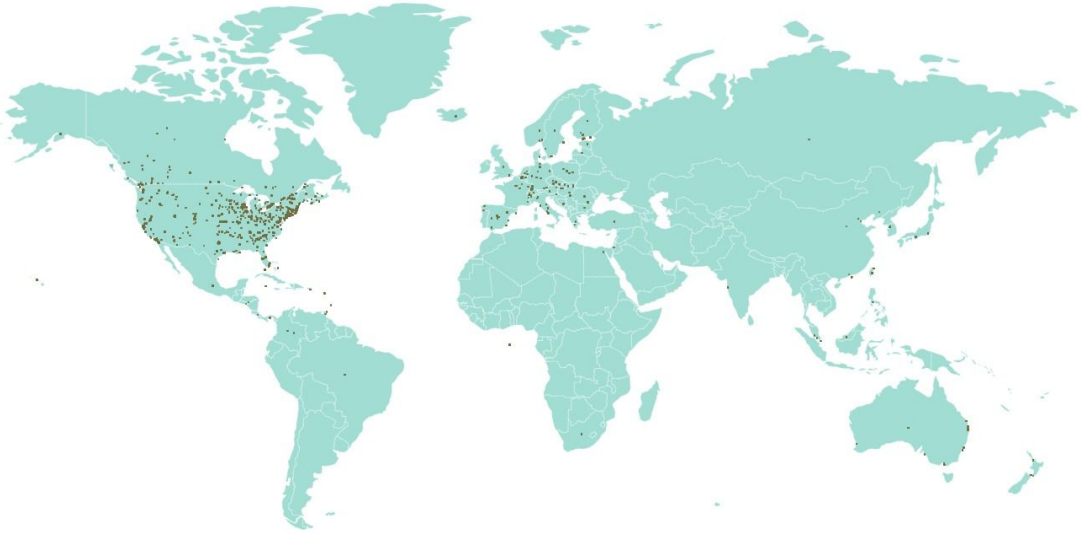
Metrics for ICS device security

Harikrishnan Manikandan - 4613201
Harm Griffioen - 4303598
Hans Schouten - 4314891
Lars van de Kamp - 4501829
Git: https://github.com/HarmGriffioen/EconomicsOfSecurity_ICS



General overview of publicly accessible ICS devices

Introduction

ICS devices have been around for a very long time, controlling some of the most critical infrastructure in the world. However, due to the uprise of the internet, people realized how beneficial it would be to connect the devices to the web for remote control and monitoring, thereby drastically reducing costs. By doing this, the devices became exposed to a whole new area of threats that were not present before. The devices were not designed to be connected to the web in a safe manner. It is therefore vital to measure how secure the critical infrastructure is, and what can be done in order to improve it. This document will give an overview of different metrics that can be used in order to measure this security.

The security issue at hand

Industrial Control Systems are part of critical infrastructure that companies, and society as a whole, heavily rely on. When access to these systems is granted to the wrong people, this could lead to large disruptions in everyday life. The Shodan ICS [Shodan ICS] dataset lists numerous Industrial Control Systems that are publicly accessible via the internet. Even when these are password protected, the systems should not be accessible from anywhere else in the world other than on-site or from specific control rooms. In addition, using password bruteforcing or exploitation of vulnerabilities in the firmware any publicly accessible Industrial Control System could get compromised. The high impact that the underlying infrastructure makes on civilization as a whole, makes the analysis crucial to fortify the defenses around the globe. This analysis on the threat on ICS devices is done by analyzing the situation from the perspective of the organizations that own the equipment. The actor that we therefore define for the use in this document are the organizations that use the ICS devices.

Ideal metrics for security decision makers

An ideal framework for security decision makers should take into account the:

1. Controls
2. Vulnerabilities
3. Incidents
4. (Prevented) losses.

Metrics on network hygiene indicate whether the appropriate configurations and controls are in place (aspect 1). Furthermore, these metrics are complemented with metrics on the level of software patching, attempting to fix current vulnerabilities (2). Besides the opportunities to attack a system, also the incentives to do so should be taken into account. Therefore, incident metrics are used to consider the entire threat landscape by mapping the attacks that did actually occur, or are most likely to occur using incident data like the Shodan ICS datasets (3). Finally, metrics need to be defined with regards to the losses and prevented losses (4). Both occurred and prevented incidents should be analyzed and quantified in such a way that the cost benefit analysis becomes clear due to the use of the chosen security policy.

