection and performance reduction is ideal for a vendor. Other metrics should allow vendors to compare cost and benefits of different security controls. For instance the vendor may want to pinpoint the real problem with default passwords. Metrics to present the percentage of devices that are compromised during the installation process (before fully configuring the device and changing the password) or as a result of not changing the default password at all or even changing the password but using a weak password can help the vendor to pinpoint the real problem. Finally another interesting metric would be showing the number of bruteforce attempts per device and probability of success in case the default or simple password is not used.

What are the metrics that exist in practice?

Table 1 reports the security metrics that are used in practice based on their metric category:

Incident metrics	Intrusion attempts
	Spam detected/filtered
	Unauthorized website access (content filtering)
	Invalid logins (failed username)
	Viruses detected on websites
	Unauthorized access attempts (internal)
	Admin violations (unauthorized changes)
	Intrusion successes
	Unauthorized information disclosures
	Spam not detected (missed)
	Spam false positives

	Other
Control metrics	Firewall used
	Antivirus used
	Ids used
	Last update of antivirus software
	patching frequency
	insufficient authentication
Vulnerability metrics	unpatched devices
	insecure firmware
	devices with default credentials
	devices with hardcoded credentials
	devices with no authentication
	number of exposed devices

Table 1: metrics used in practice

[1] https://thehackernews.com/2016/09/ddos-attack-iot.html

[2] https://www.shodan.io/

A definition of the metrics you can design from the dataset

It's important to note that our dataset is divided into two different but related data files. One part of data sheds light on incident metrics including the details of attacks over time. For instance, *number of intrusions*, *Invalid logins and Intrusion successes* can be easily extracted from one of the datasets. The first one can be measured by counting the average number of telnet trials per destination IP. The second one can be determined by measuring the number of telnet trials without any commands after the password argument. The third can be extracted by counting the number of records with commands after the password argument of the telnet. These data are valuable to analyze the frequency of attacks against devices and the probability of success. That being said, this data cannot shed light on the real world because it is taken from the honeypot. For instance our analysis shows that 31% of the attacks use root:xc3511 pair as credentials. This can suggest two conclusions: 1) the attackers use this pair more frequently for some reasons 2) the device that has this default password is broadly used in the market. Unfortunately, such conclusions can not be deducted because the dataset is the result of logs from honeypots. This means such frequency could be simply because of honeypot owners tendency to set such passwords for their device. This credential belongs to a CCTV model and our analysis of the other file cannot prove wide usage of this CCTV model in the wild (Figure 1). Again our inability to determine this can be originated from the dataset deficiency in determining the device model using the banner data.

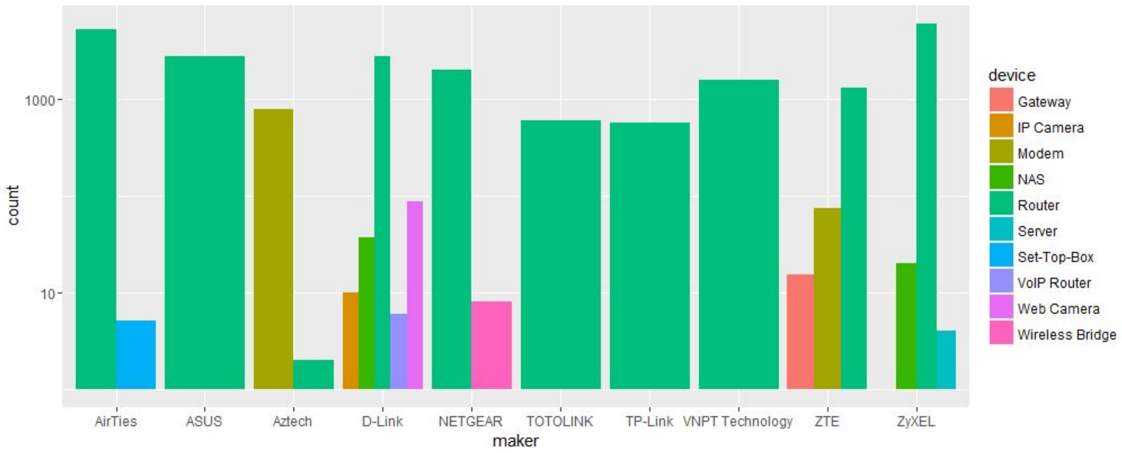

Figure 1: Vendor frequencies and device frequencies based on the back scans

From the control metrics category the only possibility is the extraction of used firewalls by examining iot_device_2m dataset. This can be determined from the banner response in some cases. This metric helps vendors to calculate the probability of infection under the condition of having a security control like firewall while having default credential[1]. Another metric that can be extracted from our dataset is the number of compromised devices with secure protocols from the population of all the compromised devices that contacted the underlying honeypots. This metric could be very useful if it was incorporated with the total number of devices that have secure protocols and the total number of devices that are compromised (we discuss the importance of these data in Evaluation section). With such data, decision

[1] we do not formulate how to calculate such conditional probability for this metric but the general idea based on Bayes theorem used for other metrics can be applied for this metric too.

makers could easily decide if incorporating security protocols with default credentials help at all. If not, instead vendors could invest the money used for security protocols in resolving default password problems. That being said, using our dataset vendors have to make assumptions on the total number of devices having security protocols and default passwords because our data represents a sample of the compromised devices and total population.

