Security Posture Based Incident Forecasting

A Thesis

Submitted to the Faculty

of

Drexel University

by

Dagmawi Mulugeta

in partial fulfillment of the

requirements for the degree

of

Master of Science

June 2019

Acknowledgements

It is through the support and counsel of numerous individuals that I have compiled the work below. Firstly, to my advisors, Dr. Steven Weber and Ben Goodman, I extend my deepest gratitude for their invaluable guidance. They constantly encouraged me to challenge myself in new and interesting ways. I will, without a doubt, continue to use the lessons from our biweekly meetings in the future. Secondly, I would like to thank Raymond Canzanese for being an outstanding mentor, both in research and life. His feedback has shaped me into the engineer that I am today. Thirdly, I would like to thank my parents (Mesfin and Hanna), my siblings (Nola and Eto), and rest of my family for being daring enough to send me to school overseas. Their support has made the time apart slightly more bearable.

I would also like to thank Pushkin, for always demanding that I focus on the bigger picture, which I hope to someday pay forward to someone else; Yoan and Pranay, for being superb friends, challenging me, and believing in my potential when I wasn't a believer; Jordane and Sena, for consistently checking up on me and providing emotional support through what seemed to be a challenging year; Anishi, Saloni, and Mahshid, for helping solve many problems (both in and out of the lab) through the majority of this process; Hari and Abhinanda, for imparting many priceless tokens on graduate student life; Nick, Mark, Tyler, Rob, Marco, Liz, Meghan, Peter, Alex, Lydia, and Jordan, for keeping me sane all those late nights I was up fixing a gruelling issue; CyberDragons (Colbert, Ryan, Nahid, and many others) for aiding in the understanding of many cyber security concepts and affording me the pleasure of being a CyberDragon.

Lastly, and most importantly, I would like to thank **God**, as it is through his grace, wisdom, and power that I have been able to successfully finish this thesis.

I dedicate this thesis to my role model and loving uncle, **Girma Deyaso**. You were a sincere soul that was taken from us too soon.

Contents

List of Tables

List of Figures

Abstract

Security Posture Based Incident Forecasting
Dagmawi Mulugeta
Steven Weber, Ph.D. and Ben Goodman

The pervasiveness of technology (e.g., the internet) is coupled with an expansion of the threat landscape. The numerous network services, versions of these services, and possible configurations make it difficult to predict the likelihood of a security incident (e.g., data breach) for hosts on the public internet. We attempt to explore this problem by analyzing data from Censys, a database of internet-wide scans, to collect network configurations for organizations that reported security incidents in 2017-2018. We seek to determine which common patterns in network configurations are associated with likelihood of reporting a security incident by providing a comparison between victim and non-victim organizations' hosts.

We design a data pipeline that extracts 1,386 features from each host machine, enabling us to build upon previous academic approaches by utilizing a more holistic feature space. We then use an Isolation Forest (Outlier Detection) algorithm, a novel addition to the problem domain, to identify outlying hosts in organizations' networks and effectively reduce the data space. We find that we can identify outlying hosts with 0.84 ± 0.01 accuracy, 0.84 ± 0.01 f1-score, and 0.18 ± 0.04 fpr. We then present the important properties that make a host an outlier in the feature space. For example, we find that Diffie-Hellman on https protocol and presence of SSH protocol are important indicators of outlying machines.

These representative liers (outliers and inliers) are then used to create risk vectors for the victim and non-victim organizations. Using these risk vectors, we are able to discriminate between organizations that report security incidents with an average 0.73 ± 0.06 accuracy, 0.73 ± 0.06 f1-score, and 0.25 ± 0.10 fpr. Through use of these techniques, we are able to correlate between certain features and the victim label, thus demonstrating the predictive power of specific features (e.g., SSH protocol and FREAK vulnerability).

In short, we (1) introduce a novel approach to building a rich configuration-centric feature space within which we successfully (2) analyze network postures and their correlations with security incidents, while (3) reducing the data space, and simultaneously, the processing cost of this sort of analyses.

Keywords: security incident, data breach, incident prediction, outlier detection, risk profiling, cyber risk

Chapter 1: Introduction

1.1 Motivation and background

As the threat landscape evolves, security incidents continue to wreak havoc on the internet. Privacy Rights Clearinghouse [7], a data breach aggregation portal, shows that there have been 614 hacking or malware related incidents that have supposedly disclosed 914,388,535 sensitive records in 2017-18 [8]. The cost of mitigating these incidents requires massive budgets from the victim organizations. Edwards et al. projected that, in the 2016-19 time span, breaches could cost north of 179 billion USD [9]. This is only exaggerated with the growing popularity of internet commerce. With this much personal and financial security at stake, it is wise to analyze and understand the different technical aspects of a security incident. It is only after doing this that the internet community can truly begin to mitigate this issue.

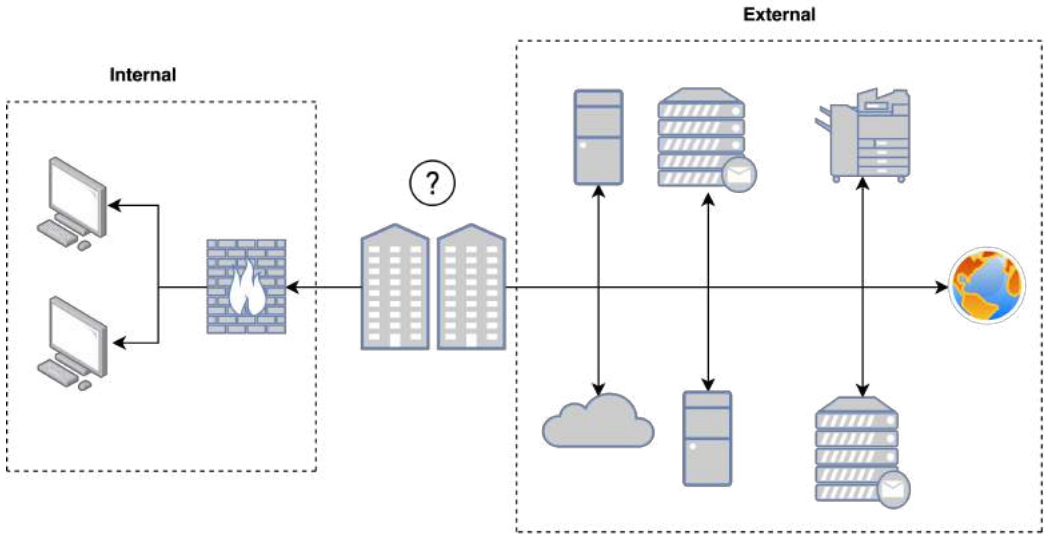


Figure 1.1: Internal and external network segments of an organization

In Figure 1.1 we can see the two different network segments (internal and external) for a given organization. In this paper, we analyze external network posture, one of the aspects of a security incident, and how it relates to victim organizations. The external network posture for an organization is the set of configurations for the hosts on the public internet. These hosts are accessible by an

individual with an internet connection. External network posture, at a high level, is comprised of configurations for different protocols, including the application behind the protocol, the version of the application and many other details around the state of these configurations. The advantage of using external network information is that it is publicly available, and it does not require internal information from the organization, which makes it relatively easier to acquire at scale.

This paper investigates the extent to which these configurations can be used to predict the likelihood of a security incident. We perform this investigation by carefully collecting a group of victim organizations (each of which reported a security incident) and non-victim organizations (none of which have reported a security incident) and comparing their external network postures (See Figure 1.2).

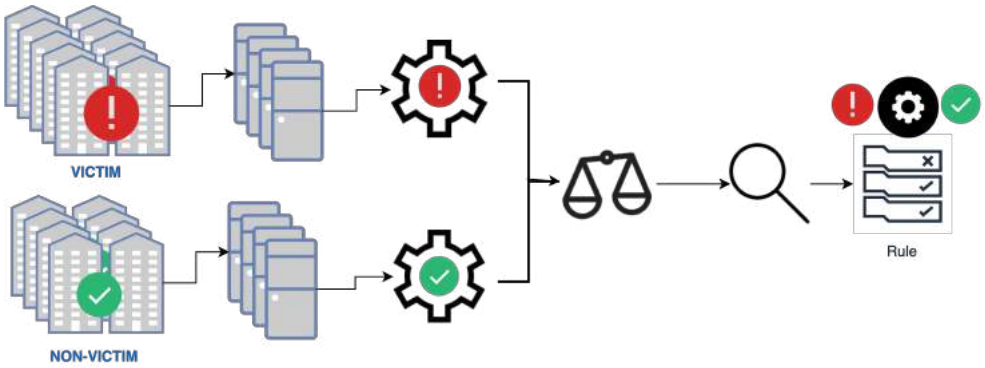


Figure 1.2: Contemporary approach in problem domain: collect victim and non-victim organizations, attribute their assets, and compare rules that discern victim configurations

The intuition here is that the configurations relate, in some degree, to hacking or malware incidents. This relationship exists either directly through an attacker exploiting a vulnerability, or indirectly through lack of adequate internal policy and implementation of controls to ensure an organization's security. This assumption that the company culture is correlated with the external internet configurations has been made by previous works as well. Zhang et al. [1] and Liu et al. [2] have shown that misconfigurations are one cause of maliciousness, and can be used to predict security incidents with high accuracy. We extend on their research by using heterogenous data, more holistic methods of representation, and novel machine learning algorithms.

The overall goal of this analysis is to provide an effective way of conducting cyber risk assessment.

Cyber risk, within the scope of this domain, is analogous to the likelihood to report a security incident. A cyber risk vector is numeric representation of this cyber risk. Hence, the (cyber) risk vector is built from organizations' external network configurations. The data sources utilized in this project are entirely public. The loose coupling of the data and technique affords users the ability to apply the techniques listed in this paper to other similar data sets. As a result of the very intuitive statistical concepts and inspections that are employed in our approach, the analysis can be easily extended to the real world by system administrators and security researchers.

1.2 Problem definition and scope

Again, the goal of this paper is to conduct risk assessment by profiling network configurations that are associated with organizations that report security incidents. To accomplish this, we utilize a numeric interpretation of the configurations for these hosts. We then set up a machine learning problem using the configurations as features and the report label as the target. These feature vectors are then used to build the organizations' risk vectors and predict whether an organization will report a security incident.

Censys, a public internet search engine, actively scans all the internet hosts, and services on these hosts [4]. It then curates and annotates the scans before storing it in a database for research use. This project utilizes Censys's database of internet scans to collect host configuration information. With the appearance of security tools like Censys that scan the public IPv4 space, identifying weaknesses in organizations' network posture takes a matter of seconds. To counter this, the security community needs to use tools like Censys to adequately understand the common vulnerabilities of the public internet and identify these misconfigurations before it is too late.

This project **does** perform the following:

- Attempt to advance the state of the art in cyber risk assessment, with respect to contemporary peer reviewed research, on attributing digital assets to corporations
- Analyze these attributed assets over a time span to find associations with reported security incidents

- Present a method to reduce the data space involved with similar analyses

This project **does not** perform the following (similar to previous works [10, 2]):

- Vulnerability analysis / Intrusion Point Detection: In this project, we do not aim to identify vulnerabilities in internet hosts, nor do we attempt to uncover possible attack vectors used by malicious agents. Despite this fact, our results and previous works [10, 2] reveal that inadequately-managed systems are correlated with organizations that report security incidents.
- Collect Measurements: This project brings together numerous data sources, but other than digital footprinting, it does not conduct novel measurements, i.e., we do not scan the organizations' network ourselves.
- Utilize Internal data: We do not analyze internal application logs, or other information that is not available to an external attacker. Instead, our analysis focuses on outside-in data. Outside-in data refers to external network configurations for an organization, e.g., DNS server configurations, mail configurations, and web server configurations.

Due to time and resource constraints, we leave many challenges that are encountered as direction for future work. These constraints, challenges, and design decisions will be mentioned through out the paper.

1.3 Contributions

Current works, both in academia [11, 2, 1] and industry [12, 13, 14], utilize a similar pipeline for collecting organizations, forming the risk vectors, and analyzing the most effective discrimination rules [15]. This pipeline can be seen in Figure 1.2. In this pipeline, we first identify a set of victim and non-victim organizations. This is then followed by attributing their digital assets, and configurations on these assets. These configurations are then compared to identify the rules that provide the best discrimination.