In addition to the four pillars mentioned above, a metric needs to clearly resemble the level of security to all involved stakeholders. Everyone, ranging from a maintenance employee all the way to c level executives should be able to grasp the situation regarding the security infrastructure. By allowing all individuals to understand the situation, topics such as renewed investments or policy changes are more easily discussed. It provides context to the current situation, that in turn could validate new rounds of funding whereas originally it could have been seen as having low intrinsic value and thus a low return on investment. These discussions can generate consensus throughout the organization, highlighting the costs and benefits of security programs, thus helping security decision makers in doing their job.

Some examples of ideal metrics are described below:

- **The cost of a system being compromised**. Ideally, the security decision makers would have access to the exact costs and the distribution thereof, if a cyber attack is conducted on their facility. This would help in the estimation of prevented losses and

also form the foundation of a cost-benefit analysis allowing validation throughout the company of the needed decisions.

- **The vulnerabilities present within a device, and how they can be exploited.** When a company knows the different vulnerabilities that are present within a device and how they can be exploited, the needed security measures can be deployed to protect against those specific attack types. These vulnerabilities also include the current performance of physical access controls, as well as side channel attacks such as magnetic radiation analysis or LED pattern recognition. The cost of the preventive measures can then be compared to the potential cost of someone exploiting a specific vulnerability.
- **Knowing which controls, how much security, is needed.** In the previous point it is assumed that we know how to prevent certain attacks from being successful. In practice however, this often is not the case. Therefore, knowing which controls need to be implemented, what level of security is needed for protection of a certain threat, would be ideal for a security decision maker.
- **Likelihood of being target by different malicious actors.** When an organization is able to see how likely it is that someone would attack their systems, it can be used to turn the total cost of an attack into the expected cost. In addition, knowing who these actors are is crucial to protect against the right level of expertise and resources.
- **The security score.** All factors, including the earlier mentioned ones, should be combined into a regularly updated overall security score of the facility. This will give the organization insight into the status of the different managed systems and know which ones need to be updated. By performing regular penetration tests, this score can be updated to validate the security and see if all controls are also implemented correctly.

Metrics used in practice

In this section, we will list and discuss various metrics and security assessment frameworks that are used in practice.

The US National Security Agency (NSA) constructed a framework for measuring ICS security, focussing on the potential impact and loss relating to possible attack scenarios [NSA, 2010]. The first step is to create an accurate map reflecting all networked assets and the digital communication links that connect them. The next step is to perform a loss assessment which aims to identify actions that can be performed by an adversary via unauthorized access to the device. In the final step, the potential loss are determined based on the achievable malicious actions combined with the expected frequency of such incidents. Their framework offers a cost-benefit analysis approach which will allow organisations to prioritize their defensive efforts by identifying network security improvements that provide the greatest benefit for a given cost.

Boyer and McQueen define a set of seven ideal-based metrics for Industrial Control Systems [Boyer McQueen, 2008]. Each ideal is associated with an abstract dimension of security and represents a system condition at a given point in time. In a perfectly secure system, all ideals

are satisfied. When attempting to assess a complete ICS system in every possible dimension, an overarching measure is computed based on the individual metrics defined for each ideal.

NIST has also published a paper where they introduce the FISMA Risk Management Framework [Stouffer, 2011]. This framework consists of six steps in order to manage the security risks on ICS devices. Step four in this framework is the assessment of the security controls. In this step, the correct implementation and workings of the controls is tested. By implementing these controls correctly, security certifications can be obtained. These certifications show a certain level of security on the ICS security of a company.

CORAS [Solhaug, 2013] is another security risk analysis framework. It provides a language that helps modeling threats and risks. In order to easily graph the threats and risks, the Unified Modelling Language (UML) is used. Because this is a standard used by a lot of software developers, it is easy to understand for a wide variety of people.

Mukama defined security metrics based on risk analysis [Mukama, J. (2016)]. The design is called the Modified Risk Analysis Framework for ICS systems (MRAF-ICS), and is based on the NIST and CORAS framework. MRAF-ICS assigns weights to the different parts of the infrastructure, in order to emphasize the importance of said infrastructure. It uses threat modelling in order to identify vulnerabilities in the system.