From the vulnerability metrics we can just partially report numbers because again we do not have the total number of device population. Based on the bruteforce attack the compromised devices performed on the honeypots, we can conclude that the devices that exist in iot_device_2M file either have default passwords or no authentication. We can categorize these devices based on their vendor and plot their growth chart. However, since we do not have the number of total devices with default passwords, no conclusion about the probability of compromise with default password can be made solely on the number of devices. In other words the dataset does not represent the number of devices with default credentials or those without authentication but the number of compromised devices with default credentials or no authentication. That being said we can report the number of compromised devices that had no authentication and the total number of devices that had default passwords. From vendor point of view this data can be valuable because they already have their total number of product sales and by approximating the probability of being targeted by a botnet they can deduct the number of users that do not change their passwords.

Apart from the aforementioned existing metrics that can be extracted from the dataset we propose some other metrics. The number of compromised devices based on their vendor and their type help vendors to prioritize their actions. Since they have their total number of category sales, by using this number and approximating the probability of being targeted by a botnet they can see if a type of devices needs more security. In addition to the previous metric, we propose a metric to show all the credentials used in the attacks and their frequency of usage. This metric has two benefits for the vendors. First they can see if their default credential is in danger. It is common that a credential is used by more than one vendor. For instance our analysis shows that “admin , password” is used by Cisco, Netgear and 3COM (and probably more vendors). Another example is “root, root” pair that is used by Axis, Ambit, CTC, Avaya, Microplex, and Milan vendors. On 1st of July the latter has been used almost in 39% of the attacks to the honeypot. This can again be attributed to the tendency of honeypot owners to set this credential. Nevertheless if a credential is more used by vendors, it is more used in attacks since it already exists in default password databases and a botnet developer embeds such password in its malware. Moreover, the trend chart of password usages over time can reveal information on the botnet behavior. For instance if we see the frequency of usage of passwords over time does not change frequently we may conclude that the bot new strains does not update their password dictionary and vendors can exploit this behavior for quick fixes like changing their default password to something that is not used by the botnet.

Evaluation

Based on the arguments explained in previous sections we continued our analysis based on the following metrics:

1. Number of compromised devices based on their protocols

2. Number of compromised devices

    A. Total over time

    B. partitioned based on vendors

4. Frequency of compromised devices with authentication
5. Frequency of credentials used in the attacks

Number of compromised devices based on their protocols

This metric as explained earlier helps vendors to see the number of compromised devices that have secure protocols and default passwords. Implementing security protocols for the vendor has cost and the goal to undergo such cost is security. In order to evaluate the effectiveness of implementing a security protocol while keeping default credential mechanism we define:

P(C) = Probability of compromise

P(S) = probability of having a security protocol while keeping a default password mechanism

Using P(C) and P(S) we can say effectiveness of implementing a security protocol while keeping default credential depends on the value of P(C| S). That is *probability of being hacked under the condition of having security protocol and also a default password.* Using Bayes formula we can deduct that:

$$P(C|\,S) = \frac{P(C)P(S|C)}{P(S)}$$

$P(S|C)$ is what we can calculate with the metric we defined here, which is the percentage of devices with security protocols from the total number of devices that are compromised. Security protocols here are ssl, X509, and ssh. It goes unsaid that vendors can tune the calculation by breaking the query to their “maker’ value and device type (Figure 1). Figure 2 reports the number of compromised devices with security protocols and unsecure protocol.s