Using Figure 1.2 as a starting point, we contribute to the general pipeline through the following ways (See Figure 1.3):

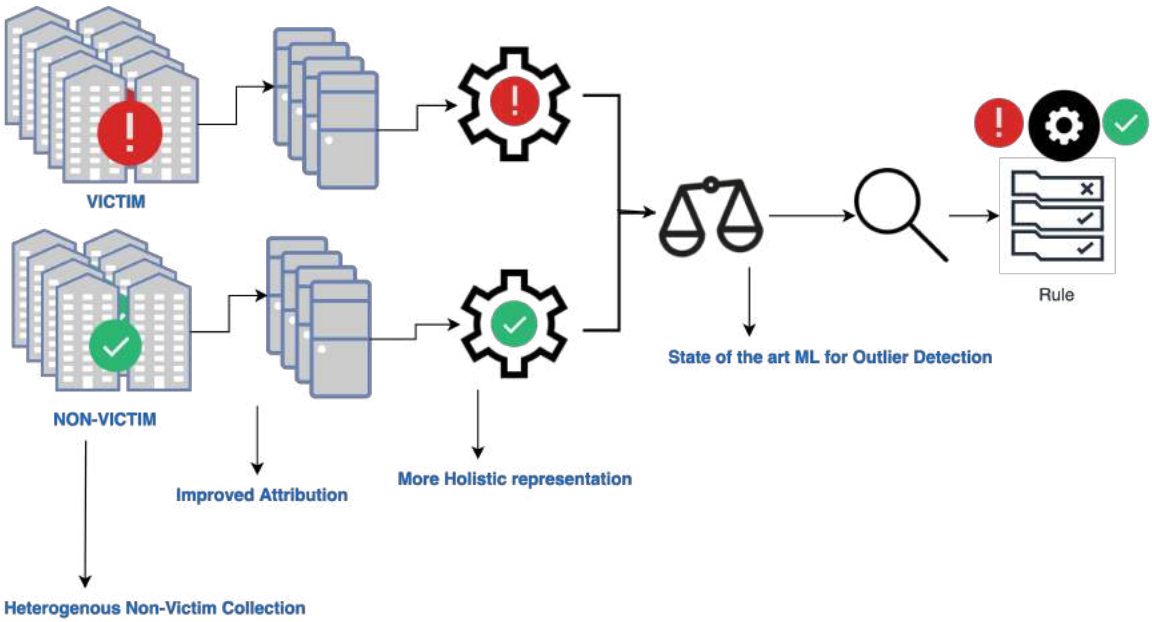


Figure 1.3: Novel contributions to the contemporary approach

1. **Heterogenous non-victim collection**: To provide an accurate analysis of victim organizations, we have to compare these victims against an exemplary set of non-victim organizations. Although defining a non-victim is simple, executing the collection is rather difficult due to selection bias. To counter this obstacle, we collect non-victim organizations using three different methods, namely

    - Method 1: Randomly sampling IPv4 addresses, and performing a reverse lookup on the DNS record to locate the domain name
    - Method 2: Randomly sampling digital certificates, and collecting the associated subject domain
    - Method 3: Collecting a subset of the Cyber security 500 organizations, a list maintained by the magazine Cyber security Ventures [16]

2. **Improved asset attribution**: Footprinting is the process of identifying the assets for an organization (more on this in Section 3.2.2). This technique will allow attackers to quickly gauge an organization's security posture [17]. Attackers, through the use of footprinting tools,

can take an unknown quantity and reduce it to a specific range of domain names, network blocks, and individual IP addresses [17]. Since there is no ground truth data source that maintains ownership and usage of digital assets on the public internet, we use these contemporary reconnaissance tools to attribute digital assets. The intuition here is that if we can not identify the owners of internet hosts with certainty, then looking at them through the same lens as an attacker will be the best attribution technique available to security researchers.

3. **More holistic host representation**: Current works [2, 1] utilize features spaces of about 260 or fewer. These are reputation black lists (public lists of known malicious IP addresses) together with 5 - 8 misconfiguration metrics to predict when organizations will experience security incidents (victim organizations) [2, 1]. In this paper, our feature space spans 1,386 features that represent detailed configurations on these hosts, which affords us the ability to conduct our analysis at a finer resolution than previously possible. To the best of our knowledge, this feature space is the largest to date.

4. **State of the art ML for Outlier Detection**: Our work introduces isolation forest (an outlier detection algorithm) into the problem domain to identify interesting configurations. This outlier detection algorithm allows us to effectively reduce the data space to a set of hosts about 12% the original size while achieving an incident prediction accuracy of 0.73 ± 0.06 , 0.73 ± 0.06 f1-score, and 0.25 ± 0.10 fpr. This is achieved by only looking at these risk vectors, which summarize the external configurations, for different organizations.

1.4 Terminology

This section presents brief definitions for terms as they are used thoughout this paper

1. **Security Incident**: an act of violating an explicit or implied security policy. Informally, it is an event that indicates an organization's IT system may have been compromised.

2. **Data Breach**: a security incident in which sensitive, protected or confidential data is copied, transmitted, viewed, stolen or used by an individual unauthorized to do so [15].

3. **Victim and Non-Victim Organizations**: A victim organization is defined an organization that has reported a security incident. On the contrary, a non-victim organization is an organization that has not reported a security incident.

4. **Cohort**: a collection of organizations (either victim or non-victim).

5. **Protocol**: a set of procedures that govern communication in a network.

6. **Service (Network Service)**: network application running a specific protocol.

7. **Port**: a number between 0 and 65535 that identifies a network service running on a host.

8. **Internet Scanning**: a process of identifying hosts on the internet, and the corresponding network services on those hosts.

9. **Domain / Domain Name)**: a label that identifies a group of computers belonging to an entity, e.g., drexel.edu.

10. **Subdomain / Subdomain Name)**: a label that smaller group of computers belonging to a domain or another subdomain, e.g., www.drexel.edu.

11. **Domain Name System (DNS)**: system on the Internet that converts domain and subdomain names to corresponding IP addresses and vice versa.

12. **Digital Asset**: could be IPv4 address, domain name or the host behind the IPv4 address that belong to a certain entity (organization).

13. **Footprinting**: reconnaissance process to identify all the digital assets belonging to an entity.

14. **Organization Size**: the number of internet-facing or publicly exposed hosts an organization has.

15. **Subdomain Enumeration**: footprinting technique that identifies all the subdomains for a specific domain.

16. **On-premises vs. Cloud Hosted**: on premises refers to computers an organizations maintains themselves, while, cloud hosted means computers that are maintained by a cloud provider (e.g., GoDaddy, Amazon).

1.5 Outline

The rest of this paper is structured as follows. In Chapter 2 we discuss the relevant works to this analysis, including Censys (and ZMap) and past works we build upon. Chapter 3 discusses the data pipeline including how we selected the organizations, how we attributed the hosts, how we downloaded the data from Censys, and how we extracted the features from the hosts. It also discusses the design decisions and the implied consequences. Chapter 4 presents an analysis of the data through setting up the classification problem, selecting the interesting hosts, and generation of the feature importance charts. Finally, in Chapter 5, we conclude the analysis and list out some avenues for future work.

Chapter 2: Relevant works

2.1 Internet scanning

Durumeric et al. presented a tool called ZMap that can be used to scan the public IPv4 space in under an hour [18]. ZMap is a modular, open-source network scanner specifically architected to perform Internet-wide scans. It is capable of surveying the entire IPv4 address space in under 45 minutes from user space on a single machine, approaching the theoretical maximum speed of gigabit Ethernet. A typical NMap scan could take weeks, however, ZMap is 1300x faster than NMap on the most aggressive setting.

Durumeric et al. also introduced Censys, a public search engine and data processing facility based on conducting internet-wide scans using ZMap. The purpose of Censys is to help researchers answer security related questions about the IPv4 address space. Censys dramatically reduces the effort needed to investigate questions like "what fraction of HTTPS servers prefer forward-secret key exchange methods?", enabling researchers to focus on asking more important questions [4]. Censys collects structured data every 24 hours, validates this data and performs application-layer handshakes to produce structured data about each host and protocol. This extracts valuable fields and annotates handshakes with additional metadata, such as device model and software version. The end product being structured JSON data describing a certain aspect of how a host is configured and annotated with additional metadata (e.g., device manufacturer and model). Hence, Censys extracts significant values and transforms handshake data into consistent, structured records that conform to a published schema. The fact that Censys conducts scans every 24 hours ensures that the scans are a realistic representation of what is on the Internet at that given moment.

Censys supports querying fields derived from scans and generating statistical reports. The query feature supports full-text searches, regular expressions, and numeric ranges, and queries can be combined with Boolean logic. Censys not only maintains an up-to-date snapshot of the hosts and services running across the public IPv4 address space, but also exposes this data through a public

search engine, REST API, publicly accessible tables on Google BigQuery, and downloadable data sets. Google BigQuery [19] is a cloud database engine designed for performing large queries. Censys's BigQuery tables contain the daily ZDb snapshots of the IPv4 address space. Hence, raw application handshakes and daily point-in-time snapshots of the structured data can be queried using SQL through these publicly accessible BigQuery tables.

One important note is that Censys does not perform login attempts, deploy any exploits, or try to access non-public resource paths. This makes it difficult to use Censys to conduct vulnerability analysis. Censys also affords user exclusion requests and respond to requests within 24 hours. This effectively excludes certain internet hosts from an analysis that uses Censys as a data medium. This project leveraged Censys directly (ZMap indirectly) for our analysis to reduce the work needed to acquire organizations' public network stature.

2.2 Incident analysis and forecasting

Sun et al. conducted a survey of 19 core proactive security incident prediction papers [15]. The authors mention that proactive (prediction) and reactive (detection) methods have been used in academia and industry to deal with cyber incidents. They state that there has been a general shift from security incident detection to prediction. Moreover, recent years have seen predictive studies involving malicious activity [2, 15]. In essence, researchers and organizations are trying to fill the security gaps by proactively predicting incidents based on observed indicators of cyber threats [15]. Sun et al. categorizes the different data sets that are available for use in this research area into seven, namely

1. Organizational report,
2. Executable,
3. Network,
4. Synthetic,
5. Webpage,

6. Social Media, and

7. Mixed type

In our project, we handle Organizational report and Network data. The Organizational report data consist of the various incident reporting sources (more in Section 3.2.1), while the Network data includes the host and IP address information (more in Section 3.2.2).

On the Mismanagement and Maliciousness of Networks [1]

Zhang et al. conducted an analysis on how mismanaged networks are related with maliciousness [1]. They define mismanagement as the failure to adopt commonly accepted guidelines or policies when administrating and operating networks. The symptoms of interest [15] are

1. Open DNS Recursive Resolvers: poses a threat to the networks through exploitation in an amplification attack

2. DNS Source Port Randomization: randomizing source ports can prevent DNS cache poisoning to some extent, hence, source ports without randomization are considered misconfigured

3. Consistent A and PTR records: RFC standards state every Address (A) record should have a matching Pointer (PTR) record. Lack of this is considered a symptom of a mismanaged network.

4. BGP Misconfiguration: short-lived routes were detected in the Route Views project. These are used to signal BGP misconfiguration

5. Egress Filtering: networks without egress filtering are considered as misconfigured

6. Untrusted HTTPS Certificates: untrusted certificates found in the process of ZMap are used as signs of misconfiguration

7. Open SMTP Mail Relays: Open mail relays are easily abused by spammers since they do not filter messages before sending to any destination

Table 2.1: Reputation blacklists

CBL	SBL
SpamCop	WPBL
UCEPROTECT	SURBL
PhishTank	hpHosts
Darknet scanners list	DShield
OpenBL	BRBL

8. Publicly Available out-of-band Management Devices: publicly available management cards pose severe security risks and are considered misconfigured

They leverage 12 global blacklists (Table 2.1) based on spam, phishing, malware and scanning activity to infer network maliciousness. They use a Spearman's rank correlation test [20, 21] to quantify the relationship between symptoms and maliciousness. They combine these symptoms into an overall mismanagement metric. Their results show a statistically significant positive correlation (0.64) between overall mismanagement metric and maliciousness. Since the overall mismanagement metric has the strongest correlation with the maliciousness metric, they encourage researchers to consider the overall network health instead of specific vulnerabilities or symptoms [1, 15]. The authors have further shown that, by using Fast Causal Inference (FCI) [22], an inferred casual relationship exists between mismanaged networks and labelled malicious networks, if social and economic elements are controlled.