Elizabeth Chew (2008), defines various metrics for ICS devices on behalf of NIST. She defines three core metrics, implementation, effectiveness/efficiency and impact. According to the paper, implementation measures are used to demonstrate progress in implementing programs, Effectiveness/efficiency measures are used to monitor if program-level processes and system level security controls are implemented correctly, and Impact measures are used to articulate the impact of information security on the mission. Examples of those metrics as defined by paper are:

1. Implementation metrics can for example be: "*Review the percentage of information systems with approved system security plans*", or: "*The percentage of information systems with password policies configured as required*".
2. Effectiveness/efficiency metrics can for example be: "*The percentage of enterprise operating system vulnerabilities for which patches have been applied or have been otherwise mitigated*". Or, focussing more on the effectiveness: "*Percentage of information security incidents caused by improperly configured access controls*", while a metric focussing on efficiency could be: "*The percentage of system components that undergo maintenance on schedule*".
3. Two impact metrics are for example: "*The percentage of the agency's information system budget devoted to information security*" and "*The number of information system budget devoted to information security*".

We had the pleasure and the opportunity to ask this question to a professional working at Applied Risk. He told us that Applied Risk in practice does not have concrete security metrics in place. However, they do request network drawings from companies and the documentation on the patch management, as it helps them to estimate the security of

devices in the network. They also look at the protocols used throughout the entire operation, as some may require more expertise to operate providing a limited threat surface. In addition, the default settings on these protocols can have a large impact on the security situation. For example, modbus, by default, does not have any form of authentication.

Metrics that can be designed from the dataset

In this section, we define different metrics for the dataset in order to measure security. In order to define the metrics, we first have to define what a metric is. We have taken the following definition of a metric:

"Metrics allow us to measure attributes and behaviors of interest. A meter, for example, is a metric that allows us to measure length, while the number of defects per shipment is a metric that allows us to measure quality." [Cyber Threat Metrics, 2012]

Following this definition, version number itself is not a metric, but the number of vulnerabilities known for a certain version number is.

After analyzing the dataset, the following metrics have been devised. The goal is to create the most useful metrics to assess the security situation, within the constraints of the provided data. These metrics are the closest attempt to the ideal ones as possible with the currently available information.

Number of publicly accessible devices over time. This shows improvements or declines in security of the different types of ICS. In addition, it could also reflect the interest third parties have in specific industries.

Update frequency. Using the version number and the timestamps, a metric can be created for the update frequency of the devices. The lower the update frequency, the more susceptible the device is against existing and future exploits.

The number of firmware vulnerabilities. By cross-referencing the current version numbers with a list of known vulnerabilities for specific versions [CVE Details, NVD], the score of the most critical vulnerability can be used as a metric defining the lack of security in a specific system.

Manufacturer. The manufacturer of the device can also be taken into account. This can be cross-referenced with a list of 'vulnerable' manufacturers, those who originate from a certain country or have a poor reputation. The vulnerabilities known per manufacturer will be the metric of this trust score per manufacturer.

The number of accessible devices per location. While the dataset only shows publicly accessible devices, this might still be a useful metric to see which locations are most vulnerable to attacks. These area's will have a higher attack vector than other areas, which need to be combined into a likelihood score per location.

The number of accessible devices per organization (type). Another highly influential factor is the type of organization that is behind the specific accessible system. Certain high

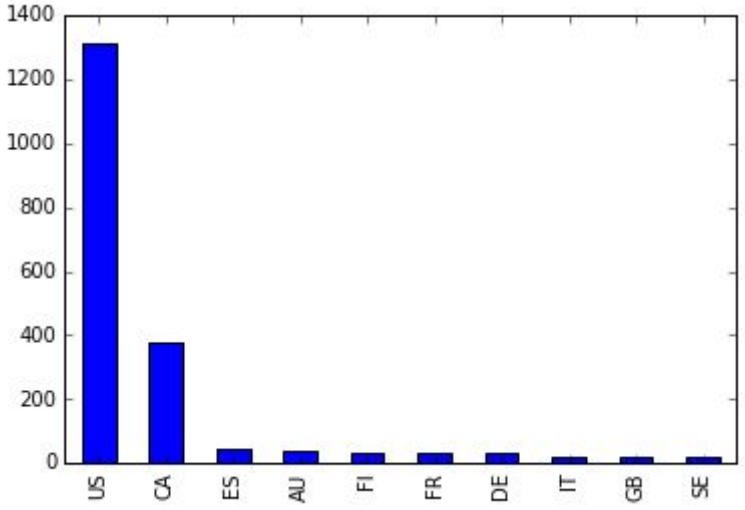
profile institutions might be more attractive targets due to their larger potential impact on society. Therefore, industry that belongs to an instance could indicate the likelihood of being targeted and thus the number of incidents and perhaps even the number of known vulnerabilities that are known. The high-profile nature of a system, could lead to more time being invested in finding vulnerabilities and exploiting these since there is also a larger return when successful.