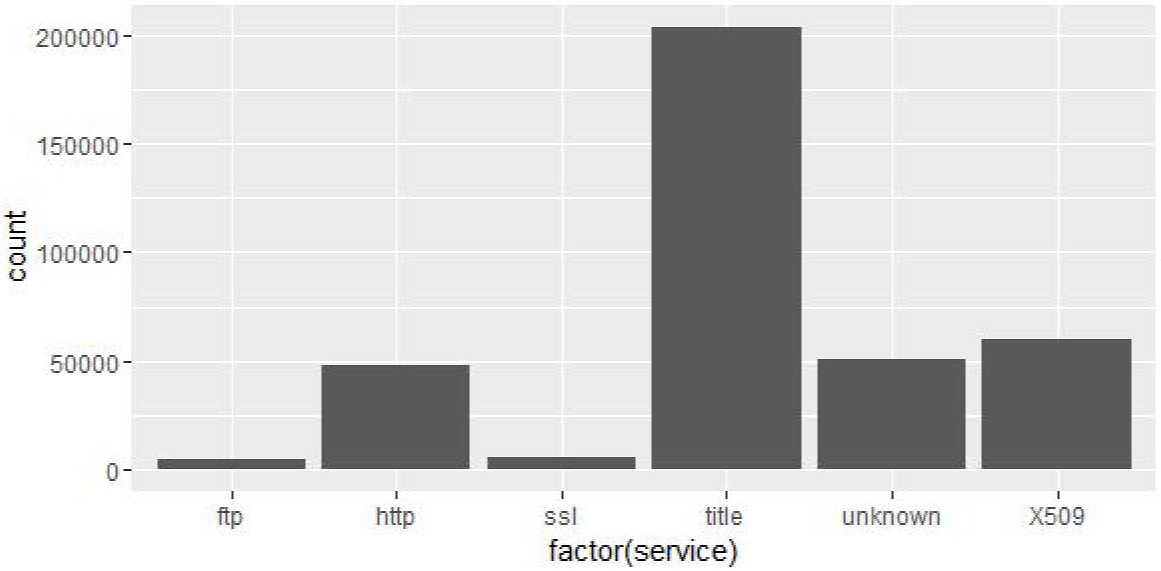

Figure 2: # compromised based on service

Number of compromised devices

In order to understand the importance of this metric, let us assume vendors are interested in knowing the *probability of being hacked under the condition of having default password*. In order to clarify the concept we define:

P(D): probability of having default password

P(C): probability of being compromised(hacked)

P(C|D): probability of being hacked under the condition of having default password

Using Bayes theorem we have:

$$P(C|\,D) \ = \frac{P(C)P(D|C)}{P(D)}$$

From above formula, $P(D|C)$ can be computed by counting the number of devices with authentication (calculated in the next section) in iot_device_2m dataset. Since the attack is password brutforce, either the compromised device had a default password or it did not need authentication. Using the numbers we present in the course of next metric, how one can compute $P(D|C)$ by simply computing devices with authentication in place on total compromises and subtracting the result from 1. Vendors can compute P(D) by simply sampling a small population of their customers or doing survey (for high precision, a whole scan of the internet is required!). P(C) on the other hand needs the total number of compromises and this is where the metric we defined here becomes valuable. To compute the divisor of P(C), the vendor can assume the total population as the total number of their sold products in case total number of compromise does not seem Epsilon. In order to cross validate the result, vendors can use the data published by security companies like Symantec to compute total number of compromises per total number of vendor devices in the market.

Total over time

In addition to the previous argument a scatter plot of total number of compromised devices is important to understand the difference in the number of infections per vendors as we see in the next subsection. In order to calculate the numbers and plot the graph we grouped iot_device_2m data based on scan date and then counted the number of unique destination ip addresses, the result of which can be seen in Figure 3.