Our project is similar to Zhang et al.'s analysis in the following ways:

- We both do not conduct vulnerability analysis (e.g., CVE-XX results in remote code execution in this web server). Instead we look at features of a network relate to an organization's profile. This is done to answer the question what relationships exist between network management and apparent network incident susceptibility.

- We both use Spearman's correlation test because it is a nonparametric measure that does not require data from a normal distribution.

- Both of our analyzes encountered issues due to utilizing a large number of external data sources that were collected from multiple networks during multiple time frames.

Our project is different from their analysis in the following ways:

- We analyze a much larger feature space than the one described in the Zhang et al. paper. We do this by looking at all of an organization's hosts (about 1,400 features per host) that are exposed to the Internet.
- We try to eliminate any biases that might exist by selectively using mismanagement features. We do this by attempting to extract features from a host's configuration in its entirety. This is because we do not preemptively assume what specific features on a host are associated with victim organization host profiles.
- We try to mitigate an issue with their collection methodology where the coverage and time frames are inconsistent. We do this by looking at incident dates and using those as our lookup dates, effectively reducing the analysis to a network snapshot closest to the incident.
- They aggregate these misconfigured systems at the autonomous system (AS) level while we are performing organization level analysis. The authors mention that this is a limitation in their analysis [1], since an organization can have multiple autonomous systems and vice versa. We mitigate some of the challenges that are associated with a more granular organization level aggregation in hopes of eliminating this limitation.

Cloudy with a Chance of Breach: Forecasting Cyber Security Incidents [2]

Liu et al. conducted a study with a very similar motivation to our project [2]. Their goal was to proactively predict security incidents, such as those referenced by Verizon in its annual Data Breach Investigations Reports (DBIR), using externally observable properties of an organization's network. They analyzed a victim data set of more than 1000 security incident reports which was comprised of 700 incidents from VERIS [23], 300 incidents from Hackmagedon [24], and 150 incidents from Web Hacking Incidents Database [25] ranging from mid-2013 to 2014. Their analysis built the non-victim data set using entries gathered from the regional internet registries. These two groups form the ground truth for their analysis. From these groups of organizations, the authors setup the research problem as a binary classification task to identify whether an organization will encounter

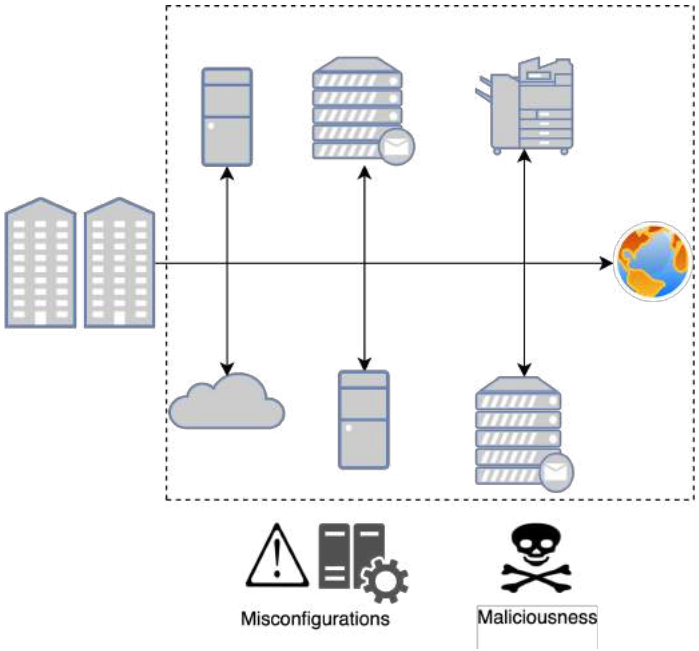


Figure 2.1: Contemporary external network posture using misconfiguration and maliciousness [1, 2]

a security incident based on external network posture information (as opposed to internal network information). To answer this research problem, they collected data that captures different aspects of a network's security posture, ranging from the behavioral (externally observed malicious activities originating from a network) to relational (such as misconfigurations in a network that deviate from known best practices, See Figure 2.1).

Their feature space comprises of 258 diverse set of externally measurable features of a network's security posture. These are : 1 organization size feature, 5 misconfiguration features, 180 raw time series maliciousness features, and 72 analyzed time series maliciousness features. The organization size, similar to this project, is the number of IP addresses within that organization's aggregation unit. The five misconfiguration / mismanagement symptoms (a subset of Zhang et al.'s misconfiguration features [1]) are measurements on a network's misconfigurations or deviations from standards and other operational recommendations. These are

1. Open DNS Resolvers,

2. DNS Port Randomization,

3. BGP misconfiguration,

4. untrusted HTTPS Certificate, and

5. Open SMTP Mail Relays

These misconfiguration features were collected from the Open Resolver Project (Open DNS Resolvers), Route Views Project (BGP misconfiguration), and previous works [1] (DNS Port Randomization, Untrusted HTTPS, and Open SMTP Server). The time series malicious features were collected from reputation blacklists (Table 2.1) as boolean features that quantify whether or not a certain IP address has been associated with spam, phishing, or malicious scanning activity. These are measurements of malicious activities seen to originate from that network, similar to Zhang et al.'s work in [1]. Their results outperform results to date with 90% TPR, 90% accuracy, and 10% FPR.

Our project is similar to their analysis in many ways including:

1. We both verify the incident contexts and dropped the incidents that were unrelated to cyber security.

2. We both use a Random Forest algorithm for classification since it is known to work well with large and diverse feature sets.

3. We both use an ROC curve to find the optimal operating point.

Our project is different from their analysis in the following ways:

1. Although, their feature space captures some of the synchronized and dynamic behavior over time within an organization, it does not provide a thorough list of services and hosts that are associated with security incidents. In this project, we use a feature extraction engine that represents 26 different protocols that could be present on a certain host. Hence, we are capable of providing over 6,000 features (in this project, we only analyze 1,400 of these) from these 26 protocols in an attempt to better identify what configurations are associated with victim organizations' hosts.

2. They utilized reputation black lists (RBLs) in their analysis. However, we do not use RBLs for the following reasons:

   - Presence of dependency for external security researchers to report malicious IP addresses. Moreover, these lists need to be updated frequently.
   - For organizations (that are not part of the study) that wish to calculate their own risk profiles, they would need all 12 RBL sources along with the model to run this against their own network. However in our approach, an organization would only require our model and the features from their own external network.
   - Shared hosting environments may blacklist a shared IP so attribution errors can be introduced.

   This RBL dependency is not a big deterrent since the threat intelligence community is keeping the lists up-to-date, but in our analysis we eliminate this dependency altogether. We do not expect our model to perform better than one that uses RBLs since it is much easier to predict a security incident if hosts on your network have been associated with malicious activity. Instead, we attempt to predict organization's susceptibility to report a security incident by profiling the organization's internet hosts.

3. Liu et. al conducted two different prediction scenarios, namely short-term and long-term prediction. These scenarios are separated based on how far back we look at the time series features (14 days for short-tem and 60 days for long-term prediction). We do not conduct a different temporal based prediction, since we do not collect time series features.

4. In their asset attribution step, they map an organization reported in an incident to a set of IP addresses through a sample IP address of the organization involved in the incident. These sample IP addresses were obtained by manually processing each incident report. One common sample IP address was the website of the organization involved in the incident. This sample IP address is used to query the Regional Internet Registry (RIR) for the public IP address range / prefix, which is then attributed to the organization. However, this is a flawed

approach as the sheer amount of public cloud providers would make it unreliable. They also mention that inclusion of such cases is a tradeoff as excluding them would have left too few samples to perform a meaningful study [2]. This is a sign that cloud based resources are not negligible edge cases but rather a big part of the internet that need to be accounted for. In our analysis, we will attempt to improve this method by including contemporary reconnaissance tactics [17, 26, 27] in combination with manually verified ARIN lookups of digital assets. This enables us to accurately locate resources on the public cloud and reduce bias away from large organizations that have their own ARIN blocks. It also allows us to locate IP address blocks under multiple owner IDs, and effectively map them to the same organization. Moreover the inclusion of cloud provider IP addresses that are associated with an organization's domain enables us to see their network under the same eyes as a threat agent.

5. The training-testing split for their analysis was done chronologically, where earlier incidents were used to train a model for future ones. However, due to difference in data set sizes (our data size being smaller), our analysis employs 5-fold cross-validation. One direction for future work is to use our analysis for a similar chronologically separated prediction.

6. An analysis that Liu et al. conducted, but still remains a direction for future work [15], is around adversarial machine learning. An organization might exist that has exemplary network configurations but has reported a security incident, and vice versa. This would normally be an impact to the analysis but the authors have shown that this error can be ignored in their analysis. We do not conduct a similar analysis but this is a good direction for future work, depending on whether one can identify the ground truth with a certain confidence value.

Other relevant incident prediction works

Other malicious activity prediction studies include

1. Sarabi et al. examine the extent that business details about an organization can help forecast its risk of experiencing different types of data incidents [28]. They show that it is difficult to assert with certainty the types of incident an organization is likely to face.

2. Liu et al. applied a Support Vector Machine to a set of reputation blacklists to generate predictions for future security incidents that may happen to a network [29].

3. Vasek et al. analyzed features from sampled web servers to identify risk factors for web server compromise [30].

4. Thonnard et al. looked at organization risk factors (number of employees and business sector) and individual level factors (job type and location) that are related with experiencing spear phishing targeted attacks [31].

5. Canali et al. analyzed user browsing behavior to predict whether a user will encounter a malicious page achieving 87% accuracy [32].

6. Edwards et al. have shown that that neither size nor frequency of breaches has increased over the last decade [9]. They combine two different cost models to project that in the next 3 years breached could cost north of 179 billion USD.

7. Aditya et al. presented a tool, RiskWriter, to assess the internal security posture of an enterprise using only external and business data [11].

8. Sarabi et al. present a framework to convert Censys scan data to a more numerical and low-dimensional representations [33]. Although their approach is similar to ours, we do not use a variational autoencoder (an unsupervised neural network model) to analyze the hosts. We use a Random Forest to conduct our analysis, and identify features that are important in the classification stage.

9. Soska el al. apply machine learning to predict whether a web site may turn malicious, and show that their method can achieve 67% true positive and 17% false positive [34].

10. Qian et al. and Ramachandran et al. used machine learning concepts to identify SPAM emails [35, 36].

2.3 Industry cyber risk analysis

In the Industry, cyber risk is defined as "any risk of financial loss, disruption or damage to the reputation of an organization from some sort of failure of its information technology systems" [37]. Before the wide-spread usage of cyber data analysis, most insurers focus on industry class (banks, healthcare, education) as well as revenue and number of data records held to evaluate cyber risk. Moreover, the industry-standard practice was to send security questionnaires for prospective customers to fill out [12]. Now, cyber risk analysis is being provided as an automated service by many companies including:

1. FICO ESS [12],
2. BitSight [13],
3. SecurityScorecard [14], and
4. UpGuard [38]

FICO's ESS is the industry method that is most similar to our analysis, as it has its origins from Liu et al.'s [2] work. FICO's Cyber Risk Score relies on a diverse set of Internet scale security signals to determine the risk profile of any organization. This information is used to train a machine learning model that produces a risk score that forecasts the likelihood of a future breach event [12].

SecurityScorecard analyzes the cyber risk of an organization using outside-in data. The platform gathers security data and grades organizations from A to F across ten security categories, namely

- Web Application Security
- Network Security
- Endpoint Security
- DNS Health
- Patching Cadence
- Hacker Chatter

- IP Reputation
- Leaked Credentials
- Social Engineering
- Cubit Score

SecurityScorecard's goal is to discover organizations' external security posture from the point of view of: a hacker, a business partner, or a customer.