Metric evaluation

The metrics suggested in the above section will be evaluated along with the information contained in the dataset to discuss their efficacy and viability in this section.

The required attributes are removed separately using Pandas (Python) to visualize the required metrics. A simple script [Json_parser.py] which parses the JSON files to obtain the data relevant to the metric is used. Some of the attributes are not populated for all the entries which is why some of the suggested metrics are unlikely to be useful with this dataset.

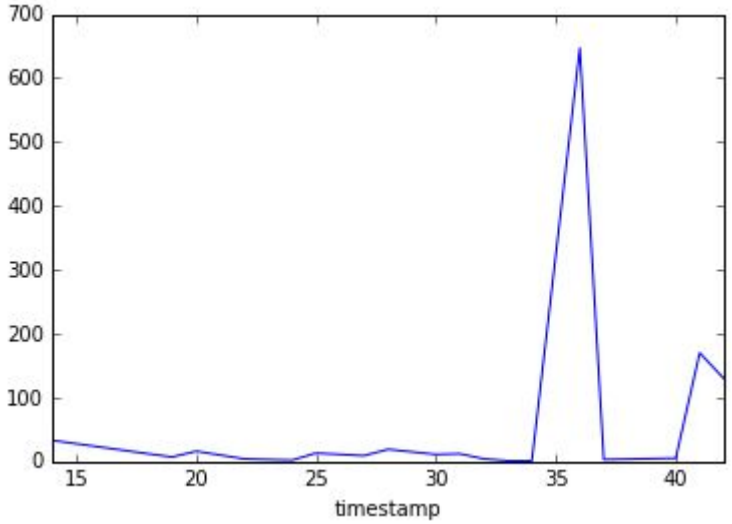
The number of accessible devices per location.
A metric of number of publicly accessible ICS per country can be used as a score to assess threat level per region. This isn't a deterministic indicator, as there are many reasons why some countries are more susceptible than others. A graph of the publicly accessible devices per top 10 countries is shown below:



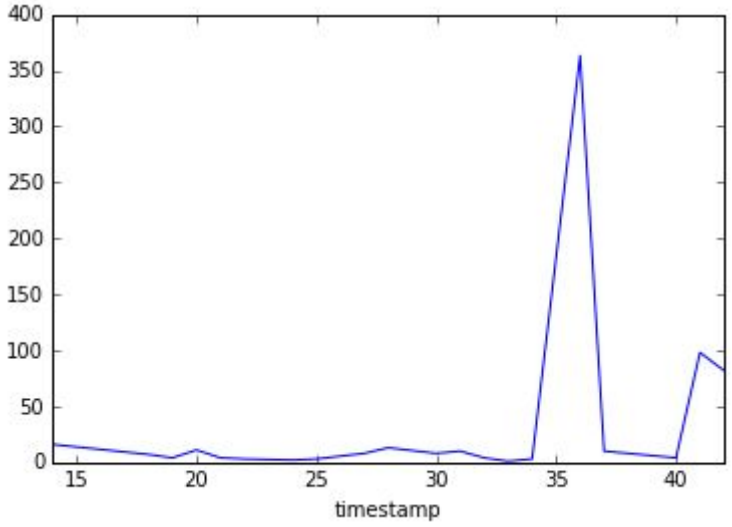
This data has been normalised by dividing the number of devices by the Global ICT development index from 2015. Taiwan has been removed from the list as it does not have a rating or ranking in the index.