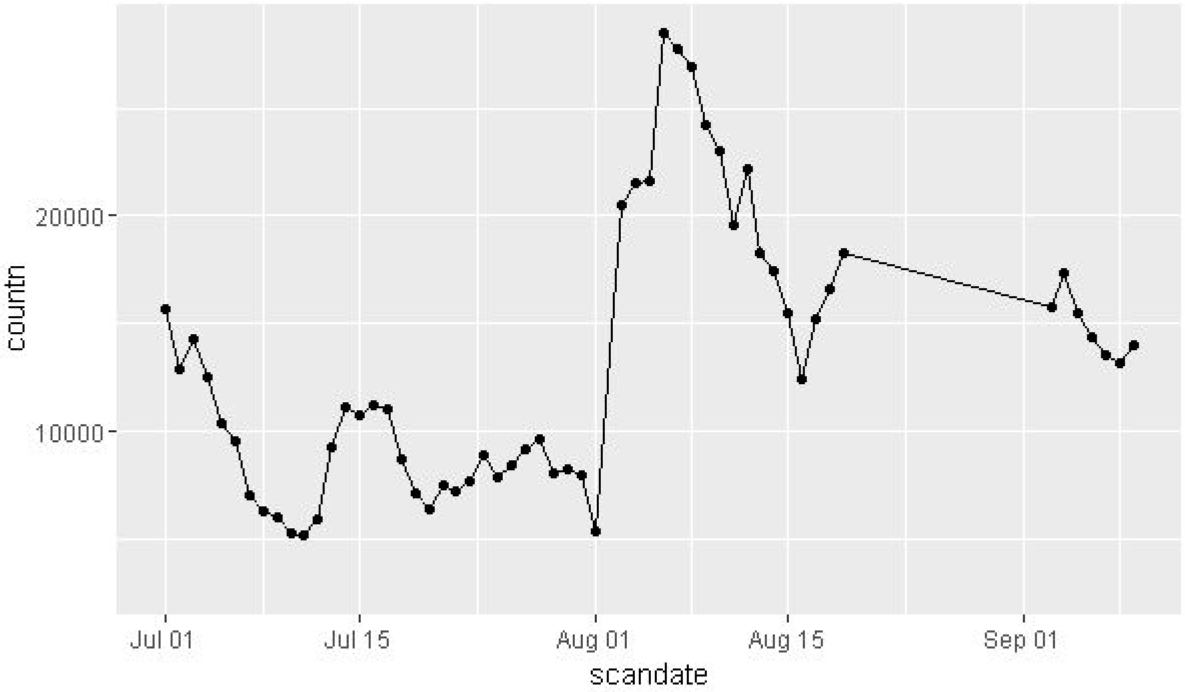

Figure 3: # of unique compromised machines based on date

Partitioned based on vendors

We were curious if a specific vendor devices compromise also followed the same pattern. This is important to understand if the distribution of compromised devices is the same. If distribution is uniform the approximation of probabilities discussed in the previous metrics would be much easier. Based on such arguments we plotted the scatter plot of compromised devices for the most 25 frequent vendors in the dataset. Figure 4 shows the graph.

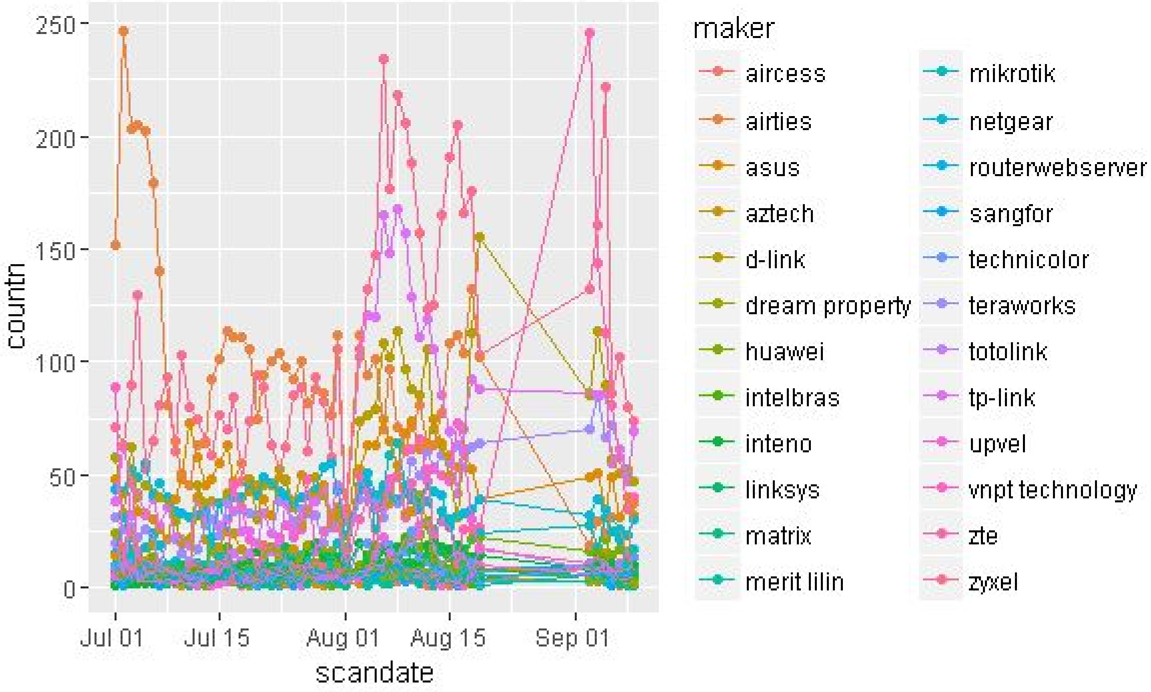

Figure 4: # of unique compromised machines based on date and vendor

At the first glance it seems that the same pattern that we observed on the number of total compromise can be observed here. Because of the density we tried to plot the graph for the top vendors: zyxel, TP-Link and D-Link. The result is shown in figure 5.