BitSight uses externally observable data on compromised systems, security diligence, user behavior, and public disclosures to compute an organization's security rating. BitSight's Security Ratings are comprised of two types of data, namely

1. Event Data: evidence of botnet infections, spam messages, malware servers, unsolicited communication, and other indicators of compromise
2. Diligence Data: information about security diligence practices, such as SSL, SPF, and DKIM configurations

BitSight does not use the following types of information in the analysis

- Budget
- Franchise Locations
- Beta Risk Vectors (DNSSEC and Disclosed Credentials)
- Compliance

UpGuard uses most of the above same data types but also monitors so-called hacker chatter (e.g., social networks).

In addition to network security, these cyber risk services are also being used to explicitly require business partners to have a certain level of cyber insurance coverage [12]. However, an outstanding issue still present in Academia and Industry is how to effectively attribute digital assets [2, 1]. Most of the industry professionals attempt to solve this by manual confirmation of an organization's

internet-visible network assets. There is no obvious automated method to gather an organization's digital resources, however, in our analysis we attempt to use contemporary reconnaissance tactics [17, 26, 27] to mitigate this issue.

Chapter 3: Data and design

3.1 Development tools

There are numerous development tools that are used in this project. These tools are used to write the source code, aid in the analysis, and store the large amounts of data.

Some of these tools are:

- Python2.7 [39]: This is the main programming language used in this analysis. The immense amount of community support and variety of libraries made this the front-runner among contemporary languages.
- PyCharm [40]: This is the integrated development environment that is used for this project. The remote debugging assistance feature aided in running the code across the remote servers.
- Elastic Stack [41]: This is a search engine that is invoked with an HTTP API. This is used to search through, store, and visualize the large amount of data that is used in this analysis.

3.1.1 Data science libraries

Since the main programming language used for this project is Python 2.7 [39], the host learning library the project used is scikit-learn [42]. This library abstracted out the implementations in algorithms like Isolation Forest [5] and Random Forest [43] that are used in this analysis. The data manipulation libraries seaborn [44], pandas [45], numpy [46], and matplotlib [47] are also heavily utilized during the large scale data analysis.

3.2 Data collection

The research problem requires a large sample of data combined with an in depth analysis. This is accomplished through a selection of data sources that are interfaced through APIs and invoked using tools. A data source is information that is utilized to accomplish a certain task. The data sources mentioned in this project aid in identifying organizations of interest, attributing hosts to

these organizations, and collecting the attributed hosts. An Application Programming Interface (API) is a set of functions and subroutines that are used for building software; essentially providing an abstraction from the underlying implementation details. These APIs are used to interface the above mentioned data sources. A tool, in the context of this analysis, refers to a utility or a piece of code that invokes an API or taps into a data source. There are many more data sources, APIs, and tools outside of the ones mentioned in this chapter that could be used to address the relevant research problem. The ones below are selected on the basis of integration capability, however, the project can scale to new data source, APIs, and tools with minor changes.

The above concepts are combined to form the data collection pipeline as seen in Figure 3.1. The pipeline is broken down into the following steps:

1. Cohort Selection
2. Host Attribution
3. Host Collection
4. Feature Engineering

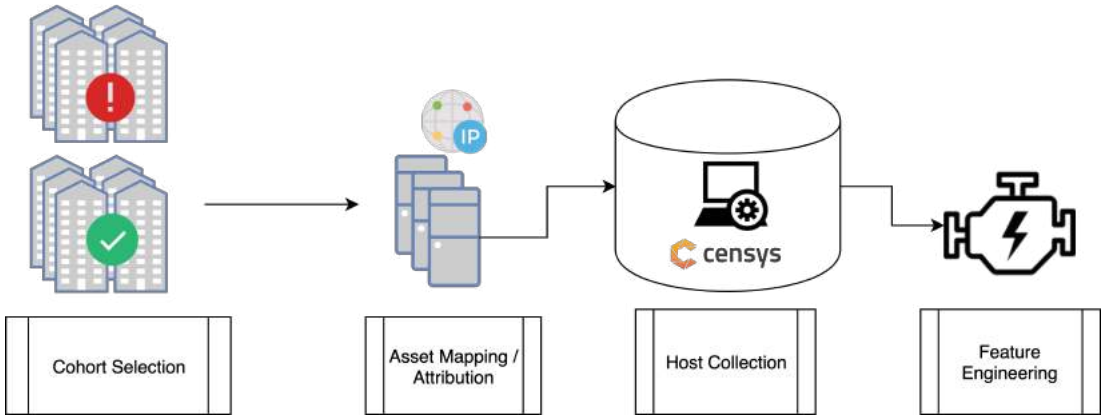


Figure 3.1: Pipeline that collects large scale victim and non-victim data to map, collect, and extract features from their assets

3.2.1 Cohort selection

The time range utilized in this analysis is January 1, 2017 - January 1, 2019. This is the time window that we used to define an organization as a victim or non-victim. This range will be referred to as

the (incident) report window. In this phase of data pipeline, victim and non-victim organizations are located to form a cohort within this time range. The cohort set consisted of one victim and three non-victim subsets, namely,

- Security Incident organizations / victims (BREACH)
- Cyber Security 500 non-victims (SEC500)
- DNS Sampled (reverse IP address) non-victims (DNS)
- CERT Sampled (certificate subjects) non-victims (CERT)

Victim subset

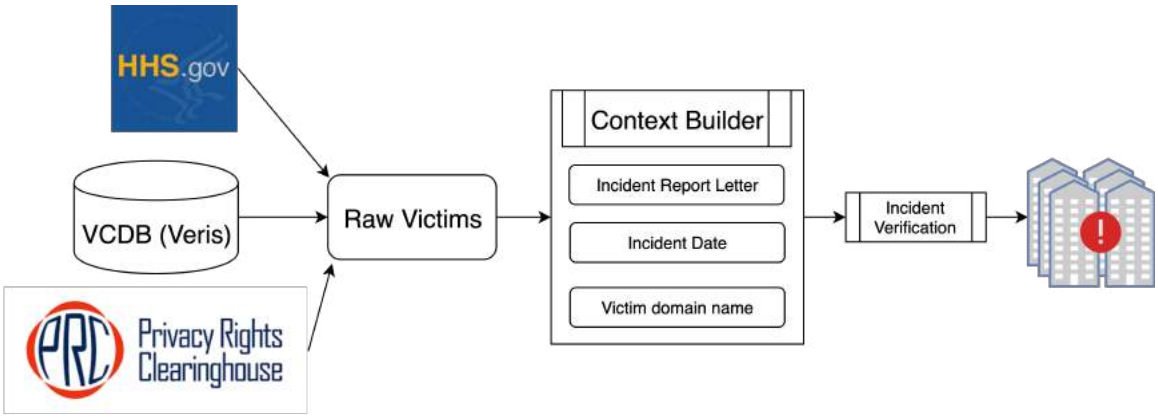


Figure 3.2: Victim organization collection pipeline

Many existing security incident data sources can be used to locate victim organizations. Security incident data sources contain ground truth security incident information including the organization that reported the security incident, the report date, and many more fields that add context to the incident. Among the many such sources that provide this information, three are selected on the basis of update frequency, data format, ease of access, and overall integration capability. The selected security incident data sources are

- Privacy Rights Clearinghouse (PRC)
- Veris Community Database (VCDB)
- Department of Health and Human Services (HHS)

There is a bias that occurs if only one of the above sources is selected. e.g., HHS only requires breaches that have more than 500 affected records. The presence of three different security incident data sources reduced this bias towards any particular reporting agency, similar to [2].

Similar to Liu et al. [2], our incident selection process is a conservative one, since all three victim data sources contain incidents that are irrelevant to the research problem, e.g., possible data breach through theft of organization laptop, or physical break in. Hence, there is a need to filter the irrelevant incidents from the data source before further analysis. This filter is applied as criteria that incidents have to meet to be included in the analysis. These criteria are grouped into general (common to all the subsets) and specific (unique to the data source). The general criteria for the three victim subsets are:

- Incident report date is within the report window (January 1, 2017 - January 1, 2019)
- Incident is reported in the United States. This meant the project only looked at security incidents that have a submitted report to a U.S. governing body.
- Incident cause is some variant of a network intrusion
- Repeat incident reports from the same organization are not considered as separate victim entries. This means organizations that report a security incident on multiple occasions count as one victim entry with the earlier reported security incident date. This is done to ensure there is no biasing for organizations that have more than one security incident in our report window. Moreover, sometimes the second incident report is a continuation of the first incident. In addition to this, this project is not looking at a non-victim organizations more than once, so it would be "unbalanced" to repeat victim organizations.

Privacy Rights Clearinghouse (PRC)

Privacy Rights Clearinghouse is a California based nonprofit organization that is focused on protecting people's privacy [7]. PRC has a data source that spans the categories of Identity Theft, Fraud, Banking, and Finance issues as well as reported data breaches dating as far back as 2005. At the current time, this subset contains 3,042 entries in the report window. For every incident,

PRC contains the date the incident was made public, the name of the entity responsible for the data, the type of entity involved, a classification of the type of incident, the total number of records, the location (city and state) where the entity operates, information on the source of the data, and a short description of the incident [9, 7]. Among all these incidents, the cyber security / security incident entries are collected through CSV format.

The PRC-specific selection criteria are:

- Source is not media. Media-based sources are unreliable as there is no viable report to verify a real incident.
- Incident Type is 'Hacking': The goal of this study is to focus on organizations that had comparatively poor network stature. Looking at 'Physical Theft' and other non-network posture related incidents would not be relevant to this analysis.

Some sample PRC incidents can be seen in Table 3.1

Table 3.1: Sample PRC incident entries

Field	Sample 1	Sample 2
Date Made Public	June/12/2018	October/1/2018
Company	University at Buffalo	Chegg
City	Buffalo	Santa Clara
State	New York	California
Type of breach	HACK	HACK
Type of organization	EDU	EDU
Description of incident	WIVB4 is reporting...	According to a filing ...
Information Source	Media	Government Agency
Source URL	http://...	https://...
Year of Breach	2018	2018

Vocabulary for Event Recording and Incident Sharing Community Database (VCDB)

VERIS Community Database (VCDB) is a breach aggregation portal with an aim of providing a public data set of security incidents that is capable of supporting community research [23]. It collects and disseminates information for all publicly disclosed data breaches. The sources that VCDB pulls from include the Department of Health and Human Services (HHS) incidents, the sites of the various Attorneys General that provide breach notification documents, media reports, and press releases. At the current time, this subset contains 8,158 entries in the report window. The data is provided

in a model that lends itself to ease of data manipulation and transformation. This data model is available in JSON format from the GitHub repository [23].

The VERIS-specific filter criteria are:

- Environmental and physical related security incidents are dropped. These are not related to the research problem at hand.
- Security incidents that did not have a incident date are dropped.

A sample VCDB entry can be seen in Listing 3.1.

Listing 3.1: Sample VCDB entry

```
{
    "action": {
      "hacking": {
        "variety": [ "Unknown" ],
        "vector": [ "Web application" ]
      },
      "malware": {
        "variety": [ "Capture app data" ],
        "vector": [ "Direct install" ]
      }
    },
    "actor": { "external": { "motive": [ "Financial" ] } },
    "asset": { "assets": [ { "variety": "S - Web application" } ] },
    "attribute": {
      "confidentiality": {
        "data": [ { "variety": "Payment" } ],
        "data_disclosure": "Potentially",
        "data_victim": [ "Customer" ]
      },
      "integrity": { "variety": [ "Software installation" ]}
```

```
        },
        "discovery_notes": "Ext - Unrelated third party. Discovered by
            security researcher who made the notifications.",
        "incident_id": "00539A80-EC99-4E3E-81BA-EBAB8B2FE41E",
        "reference": "http://www.pcworld.com/article/3131040",
        "schema_version": "1.3.3",
        "security_incident": "Confirmed",
        "source_id": "vcdb",
        "summary": "Online skimmers ....",
        "timeline": { "incident": { "year": 2016 } },
        "victim": {
          "industry": "44",
          "victim_id": "everythingyamahaviking.com"
        }
      }
```


Department of Health and Human Services (HHS)

As required by the HITECH Act [48], the Department of Health and Human Services must post a list of data breaches of unsecured protected health information affecting 500 or more individuals. This list contains all the breaches that have been reported to the Secretary since 2013 [49]. At the current time, the subset contains 418 security incidents in the report window.

The HHS-specific filter criteria are:

- Data Breach is of type Hacking/IT Incident.
- Incident Location is not 'Paper/Films'.