Number of publicly accessible devices over time.
The number of publicly accessible devices per company over a period of time must also be computed and maintained to see any general trends across the sector. The number of publicly accessible devices added each week per organisation are shown in the figures

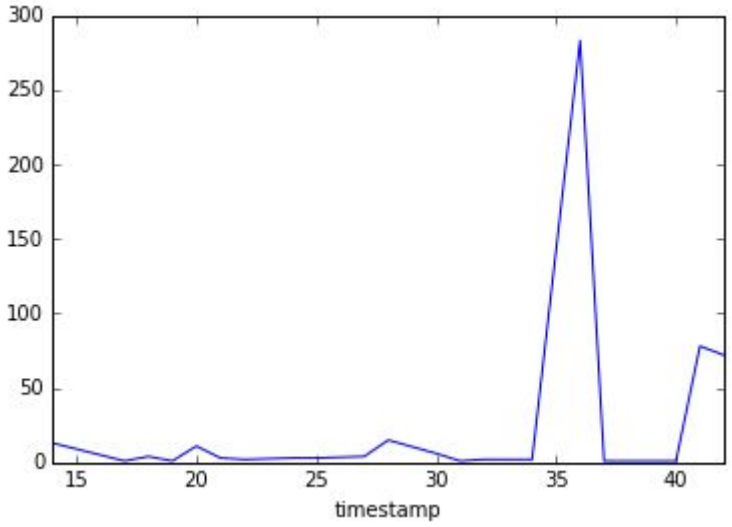
below. As we can see, the same trend appears throughout the different organizations. The last plot shows the overall trend containing all instances.