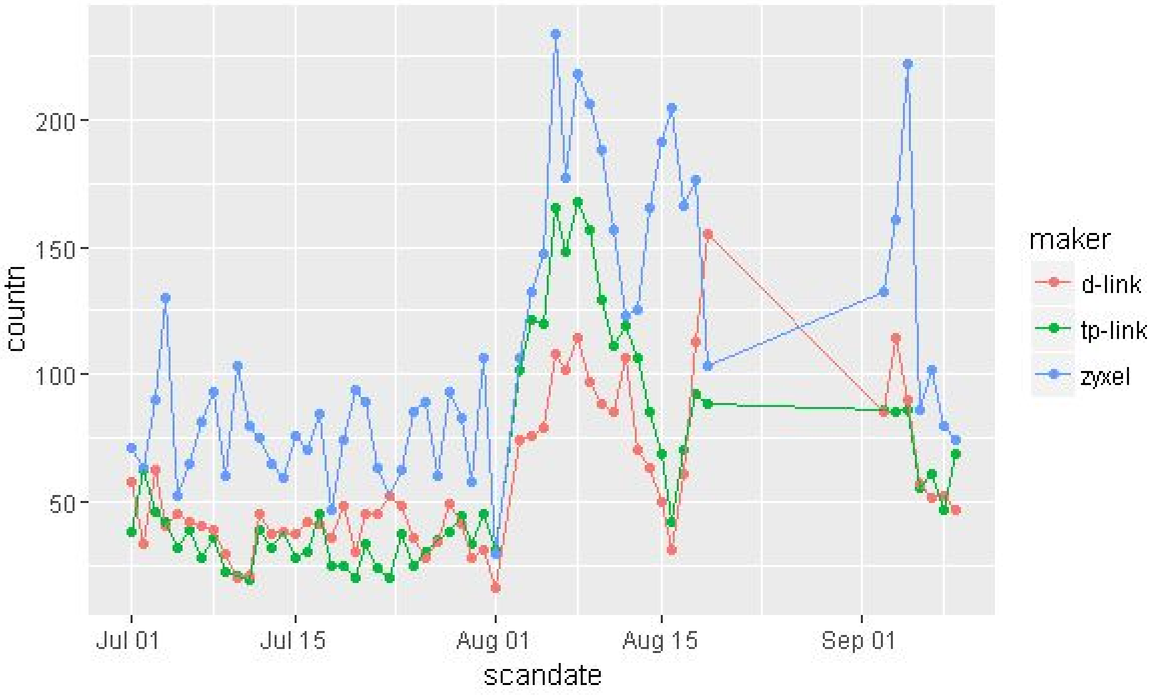

Figure 5: # of unique compromised machines of top 3 vendors based on date

As it can be seen in Figure 5 again the same pattern holds and this proves that the compromise pattern is independent of the vendor. In other words no vendor from our dataset is significantly doing better or worse in terms of default password. This means the difference in the frequency of infections for a vendor can be attributed to other reasons such as higher number of a vendor product sale. Moreover we can conclude that during these two months no security update for a specific vendor to address default password has been released or the release had not been effective. Otherwise the pattern of infection for a vendor would have been different.

Frequency of compromised devices with authentication

The devices are categorized into separate categories, those requiring credentials to be provided and those without to access some service. Inside the ip_device_2M excel file, the scans show which ports and which services were used as well as the banner returned from the scanned devices. Using the services and the banner the categories are broken down into devices that have authentication requirements on four services: FTP, HTTP, Title, and an unknown service. The unknown services were using port 23, which means the usage of telnet. Figure 6 shows the number of unique machines scanned that required authentication to use at least one of the services mentioned. Figure 7 shows the number of unique devices scanned that did not require any sort of authentication to use the services. In Figure 8, there is a comparison between the percentage of services running on the indicated machines that require some authentication. To place each device into a category the banner returned from the scan is used since it contained information as to whether someone or something is unauthorized when they try to connect, if they have to log in first, or if they are taken straight to the resource without needing any credentials. Vendors can use this data to look at which services were used the most to connect to the device without any authentication requirements. They can decide whether to disable some service by default because it may not of any use and having the service open when it is not used is a risk.