This data set is available for public usage through CSV format, and a few samples entries can be seen in Table 3.2.

After collecting and aggregating the security incidents, they are put through a set of verification steps. During the following steps, the incidents not matching the verification criteria are dropped.

Table 3.2: Sample HHS data breach entries

Entity	State	Type	Count	Date	Type of Breach	Location	BA
UW Medicine	WA	H.care	9730	02/20/2019	Hacking/IT	Server	No
Fred Eaglstein	FL	H.care	2000	05/30/2017	Unauth Access	EMR	No

1. Ensure that the incident contains an organization name. For example, "Identity theft at a car dealership" is invalid

2. Organization name is associated with a valid domain name. The two search engines that are used in the domain attribution process are Microsoft's Bing [50] and Google [51]. These engines are used to identify domain names for organizations and vice versa. We perform a search [51, 50] using the organization name to locate the domain name. This approach, by itself, is not always accurate. However, when combined with manual verification, it provides a reliable and efficient way to associate organizations to domains Some organizations did not have a domain name, e.g., route paths (https://www.example.com/organization_name) would not be associated with "organization_name" but would be "example.com". Organizations that did not have a viable domain name are dropped, as there is no reliable way to attribute digital resources to a organization that did not have a domain name.

3. Once the domain name is located, we perform a quick Censys search for the domain. The incident / domain is dropped here if there are no matches. The assumption here is that if an organization does not have data in Censys today, then it is not likely to have data for the reported incident date. This might not be a valid assumption, since organizations might still have digital resources, even if they do not have a domain name, or even if Censys does not have that domain but has IP addresses that are associated with that organization. For this project, these incidents are dropped, however a possible direction for future work is to include organizations that do not have domain names.

After the above steps, this project had collected 373 incidents that met the filter criteria. We then proceed to verify the security incident and collect more in-depth incident information for each entry. Context is increased by a viable incident report that is submitted to a United States

government entity (HHS, residing state's attorney general's office, etc.) Sources like HIPAA journal and databreaches.net are also considered viable sources of security incident confirmation. The steps taken to increase the incident context are:

1. Step 1 : Locate the source url of the incident report. This step mitigates the false report issue, whereby a incident is erroneously entered in one of these data sources.

2. Step 2 : Locate the incident report letter. The incident report contains a sample notification letter sent to individual's whose information is suspected to have been exposed. The sample letter or the incident report briefly states what lead to a security incident, how many records are affected, and the important dates of the incident, among other things. A sample notification letter can be seen in Figure A.1.

3. Step 3 : Verify the incident type. We ensure that the incident type is one that matches this project's interests. The types of incidents this project is interested in are phishing scams, malware attacks, and misconfigurations.

4. Step 4 : Identify the attack's main point of entry. The organization that reported the incident needed to be the main target in the incident, e.g., organization X is reporting a breach because one of its partner's, organization Y, is compromised and attackers gained access to our system through secure infrastructure setup with organization Y. This is invalid because external network posture information would not reveal anything relevant to this internal communication pipe. Hence, organization X would be dropped but organization Y would be used in the analysis.

5. Step 5 : Extract important dates. From the incident report or notification letter, we can pull the following dates:

    - Incident start date: when the incident is suspected to have started.
    - Incident discovery date: when the incident is discovered.
    - Incident end date: when incident is suspected to have ended. This is usually the same as the discovery date.

- Incident report date: when the report and incident letter are sent out

6. Step 6 : Identify the lookup date. As Liu et al. [2] have mentioned, it is very important that the features used to build the classifier/predictor reflect the condition of a network prior to the incidents. Hence, we need to select a lookup date to view the network posture snapshot accordingly. This date would need to be close to the incident occurrence date as possible to accurately represent the network state. Since the reported incident start date is the closest to this occurrence date, the incident start date is used as the lookup date. However, if we are not able to locate the incident date, the discovery date is a valid substitute.

7. Step 7 : Remove duplicate incidents and organizations. Verify that multiple data sources did not result into two duplicate entries by a manual verification step.

The report window, as previously mentioned, is the date range for which we scoped the reported security incidents. The incident window is the date range for which the incidents occurred in the victim organizations. Although the grace period (required time to report a breach) is different depending on which state the organization resides, the data affected, and the type of business the organization conducts, the general interval is 45-60 days after discovery. Above, we mentioned that a requirement for all incidents is that the report date is in the two year period. Hence, the incident window lasted 22 months; the last two months did not have data points due to the incident report grace period.

After all the above steps, there are now a total of 263 incidents. Of these entries, there are 176 PRC incidents, 84 VERIS incidents and 3 HHS incidents that match both the general and specific criteria. From these data points, we randomly sampled 200 for this study.

Non-victim dataset / organizations

This project is analyzing the traits among victim organizations, and to compare these organizations against non-victim organizations. Hence, there is a requirement to collect a list of non-victim organizations. Moreover, it is important to collect an unbiased list of non-victim organizations to ensure a sound analysis. To counter this bias, non-victim organizations are collected through three

different methods, as seen in Figure 3.3. These methods are:

1. Sampling from the CyberSecurity Ventures' Security 500 (SEC500) [16]

2. Sampling subjects from digital certificates (CERT)

3. Sampling domains from reverse IP address resolutions (DNS)

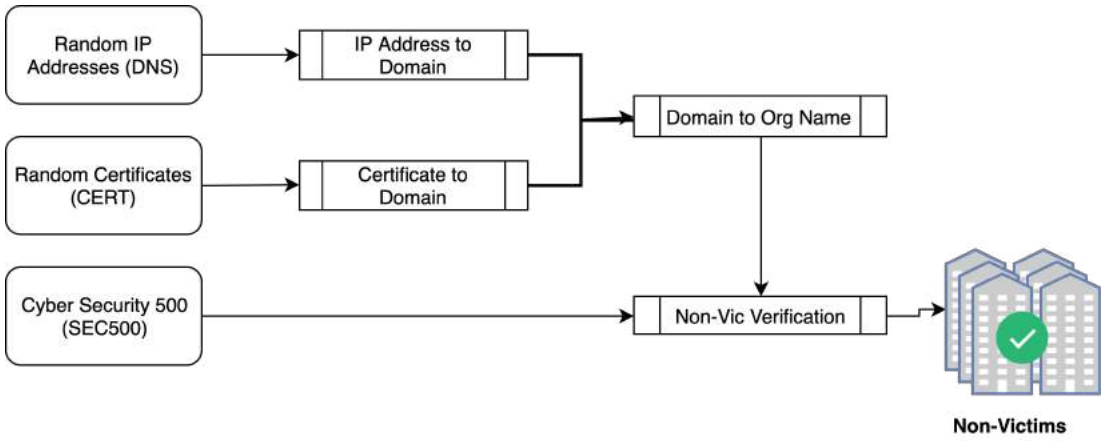


Figure 3.3: Non-victim organization collection pipeline

CyberSecurity Ventures Security 500

CyberSecurity Ventures is a security magazine that annually selects the top 500 organizations in the field of Cyber Security. The criteria for the selection and ordering of the cyber security 500 include customer base size, notable implementations, product reviews, and many others seen in [16]. This list contains security organizations in many sectors including, but not limited to, Web Application Security and Advanced CyberThreat Detection. A sample of the domain names that are present in this list are used as a subset of non-victim organizations.

Sampling digital certificates

The second method of identifying non-victim organizations is to sample subjects from digital certificates. A uniformly sampled batch of 15k IP addresses are collected from a Censys [4] table containing 160 million IP addresses. From this batch, the digital certificates on ports 25 (SMTP), 110 (POP3), 143 (IMAP), 443 (HTTPS), 1433 (MSSQL), 1521 (ORACLE), 3306 (MySQL), and

5432 (POSTGRES) are used to locate certificate subject fields. A domain name is then extracted from the subject of the certificate.

Sampling reverse IP address domains

The batch from the second method is utilized to randomly sample a host and reverse lookup the IP address. This would immediately return a domain name on the public IPV4 space.

Using each of the three methods above, we collect a set of domain names. Each domain name is then associated with a non-victim organization. Most of the same selection criteria made from the victim subset hold for these subsets too. However, there are some that are unique to the non-victim organizations. These criteria are:

- Sample organization is not a cloud provider. It is difficult to define the ownership boundary if we were to include cloud providers in our analysis, e.g., if we were to sample an AWS IP address, we would not be able to differentiate Amazon's personal resources from their clients' assets.

- Organizations that had not reported a security incident in the incident report window. This is a simple check to see if the domain name is not in the victim data set.

- The non-victim organizations need to be U.S.-based entities. This is because we retrieved the reported incident label from U.S.-based reporting agencies (HHS, PRC, and VERIS). This does not mean all of the organization's assets are based in the U.S., but rather the organization has a headquarters in a U.S. state. However, this is a very cloudy restriction, and very hard to enforce. Hence, it is a soft restriction where if a organization had a major office in the U.S. then it would also be liable to report an incident report to a U.S.-based governing entity.

After these automated filters, the non-victim organizations are put through a manual verification step. Once we have the set of domains, we proceeded to increase the organization context by adding

- Organization Name: Bing / Google search using the domain

- Date: For each non-victim organization, we assigned a random lookup date within the incident window.

We selected 200 for each method above. At this point we have collected a cohort of 800 (200 victim, 600 non-victim, 785 unique) organizations, where each organization contains a name, one sample domain and a lookup date. We will then proceed to attribute these organizations with their public digital assets.

3.2.2 Host attribution (Footprinting)

Footprinting or Host (Asset) attribution is the process of finding digital assets associated with a certain organization. In the space we are concerned with, the public internet, footprinting is taking an organization name and retrieving a domain name and a set of public IPv4 addresses. A custom component, seen in Figure 3.4, is built that collected, curated, and stored the digital assets for the cohort.

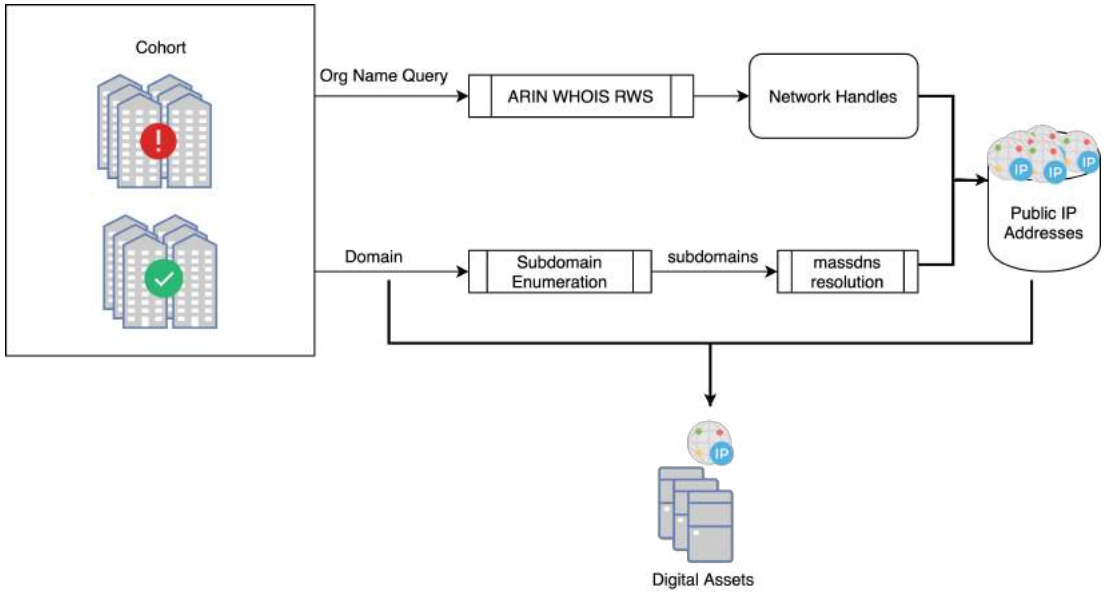


Figure 3.4: Asset attribution stage to collect a company's resources through two different methods

This component performed the following tasks:

1. Identify a domain name for a organization

2. Enumerate subdomains for the domain name

3. Resolve the subdomains to a set of IP addresses
4. Perform WHOIS lookup using the organization name to retrieve a set of IP ranges
5. Aggregate the IP addresses from the above two steps
6. Store the IP addresses along with the domain name

The component does not return an exhaustive list of assets belonging to an organization, but is rather a best effort approach to gaining insight into most of the organization's digital resources. Moreover, the tools the component uses are the same ones penetration testers, ethical hackers, and malicious actors use for reconnaissance [52] [53]. The footprinting data sources are:

- American Registry for Internet Numbers (ARIN) [3]
- Subdomain enumeration data sources

American Registry for Internet Numbers (ARIN)

ARIN is a non-profit Regional Internet Registry [3] for Canada, United States, and North Atlantic islands. The registry has numerous tasks, one of which is to the distribute IP addresses and Autonomous System (AS) numbers across these regions. ARIN "maintains a database that contains detailed records of which resources have been allocated and assigned, as well as which organizations and POCs are authoritative over those resource records" [3]. This data source can be used to conduct organization name-based searches for IPv4 ranges (network handles).