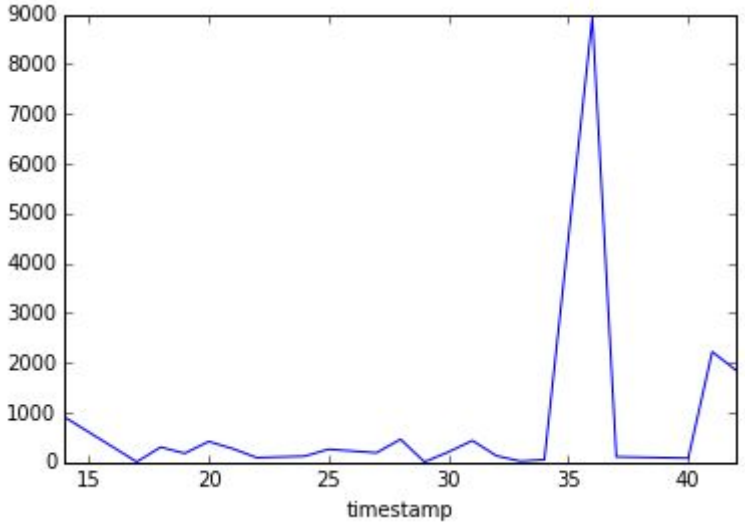
AT&T Internet Services



Comcast Business Communications



Comcast Cable



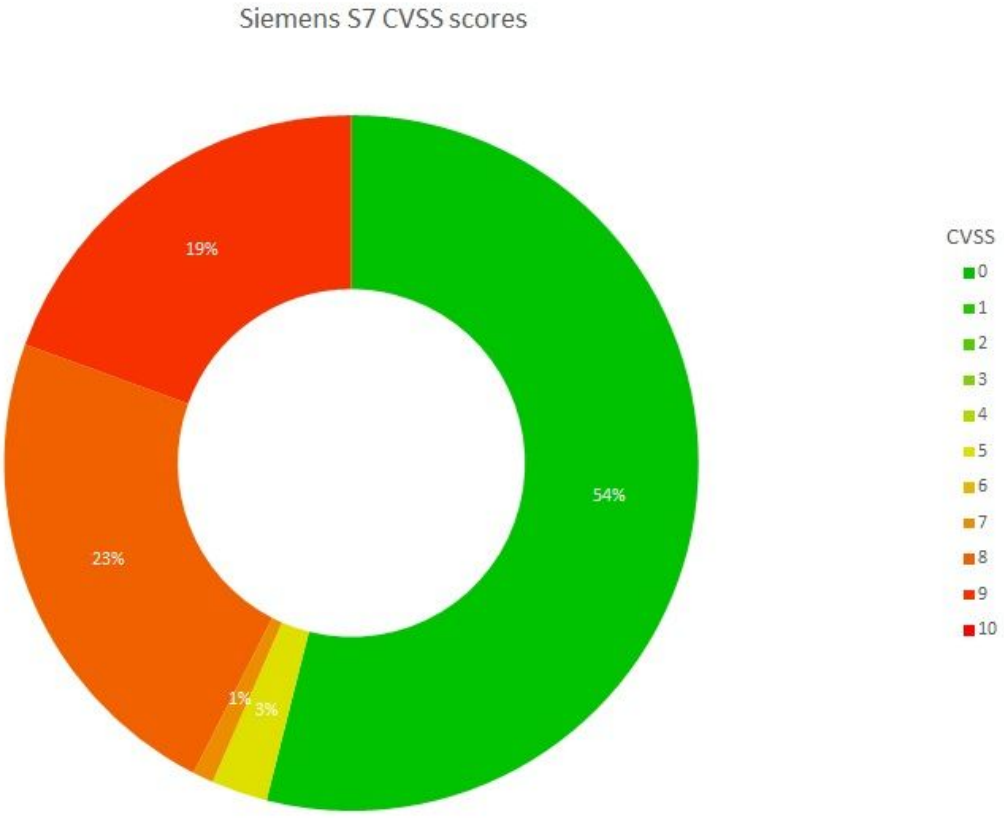
The number of firmware vulnerabilities

By cross-referencing the current version numbers with a list of known vulnerabilities for specific versions [CVE Details, NVD], the score of the most critical vulnerability can be used as a metric defining the lack of security in a specific system.

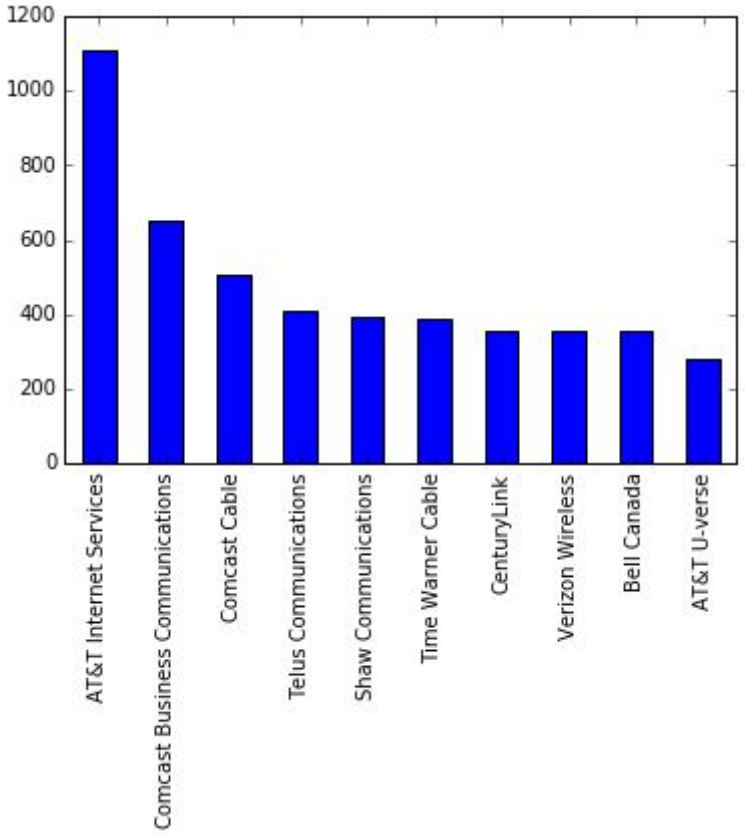
After analysing the Shodan datasets it became clear that searching for the exact firmware versions in CVE databases did not lead to any results. Some manual processing had to be done, mapping raw metadata to strings that can be entered as search queries. Since it is very time consuming to preprocess the data and search these databases, this will only be evaluated for the Siemens S7 dataset.

First, all vulnerabilities regarding Siemens hardware are requested from the US National Vulnerability Database. For each of them the description is analysed and searches through the Shodan dataset are performed to determine whether the vulnerability does apply to these systems. For all applicable vulnerabilities, detection rules are constructed in the form of a set of strings. If all strings are present in the metadata it can be derived that the device has the particular vulnerability. To make listing the vulnerabilities easier, a simple user

interface is written in Python: [Vulnerability parser]. This script converts user input to a JSON structure containing fingerprints of all considered vulnerabilities.
Next, each of the Siemens S7 devices from the Shodan dataset is matched to the created list of vulnerability fingerprints. The highest CVSS score is tracked for each of the systems. Aggregating the number of occurrences of the maximum CVSS scores of all systems gives an overview of how many systems are at each level of security deficiency. Since these amounts in itself do not yield much information, they are normalized over the total number of Siemens S7 devices in the dataset (3710). In the calculations, 676 devices were ignored as they did not contain any version information. The resulting numbers show that 19% of the successfully scanned Siemens S7 devices have at least one vulnerability of criticality 9. Since 23% of all devices have at least one vulnerability of criticality level 8, almost half of all systems have a vulnerability with high criticality on the CVSS scale.