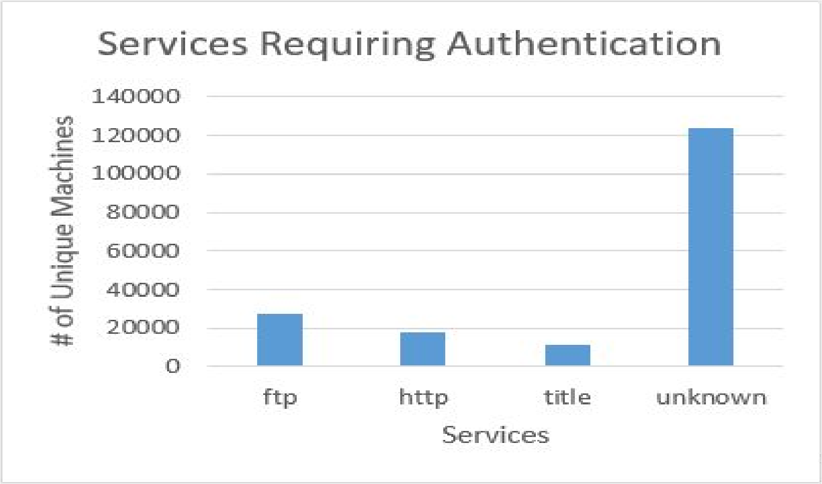

Figure 6: # of unique machines requiring authentication for different services

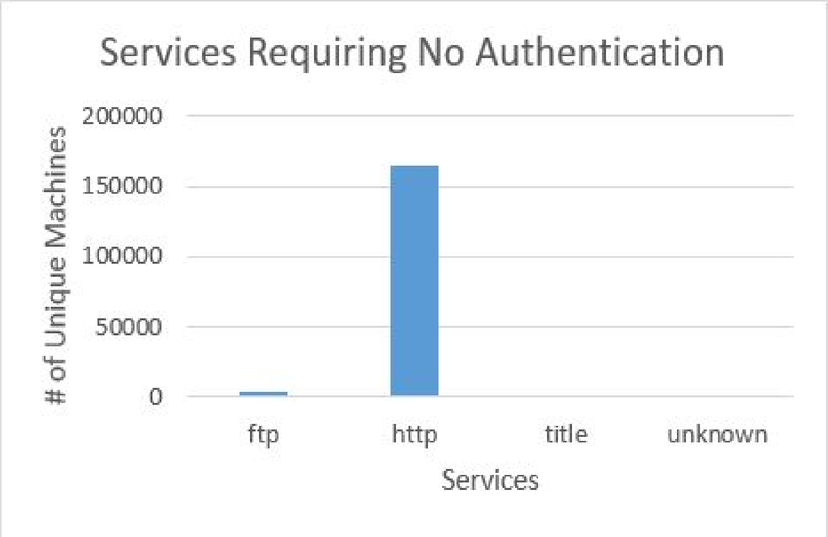

Figure 7: # of unique machines requiring no authentication for different services