ARIN provides access to its network information though a RESTful Web Service (ARIN WHOIS-RWS). A custom tool is used to interface the ARIN WHOIS-RWS API to search for IP address ranges that are owned by an organization. The tool takes in a search term as a parameter and returns a list of ARIN network handles (IP address ranges) that correspond to that organization. The search terms/queries are formulated through permutations of the organization name. The search terms are compared against the following in ARIN's database:

- ASN name
- Organization name

- Customer name

The results from these are manually verified to make sure the results of the searches are valid. e.g., '*drexel*' would return both 'BILL DREXEL' and 'Drexel University'. We would filter out the IP addresses for all 785 organizations. These resulting IP ranges (ARIN network handles) are attributed to organizations. A sample ARIN IPv4 Range (network handle) can be seen in Figure 3.5.

Network	
Net Range	173.14.79.248 - 173.14.79.255
CIDR	173.14.79.248/29
Name	DREXELUNIVERSITY
Handle	NET-173-14-79-248-1
Parent	CBC-SACRAMENTO-11 (NET-173-14-64-0-1)
Net Type	Reassigned
Origin AS	
Customer	DREXEL UNIVERSITY (C03034328)
Registration Date	2012-06-11
Last Updated	2013-12-09
Comments	
RESTful Link	https://whois.arin.net/rest/net/NET-173-14-79-248-1
See Also	Upstream network's resource POC records.
See Also	Upstream organization's POC records.
See Also	Related delegations.

Figure 3.5: Sample ARIN lookup [3]

Subdomain enumeration data sources

At this stage, we associate a set of subdomains for each domain name. e.g., if the domain is 'ibm.com', some possible subdomains are 'research.ibm.com' and 'dev.ibm.com'. This procedure of finding and assigning a set of subdomains to a domain is called subdomain enumeration.

There are numerous data sources used in the identification of subdomains for a particular domain. These data sources are tapped into either directly, through use of an API key, or indirectly, through a footprinting tool. The data sources whose API is directly invoked are tabulated in Table 3.3

Table 3.3: List of directly invoked subdomain enumeration data sources

RiskIQ (PassiveTotal) [54]	Binary Edge [55]
Security Trails [56]	VirusTotal [57]
DNSDB (FarSight Security) [58]	Shodan [59]
Google [51]	Riddler [60]
Bing [50]	crt.sh [61]
Censys [4]	

Table 3.4: List of subdomain enumeration data sources with research access

RiskIQ (PassiveTotal) [54]	Binary Edge [55]
Security Trails [56]	VirusTotal [57]
Censys [4]	

Some of the direct API data sources afforded this project research access to perform large amounts of subdomain lookups. This significantly helped the analysis by allowing our team to perform large scale subdomain enumerations in a very passive manner. The services that use research access can be seen in Table 3.4. They are implemented through OWASP's Amass [62] which will be covered in Section 3.2.2.

The exhaustive list of external data sources and search engines used in footprinting, both directly through an API invocation or indirectly though a tool's interface can be seen in Table 3.5.

Table 3.5: Exhaustive list of footprinting data sources

Archive.is	ArchiveIt	ArchiveToday	Arquivo	Ask
Baidu	BinaryEdge	Bing	BufferOver	CIRCL
Censys	CertDB	CertSpotter	CommonCrawl	CrtSearch
Crtsh	DNSDB	DNSDumpster	DNSTable	Dogpile
Entrust	Exalead	FindSubdomains	Google	HackerTarget
IPv4Info	LoCArchive	Netcraft	OpenUKArchive	PTRArchive
PassiveDNS	PassiveTotal	ReverseDNS	Riddler	Robtex
SecurityTrails	Shodan	SiteDossier	ThreatCrowd	ThreatMiner
Twitter	UKGovArchive	URLScan	Umbrella	VirusTotal
Wayback	Yahoo			

These tools used for footprinting are selected from numerous other possibilities due to the span of techniques and services. The reason for multiple tools is to avoid biased results from just one tool. Moreover, there is no one tool that uses all the standard subdomain enumeration techniques. These footprinting tools are:

1. Amass [62]
2. dnsrecon [63]
3. Sublist3r [64]
4. SubFinder [65]
5. assets-from-spf [66]
6. domains-from-csp [67]

Amass

Amass obtains the subdomain names for a certain domain by scraping data sources, tapping into web archives, and reverse DNS sweeping. It also uses the IP addresses obtained during resolution to discover associated network ranges and autonomous systems [62]. This tool is also used with the research APIs.

dnsrecon

dnsrecon is a Python port of an older Ruby project and is a common tool among many ethical hackers [53, 52]. It also comes pre installed on many offensive operating systems. It uses techniques like wild card resolution and DNS server cache lookups, among many other basic approaches, to identify subdomain names [63].

Sublist3r

Sublist3r [64] enumerates subdomain names for websites using search engines and numerous other data sources. It is an established tool, much like dnsrecon [53] [52]. One thing to note is that no brute forcing is conducted with this tool, or with underlying subbrute.

SubFinder

SubFinder is subdomain discovery tool that uses Passive Sources, Search Engines, Pastebins, Internet Archives, etc to find subdomains [65]. It has a simple modular architecture and is foreseen as a successor to the Sublist3r project. SubFinder complies with the passive sources' licenses, and usage

restrictions, as well as maintains a consistently passive model to make it useful for penetration testing [65].

assets-from-spf

Sender Policy Framework (SPF) is an email authentication method designed to detect forged sender addresses in emails (email spoofing), a technique often used in phishing and email spam. SPF allows the receiver to check that an email claiming to come from a specific domain comes from an IP address authorized by that domain's administrators. The list of authorized sending hosts and IP addresses for a domain is published in the DNS records for that domain. This small script [66] parses network ranges and domain names from SPF DNS record.

domains-from-csp

Content Security Policy (CSP) is an added layer of security that helps to detect and mitigate certain types of attacks, including Cross Site Scripting (XSS), and data injection attacks. This small script [67] parses domain names from the CSP header.

An exhaustive list of information gathering techniques that these footprinting tools used is:

- General DNS Records (MX, SOA, NS, A, AAAA, and TXT)
- PTR Record lookup for a given IP Range or CIDR
- Subdomain scraping
- Service Record (SRV) Enumeration
- Snooping cached records for A, AAAA and CNAME Records
- Certificate Transparency Logs
- Wildcard Resolution
- DNSSEC zone walking
- AXFR / Zone transfers

- mDNS records enumeration
- Sender Policy Framework (SPF) records
- Content-Security-Policy (CSP) HTTP headers

It is important to note here that the attribution process did **not** include **brute-force subdomain resolution** or **alterations / permutations of already known subdomains**. This is due to the length of time needed combined with the questionable legal implications of performing this sort of search. After collecting the subdomains for each organization, we need to resolve them to their respective IP addresses. For the total cohort, we collected 2.3M subdomains. The IP address resolution tool used in the project is **massdns** [68]. This domain resolution tool enabled efficient bulk lookups. Massdns [68] takes in a list of domain resolvers and a list of subdomains as parameters, concurrently distributes the resolution load across these resolvers, then aggregates and returns the results. From the 2.3 M subdomains found, 400 K are non-resolvable. Hence we have 1.9 M IPV4 addresses that are attributed to the cohort.

3.2.3 Host collection

Censys is an organization that periodically scans the public IPv4 address space (a portion of the Internet) [4]. These scans contain host information like the services running on a host, the ports these services are running on, and detailed fields per service. These snapshot (daily scan) results are stored in BigQuery tables (exposed as an API that is leveraged through a Python Library) [4]. For example: the scan results for December 21, 2018 are stored in table '20181221'. In essence, Censys contains historical IPv4 scan information and provides a means to lookup a host's external state close to the reported security incident date. This, close to security incident date, property is extremely important to conducting a sound analysis as straying too far from this date would not provide an accurate snapshot of an organization's external network at the time when the security incident occurred. These scans can be accessed through a Web API, which handles smaller request sizes, as well as through the BigQuery API, which is capable of handling larger requests. Due to large amount of data to be collected from Censys, the BigQuery API is invoked. Hence, the host

collection component constructs queries with the domain and IP addresses for each organization, which are then used to collect the associated hosts from Censys.

Once we located a domain name and IP addresses for an organization, the queries are formed in two ways

1. Locate the host for a certain IP address: This IP address is located through either the subdomain enumeration step or the WHOIS ranges lookup.

2. Locate the host with a digital certificate containing the subject domain: The organization domain name is looked up in following fields for all the certificates available on a given host

   - certificate.parsed.names
   - certificate.parsed.subject.common_name
   - certificate.parsed.subject_dn
   - certificate.parsed.extensions.subject_alt_name.dns_names

This process of collecting the hosts can be seen in Figure 3.6

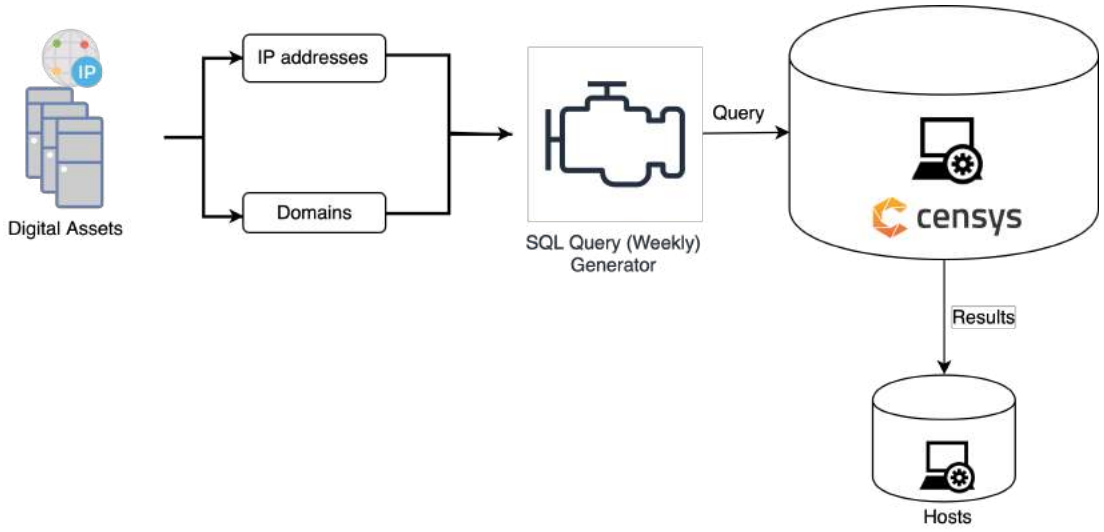


Figure 3.6: Host collection stage using Censys [4]

At this stage, the analysis has collected 714,244 hosts across all the organizations in the cohort. The distribution per cohort subset can be seen in Table 3.6.

Table 3.6: Cohort size distribution

Cohort Subset	Organizations	Hosts	Avg Org Size
VICTIM(BREACH)	199	48017	241.3
CERT	198	388552	1962.4
DNS	194	271844	1401.3
SEC500	200	55372	276.9
All	791	763785	965.6
All (Unique)	776	714244	920.4

We can see that the DNS and CERT cohort subsets had a larger average number of hosts per organization than the SEC500 or VICTIM subsets. This is because the DNS and CERT subsets are the result of sampling from the internet, a disproportionate universe. Uniformly sampling from this universe resulted in finding IP addresses that are associated with large organizations. This also means, compared to randomly sampled organizations from the internet, victims tend to have a smaller number of hosts.