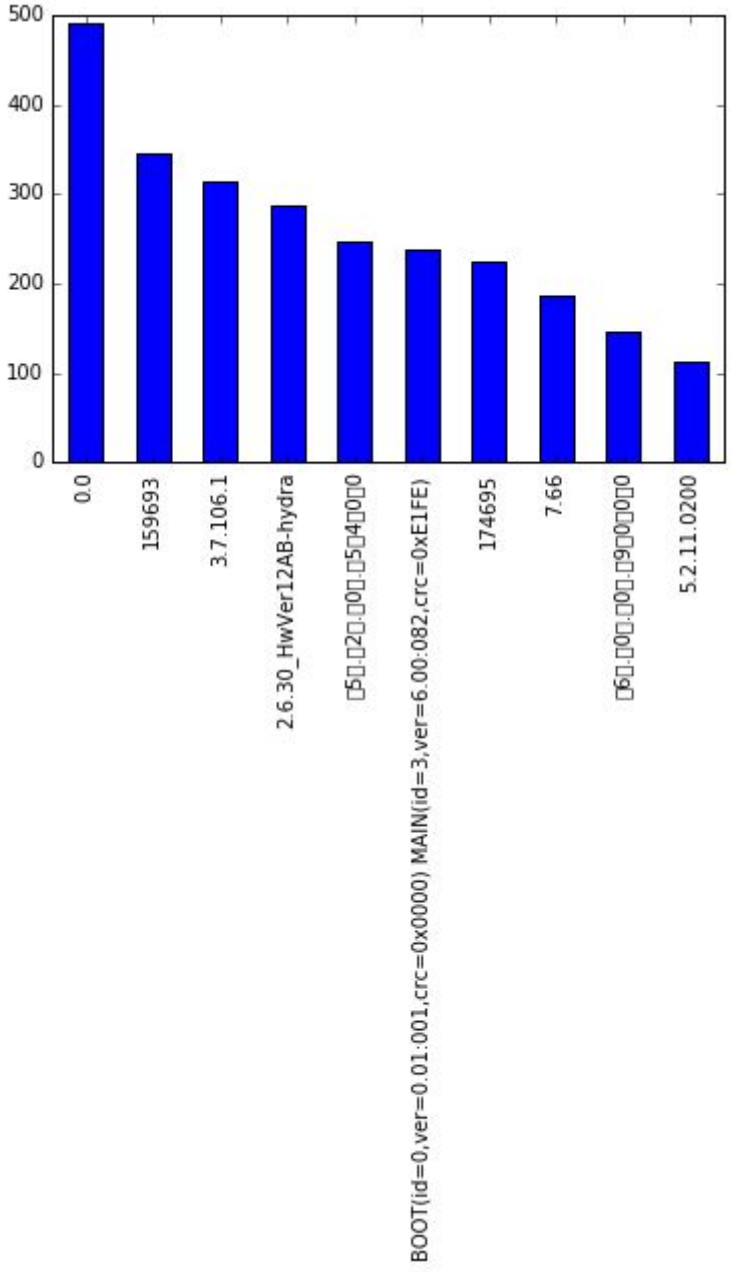


The number of accessible devices per organization (type).
The highest number of entries in the “bacnet.json” dataset. Do note that the higher the number of accessible devices does not automatically imply poorer security levels, as this could also be due to a larger company size. These need to be normalized to company size, to paint a more accurate picture. Due to the high level of granularity that is present in the dataset, large companies are split into multiple parts as can be seen below for AT&T. On these smaller parts, the available data is limited making normalization a very hard task. Still, the number of devices is a critical metric and should be included as it does portrait the general awareness of the organization, and the threat surface that is currently present. Ideally, when a more detailed dataset appears, this metric would still be normalized.