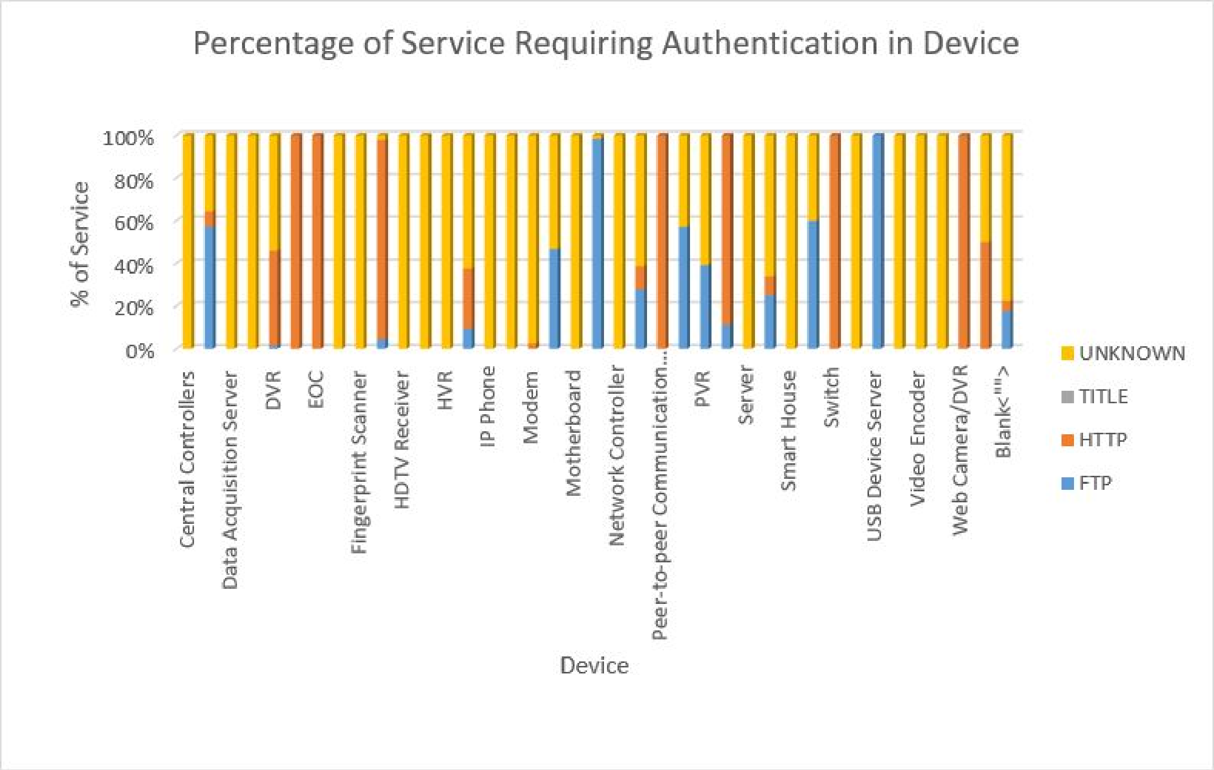

Figure 8: Device with percentage of service needing authentication

Frequency of credentials used in the attacks

Figure 9 reports the frequency analysis of passwords used for attacks from 1st of July until 1st of September. Our analysis reveals that the credentials can be divided to three categories. The first category includes passwords that are frequently used with a significant difference. This can be attributed to the choice of honeypot owners to set a special password. Assuming this, the botmaster frequently uses the same passwords to inject commands and this leads to the high frequency of usage. The second category includes passwords that are moderately used with almost a similar frequency and pattern of usage during the course of collecting data. These are other passwords the botnet uses for bruteforce but since they are not set for honeypot devices, they are used less. The third category includes password that are used a couple of times. Our guess is that, they are either targeted attacks or user mistakes in contacting wrong servers.

Based on the explained arguments, in order to eliminate the noise and concentrate on the botnet passwords only the top 30 credentials are analyzed. As it was mentioned in previous sections, one of the reasons IoT devices are such an easy target for compromise is usage of the default login credentials. In Figure 9 more than half of the attacks were completed with only 2 login password pairs. These are root:xc3511 which is a default password for a popular cheap DVR system, most other passwords are also default and never changed by many users. We made this conclusion based on manual checking of some the used credentials. The result of this manual analysis is reported in Table 2.