There were 8 companies that did not have data points in Censys from the data sets. These were 1 from the VICTIM set, 1 from the CERT set and 6 from the DNS set. This was due to the randomly assigned historical lookup dates. Some of these companies did not have anything on the internet during the lookup time. This raises an interesting other question of the company maturation as it pertains to likelihood of data breach. We leave this as an avenue for future work. One other CERT data point had over 150K machines attributed to it. We did not have the computing power nor the time to process this data point, hence we dropped it. In total, we did not have any machines available for a total of 9 out of the 785 organizations. Seeing as this is about 1% of the data space, we believe it is safe to continue with the analysis.

A sample host from Censys can be seen in Listing 3.2.

Listing 3.2: Sample Censys host

```
1 {
2     "80": {
3         "http": {
4         "get": {
```

```
            "headers": {
            "unknown": [
                { "value": "2.0.50727", "key": "x_aspnet_version" },
                { "value": "Tue, 09 Apr 2019 10:17:47 GMT", "key": "date" }
            ],
            "x_poared_by": "ASP.NET",
            "vary": "Accept-Encoding",
            "server": "Microsoft-IIS/8.5",
            "content_type": "text/html; charset=utf-8",
            "cache_control": "private"
            },
            "status_code": 200,
            "title": "Drexel University - Date/Time View",
            "status_line": "200 OK",
            "body_sha256": "8d98748cc5d3b7f",
            "metadata": {
            "product": "IIS",
            "version": "8.5",
            "description": "Microsoft IIS 8.5",
            "manufacturer": "Microsoft"
            }}}
    },
    "443": {
        "https": {
        "dhe": { "support": false },
        "dhe_export": { "support": false },
        "rsa_export": { "support": false }
        }
    },
```

```
    "ip": "144.118.39.9",
    "updated_at": "2019-04-15T14:45:21+00:00",
    "location": {
        "province": "Pennsylvania",
        "city": "Philadelphia",
        "country": "United States",
        "postal_code": "19104",
        "country_code": "U.S.",
        "timezone": "America/New_York",
        "continent": "North America"
    },
    "autonomous_system": {
        "description": "DREXEL-ASN - Drexel University",
        "rir": "unknown",
        "routed_prefix": "144.118.0.0/16",
        "country_code": "U.S.",
        "path": [ 11537, 36412, 11834 ],
        "asn": 11834,
        "name": "DREXEL-ASN - Drexel University"
    },
    "protocols": [ "80/http" ],
    "ports": [ 80 ],
    "tags": [ "http" ],
    "metadata": { "os": "Windows", "os_description": "Windows" }
}
```


3.2.4 Feature engineering

Once the hosts are collected from Censys, a feature engineering component converted the hosts into feature vectors. A feature is a mathematically analyzable property of a phenomenon or observation.

Each feature vector should ideally be a numerical representative of the configurations that are present on each host. Moreover, there is a need to convert the nested implementation seen in Listing 3.2 into a more shallow representation. In essence, the feature engineering component maps the Censys data into a different model for mathematical analysis. This process of feature engineering is broken down into two steps:

1. Field selection
2. Feature extraction

The entire process can be seen in Figure 3.7.

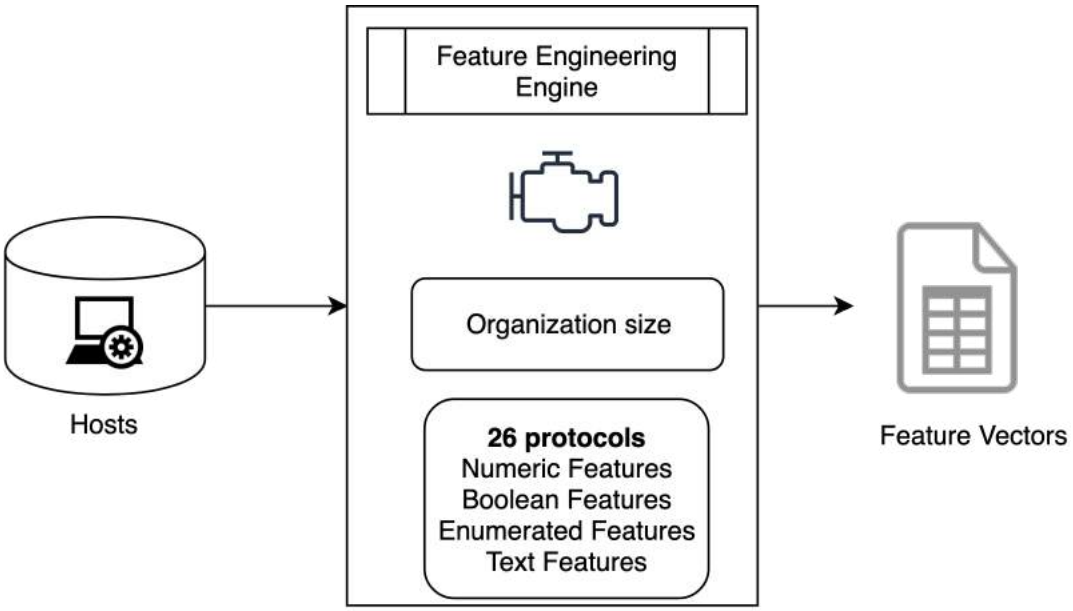


Figure 3.7: Feature engineering stage to extract features from 26 protocols

Field selection

The Censys data model, for this project, is a unique set of all the BigQuery table schemas in the incident window. This means every field in every table in the 2017 - 2018 year period was accounted for in the Censys data model. In total, this data model contained 9,899 fields across the schemas.

Although the exhaustive list can be seen in the appendix (Listing A.2), a condensed list can be seen in Table 3.7.

Table 3.7: Sample list of Censys fields

autonomous_system	ip	location	metadata	p102.s7
p110.pop3	p143.imap	p1433.mssql	p1521.oracle	p1900.upnp
p1911.fox	p21.ftp	p22.ssh	p23.telnet	p2323.telnet
p25.smtp	p3306.mysql	p443.https	p445.smb	p47808.bacnet
p502.modbus	p53.dns	p5432.postgres	p587.smtp	p631.ipp
p7547.cwmp	p80.http	p8080.http	p8888.http	p993.imaps
p995.pop3s	ports	protocols		

Feature extraction

Sun et al. state that "performance of machine learning algorithm heavily depends on the choice of features or data representation" [15]. In accordance with this statement, our feature extraction component uses the 9,899 fields available in Censys to extract up to 6,008 features. The feature types this engine handles were:

1. Numeric Fields: These are direct transfers of numeric fields from the Censys data model, e.g., the validity length in seconds for an HTTPS certificate

2. Boolean Fields: These are converted to {0, 1} for a mathematical representation, e.g., RUNNING_P22_SSH could be 1 for true or 0 for false.

3. Enumerated Fields: These are one-hot encoded for the possible values for that field, e.g., TLS_CERT_VERSION could be {1.0, 1.1, and 1.2}. These are converted to three boolean features: TLS_VERSION_1_0, TLS_VERSION_1_1, and TLS_VERSION_1_2 which have values {0, 1}.

4. Text fields: These are handled on a case by case basis. For the shorter ones, we looked at the top 10 - 20 values through the Censys report tool and treated them as enumerated fields. For the longer one, we dropped them as possible avenues for future work.

Outside of the features from the 26 protocols, we also included the following features

- ORG_SIZE: number of hosts an organization owns.

- NUM_PORTS: number of ports that are running a service on a host.

- METADATA_DESCRIPTION: synonym for operating system.
- COMPANY_NAME_IN_ASN: feature that allows us to differentiate the host as on premises or in the cloud.
- AUTONOMOUS_SYSTEM: if the host is in the cloud, checks for more common cloud providers, e.g., AWS, Google, GoDaddy etc.

The total feature distribution can be seen in Table 3.8, where the count designates the number of features to represent that particular protocol.

Table 3.8: Feature space distribution

Feature	Count	Feature	Count
P995_POP3S	77	P993_IMAPS	74
P8888_HTTP	43	P80_HTTP	96
P8080_HTTP	74	P7547_CWMP	22
P631_IPP	20	P587_SMTP	53
P5432_POSTGRES	32	P53_DNS	5
P502_MODBUS	5	P47808_BACNET	64
P445_SMB	1	P443_HTTPS	112
P3306_MYSQL	69	P25_SMTP	99
P23_TELNET	3	P2323_TELNET	2
P22_SSH	204	P21_FTP	26
P1911_FOX	54	P1900_UPNP	2
P1521_ORACLE	5	P143_IMAP	87
P1433_MSSQL	33	P110_POP3	77
P102_S7	2	ORG_SIZE	1
NUM_PORTS	1	METADATA_DESCRIPTION	17
COMPANY_NAME_IN_ASN	1	AUTONOMOUS_SYSTEM	15

The features extracted here are not, by any means, exhaustive. Censys has many more fields than the ones mentioned here and more features could be extracted from the current Censys data model. They are rather meant to be a best effort selection of mathematical elements from Censys. This is because the Censys data model is a moving target, as it continues to release newer data models. Hence, a highly promising direction for future work is to extract more features from up-to-date Censys fields to improve network posture-based risk models.

At this stage, we have collected 714,244 hosts across the 785 organizations. We have represented each host as a feature vector with a size of 1,386. This can be seen in Figure 3.8.

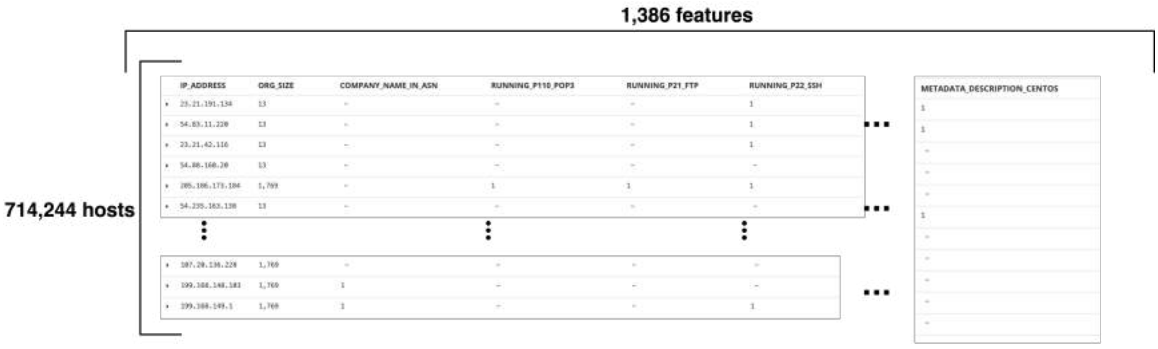


Figure 3.8: Entire data space: 714,244 hosts x 1,386 features

3.3 Design decisions

The data collection pipeline begins with a list of organizations to locate a subset of digital assets. It uses these assets to then look up the organizations' historical network configuration in Censys. This effectively returns the configurations associated with each host at a given lookup time which we then use to build our feature space.

Due to the relative novelty of this problem, there are no contemporary data collection standards. This is only exaggerated when dealing with a large data set like the one here. To effectively scope the research problem based on the project's resource constraints, the design decisions that are made along this process included some assumptions as well as certain challenges. The design decisions made in this process are categorized based on the phase of the project the challenges are encountered. These categories are:

1. Cohort selection
2. Host attribution
3. Host collection
4. Feature Engineering
5. General

Cohort selection

The design decisions made in the cohort selection stage are as follows.

1. An assumption this project makes is that organizations that have not reported a security incident are exemplary of a "good" security posture. As Liu et al. have shown, there are numerous cases where this does not hold [2]. Some of these cases are:

    - Organizations might have experienced an incident but it went unreported due to intentional or unintentional negligence An example of this could be a data breach where the data involved is not sufficient (number of records, exposed credit card info etc.) to report an incident. Another example could be if the organizations lack incident detection capabilities, or they do not have sensitive data. Hence, there might not be a requirement to report a security incident.
    - Organizations might have reported an incident close to the report window, but not within it. One of the requirements placed on victim organization is to have reported an incident in the incident report window.
    - Some Cyber Security 500 organizations had hosts that are honeypots. Honeypots are hosts that are intentionally left vulnerable to lure attackers away from other real hosts. This can lead to incorrect learning as these vulnerable host will be associated with the non-victim subset. Moreover, two organizations, namely "LookingGlass Cyber"" and "Defense Point Security", have reported security incidents within the incident report window. Although these organizations are not in the Cyber Security 500, it raises the question whether cyber security organizations are reliable non-victims.