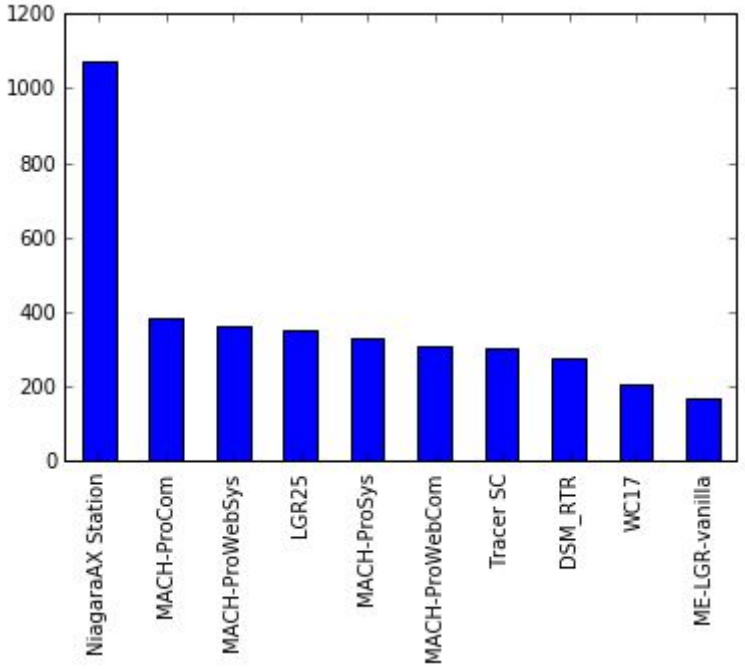
Update frequency

The version numbers are hard to track as most of the entries are empty while very few of the rest have valuable information. It seems that the data does not provide us with the needed insights to validate these metrics. The 10 most frequent version numbers are shown below:



Manufacturer

The product and manufacturer information is also hard to track since most of the entries do not have this field filled out. Due to the large variety in manufacturing companies, and the limited availability of public market share information, the manufacturer performance is not normalized. However, a high presence in these datasets could still say something about the credibility of the manufacturer and thus the overall security of the system. Therefore, we have decided to still include this metric. The list of top 10 most frequent products are shown below:



Conclusion

Even though several variables are lacking in the dataset, some of our metrics can still be derived. The location, time-series, and organization analysis have shown potential to indicate the security of a system. The manufacturer metric is partially affected by missing data, but could prove to still be useful as the top 10 does allow us to see predominant manufacturers in the industry. The CVE analysis seems to be particularly useful, as it clearly shows the severity of the security situation for the different systems. Hopefully, our metrics will improve the understanding of the different security situations, and help organizations in better protecting its critical infrastructure for many years to come.

References

National Security Agency (NSA). (2010). A Framework for Assessing and Improving the Security Posture of Industrial Control Systems (ICS)
https://scadahacker.com/library/Documents/Assessment_Guidance/NSA%20-%20Framework%20for%20Assessing%20and%20Improving%20Security%20Posture%20of%20ICS.pdf

Boyer and McQueen. (2008). Ideal Based Cyber Security Technical Metrics for Control Systems
https://inldigitallibrary.inl.gov/sites/sti/sti/3884735.pdf

Shodan map of publicly accessible Industrial Control Systems
https://icsmap.shodan.io/

Mateski et al. (2012). Cyber Threat Metrics
https://fas.org/irp/eprint/metrics.pdf

CVE Details. Current CVSS Score Distribution For All Vulnerabilities
http://www.cvedetails.com/

NIST. National Vulnerability Database
https://nvd.nist.gov/vuln/search

Vulnerability parser
https://github.com/HarmGriffioen/EconomicsOfSecurity_ICS/tree/master/metrics/number-of-vulnerabilities

Mukama, J. (2016) Risk Analysis as a Security Metric for Industrial Control Systems. Göteborg : Chalmers University of Technology

Elizabeth Chew et al., "Performance Measurement Guide for Information Security," NIST, Gaithersburg, Special Publication NIST SP-800-55, 2008.

Global ICT Development Index
http://www.itu.int/net4/ITU-D/idi/2015/

Country Codes Index
http://www.nationsonline.org/oneworld/country_code_list.htm

Stouffer, K., Falco, J., & Scarfone, K. (2011). Guide to industrial control systems (ICS) security. NIST special publication, 800(82), 16-16.

Solhaug, B., & Stølen, K. (2013). The CORAS Language-Why it is designed the way it is. In Proc. 11th International Conference on Structural Safety and Reliability (ICOSSAR'13) (pp. 3155-3162).