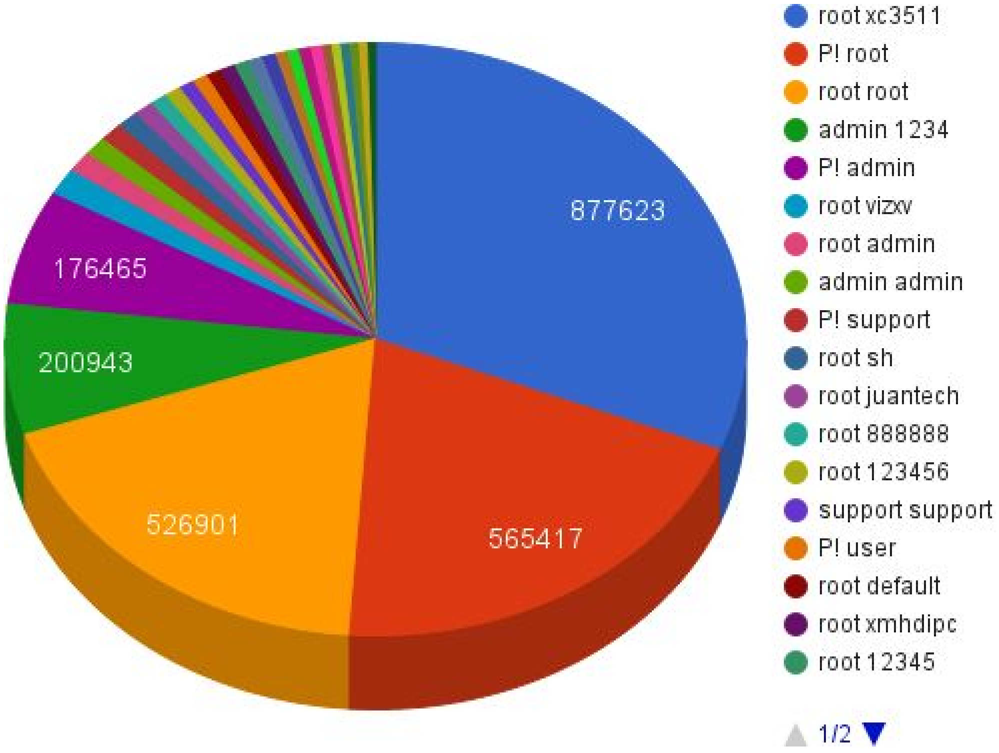

Figure 9: Frequency of the most popular credentials used in attacks over 2 month

Credential	**Vendor**	**Type**	**Device**	**Note**
root , xc3511	DVR DH - 3004	Default	CCTV	
root , root	Axis; Ambit;CTC; Avaya;Microplex;Milan	deafult	CCTV	
root , vizxv	CCTV - IPC - HFW4300S	default	CCTV	Old firmware
admin , 1234	Cisco; netgear	Default		
root , 888888	Dahua		CCTV	
netgear , netgear	NETGEAR	Simple	Router; wifi modem	
root , telnet	Telnet	Simple		
cisco , cisco	Access Point 1200 IOS	Default		
cisco , netgear	Netgear router	Simple		
cisco , password	Cisco - FreeRADIUS	Simple	switch	
pi , alphine				
admin , 7ujMko0admin	Dahua hdb3200c	Default	CCTV	New firmware
root , 1234				
admin , password	3COM Internet Firewall; Cisco; netgear	Default		
guest , maxided		Simple		
admin , maxided		Simple		
cisco , support	cisco	Simple		
root , support				
admin , 123456	Toshiba; KTI	Default; Simple		

root , admin				
root , raspberry	raspberry	Default		
oracle , cisco		Simple		
root , password	Wireless Location Appliance 2700 Series(cisco)	Default		

Table 2: Analysis of credentials used in the dataset

In Figure 10 analysis of frequency distribution over time was performed on all the datasets from the zip archive. Based on the reasons explained in the beginning of this section in order to eliminate noise and make the graph more readable all passwords that were used less than 2000 times were disregarded. Frequency distribution on a daily basis shows quite a repetitive pattern. However, from day 34 a drastic change can be observed. The amount of popular credentials being used increased just as the attack volume overall. This can be attributed to the gap between the collection of the data. An alternative reason for this sudden change can be usage of other credentials by the botnet (emergence of a new strain of botnet).

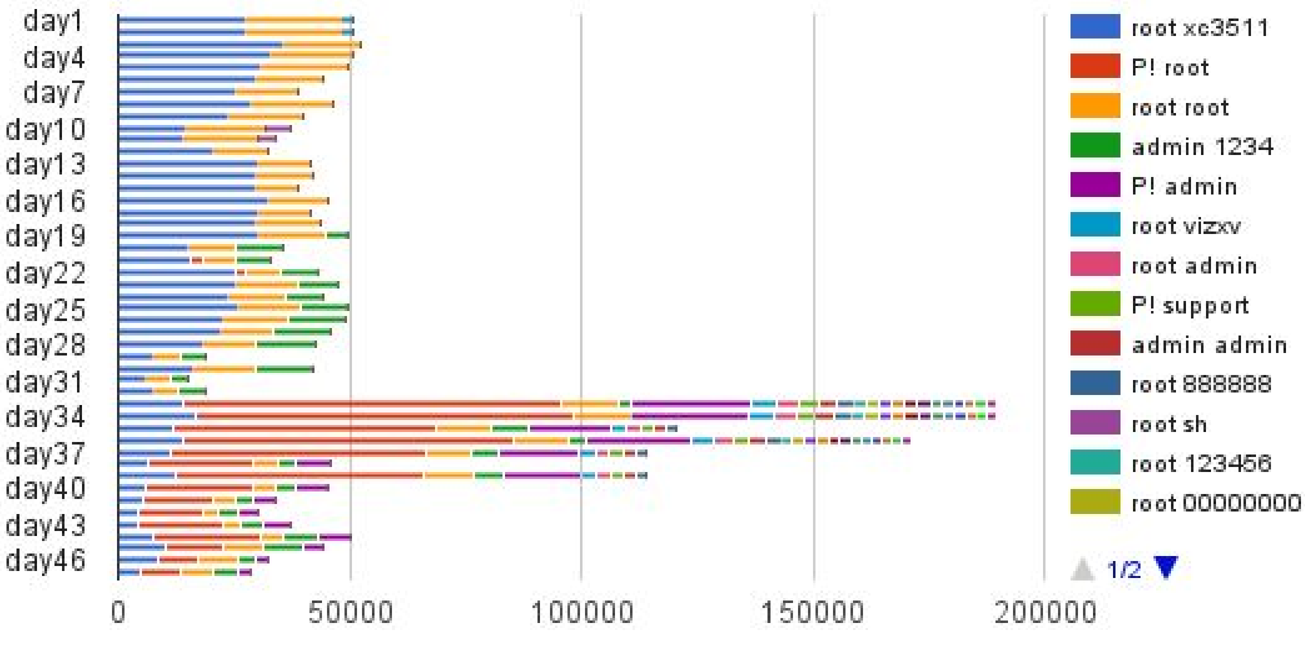

Figure 10: Frequency of the most popular credentials used in attacks on a daily basis