2. This analysis assumes that the three victim data sources are exhaustive representation of all victims that have reported security incidents. However, these lists are certainly not exhaustive. Again, the combination is an attempt to be unbiased, but there is no certainty that this project has achieved unbiasedness or exhaustiveness.

3. There are duplicates across the three victim data sources, and since these are manually reported, the same name might be written different ways, e.g., "Drexel" vs. "Drexel U." vs. "Drexel University". Hence, there needs to be a manual step to filter out these duplicates.

4. As mentioned above, there are certain issues with selecting organizations that had an international presence. To mitigate this, we only selected organizations with a major office in the U.S., regardless of where most of their business is conducted. We leave dealing with incidents outside of the United States as a direction of future work.

5. Another assumption is that "hacking" type incidents are the only ones of interest for a network posture based analysis. Hence the filter criteria are biased away from more physical-based security, or non-hacking incidents. Network posture based analysis for non-hacking incident prediction is an excellent direction for future work.

6. Another decision is to randomly assign lookup dates for the non-victim organizations off of a uniform distribution. However, the lookup dates for the victim organizations are not uniform through out the incident report window.

7. During the non-victim cohort selection, this analysis avoided cloud providers (AWS, GoDaddy etc.) The reason for this is that the "ownership" for the configurations that appear in Censys would be uncertain. Again, analyzing the risk profiles of different cloud providers is left as a direction for future work.

8. As mentioned in the above section, this project did not record an organization more than once, even if it had reported more than one incident. One argument against this can be that an organization that reports an incident more than once is a better example for a organization with "bad" network posture than an organization that reports an incident only once. Although, this is a good argument, unless the non-victim cohort is afforded the same double-sample property, the cohort will be biased.

9. One thing to note is that the victim data sample are slightly biased towards health care organizations. This might be due to the HHS data source or just the simple fact that healthcare organizations have more security incidents.

10. One challenge when collecting victim organizations is whether to include incidents that involved two organizations. The analysis handled this on a case by case basis, however, there are times

when the distinction is not clear. An example for this case is the malware incident at "Aptos", an e-commerce site that reported a data breach around December of 2016. All the customers of this solution provider reported incidents as well. The analysis normally dropped these sort of incidents but since this is a web-based platform, we took the organizations that are involved as they are associated with poor maintenance (they would have regular control checks against their internet-accessible integrations).

As the above list shows, it is quite hard to collect sample organizations that have "good" external network posture. Essentially, this is because "good" network posture is hard to define in an unbiased manner. Luckily, there is one source of ground truth, which is that victim organizations have reported an incident. Therefore, one needs a significant amount of scrutiny to ensure that the victim organizations are exemplary of "bad" network postures. This is why there is a lot of effort put in during the victim organization collection step.

Host attribution

The design decisions made during the host attribution stage are as follows:

1. The dynamic nature of the internet poses a significant challenge for asset attribution. Some of the the reasons for the challenges are

   - DHCP: this protocol changes the IP address for some of the hosts belonging to an organization.
   - Cloud providers: these organizations make it difficult to identify resources belonging to a certain organization.

2. For organizations that have rather ambiguous names (mistyped or incomplete), there is uncertainty in locating the correct domain name. This step has been automated by scripting Google and Bing custom searches. However, when a conflict arises, manual intervention is required to resolve it. A good example is an organization with name 'GNAC'. The automated step returns 'GNAC sports', however, the incident report revealed the name is actually 'Gallager NAC'.

3. Our project only looked at the top 256 ARIN network handles when conducting the organizational asset lookup. Hence, the number of results is limited for only those results due to the fiscal constraints. Moreover, since there is no automated way to ensure the network handles from ARIN are registered to the organization in question, there needs to be a manual step to verify ownership

4. The domain name attribution step is not accounting for organizations that have more than one domain. Incorporating more than one domain name per organization is left as a direction for future work. Moreover, during the subdomain enumeration step, this project looks for subdomains that only match the domain associated with an organization. For example: drexel.edu will result in test.drexel.edu, dev.drexel.net, drexel.ace.com, which are all valid. However dragons.com will be dropped even if dragons.com is owned by same organization that owns drexel.edu.

5. During the subdomain enumeration step, there is no guaranteed check to see if the subdomains found are updated / registered before or after the lookup date. Given the scope of time frame and resources, this project did not conduct a historical attribution of these IP addresses. A direction for future work would be to look through Rapid7 Open Data [69] to check for historical data points. This project 'Sonar' is a historical collection of Reverse and Forward DNS lookups. This would be an ideal data set to mitigate the issue.

6. During the subdomain resolution step, the list of resolvers used as a parameter for massdns had some malicious servers. These are name servers that are intentionally returning incorrect IPV4 addresses for the supplied subdomains. To mitigate this issue, this project collected 10 reliable name server lists and ran each subdomain against 5 randomly selected groups. If 4 or more of the 5 groups return the same IP address for a subdomain, this project takes that value, otherwise it drops that subdomain; similar to a majority vote problem. Again, this back of the envelope calculation combined with some quick testing yielded good results. However, this is left as a direction for future work as well.

7. Although our scans are not designed to be intrusive, there is a possibility subdomain enumeration will overload an organization's domain server. This runs the risk of our analysis host being blacklisted. To reduce the chances of this happening, the subdomain list is shuffled at the beginning of this step.

8. This project relies heavily on the organization and domain names. However these can be inaccurate and susceptible to error on both ends (incident reporter and analyzer). A universal organization identifier (analogous to Legal Entity Identifier for financial organizations) would mitigate this issue. In an ideal world all of the data sets have this ID and it can be used to lookup their network resources. However, this is highly unlikely. A viable alternative could be if organizations like Censys and Binary Edge have an organization attribution with a certain confidence level.

Ideally, this footprinting technique should efficiently convert an organization name into a domain, or set of domains, that has a high level of confidence of correctly attributing all the public digital resources. However, even with the methods described here, there is still a possibility of false attributions and even higher likelihood of missed detection when it comes to accurately identifying the assets for an organization. The question "How do you find all the IP addresses associated with an organization?" is a very open-ended research problem. This is because locating all the assets for an organization is a rather difficult endeavor. There are a variety of reasons why attribution is difficult, one example reason is failure to locate all resources during acquisitions [70]. Another, more important, reason is that there is no ground truth to compare to. The ground truth here is an exhaustive list of the IP addresses associated to an organization at a given time. To the extent of our knowledge, no one maintains this sort of information.

Host collection

The design decisions made in the host collection stage are:

1. An assumption made here is that the hosts that have been attributed to the organizations exists in Censys. However, this might not be the case. The criteria changes depending on

what "associated" means, what configurations are of interest, and a myriad of other criteria. This issue is prevalent when dealing with cloud providers, and if the company has blacklisted Censys's probes. This issue can be mitigated early on by dropping organizations with domain names not in Censys. However, eight of the organizations (out of 785) still did not have host information in Censys during the lookup date.

2. Another issue is for organizations that are too large to be successfully processed through the entire data collection pipeline. There was only one organizations that is currently too large, however, the issue should be fixed before releasing to production.

3. To ensure updates to the Censys data model are not an issue, this analysis aggregated the table schemas inside the incident window. However, this might not be the best way to handle the issue of changing data model. The randomly assigned lookup date ensure this is not a problem during classification, since there will be an organization pair that have similar data models. This is a good direction for future work.

4. There is a cost associated with running a query against BigQuery's API. Due to the fiscal constraints imposed on this project, running daily queries is not a feasible option. To make this step more economically attainable, an aggregated weekly lookup is performed instead of daily lookups. Weekly lookups are selected as the span because it matched the budgetary constraints of the project. This means if two lookup dates are Tuesday and Friday, the component would look at the hosts form these organizations together on Monday. The component would then separate the hosts for each organization after the hosts are collected. The assumption here is that weekly lookups are close enough to daily looks not to skew the results, however, there are no scientific tests to ensure this is the case.

Feature engineering

During the feature engineering stage, the following design decisions are made:

1. There lies a subtle issue in the feature engineering scheme this project utilizes. The lack of features that describe certain protocols as deeply as others leaves a certain imbalance around

which protocols dominate the feature space. The protocols that have TLS certificates (web and mail servers), and SSH comprised around 22% of the total feature count. This does not a necessarily mean the results are inaccurate. It means there is still a lot of space for future work to define a fair scheme to extract reliable features from outside-in network posture data.

2. This analysis focused on host-based features. However, inter-host-based (organization level) features, like the ratio of hosts that are on the cloud, could prove useful and are a good direction for future work.

General

Other than the ones mentioned above, some general design decisions are:

1. A huge challenge, also mentioned in Liu et al.'s work [2] is the challenge in acquisition of high quality incident data. This results in a feature space we can not more deeply analyze with confidence. As resulted in [2], this would not be an issue if there was a more systematic and uniform incident reporting model in place.

2. An assumption we are making is that the code written to automate the procedure is free from bugs. This is definitely not the case. Although best practices and software design controls are implemented, software bugs are inevitable. One direction for future work is to open source the project to receive community support and a possible open source product

3. The numerous manual verification steps are susceptible to error. For the same reasoning given above, open sourcing the code might prove helpful on this front as well.

4. Given the resource constraints, this project did not have the time or resources to do a full reconnaissance on the organizations. However, with more tools and techniques we could provide visibility into much more than just their external network posture. This includes using tools other than Censys and the tools used in this pipeline.

5. The victim organization selection step would have been much smoother if there were a standard format for security incident reports. The sheer lack of incident context makes it impossible

to come up with a well-defined and reproducible selection criteria. Ideally, this report would have about 30-40 fields, including presence of external source, name of external source, all 4 dates involved in the security incident, etc. However, with the current reporting models, the only viable way is to read each breach report and manually construct these fields. This is a time consuming and possibly biased selection procedure. Another direction for future work is a tool that uses natural language processing to extract dates and other required features from breach reports.

Finally, although this project encountered a lot of obstacles, having so many sources of noise makes the model resilient to adversaries for same reasons as [2].

Chapter 4: Results and analysis

Once we have selected the cohort and attributed the internet hosts, we apply numerous statistical tools, models, and algorithms in order to categorize/label the data.. We then analyze the performance of the models and the rules that are most effective in discriminating between the two types of organizations and their host machines.

4.1 Algorithms

This analysis utilizes numerous algorithms to identify interesting hosts, classify these hosts as belonging to victim or non-victim organization, and locate the features that are important for discrimination. Of these many algorithms, some of the more substantial ones are described here.

4.1.1 Spearman correlation

A rank correlation is defined as a statistic that measures an ordinal association (relationship of ranking between different variables). The Spearman correlation coefficient [21, 20] is a rank correlation that measures the strength and direction of association between two variables. It is also a nonparametric measure that does not require data from a normal distribution. It describes the relationship between two variables using a monotonic function, as opposed to a plain linear relationship. The result is a value between -1 and 1, denoting the maximum negative and maximum positive relationship between the variables respectively. It is sometimes accompanied with a hypothesis test to ensure the values are prevalent in the data set and are not edge cases. According to the general guidelines (also used in [1]), an absolute value from

- 0.1 to 0.3 is slightly correlated
- 0.3 to 0.5 is moderately correlated
- 0.5 to 1.0 is strongly correlated

4.1.2 Cross-validation

Cross-validation (out-of-sample testing) is a statistical technique that is used to gauge the real-world performance of a model. It does this through a generalization test against an independent data set. In a prediction problem, a model is usually given a data set of known data which is used for training, and a data set of unknown data which is used for testing. Hence, the goal of cross-validation is to test the model's ability to predict the label for novel data (not used in training), in order to identify problems like overfitting or selection bias. It is popular because of its ease in understanding, and more realistic performance compared with the traditional train/test split. A given round of cross-validation involves partitioning the data set into equivalent training and testing sets, training the model on the training data, and predicting the labels for the testing data. Depending on the type of cross-validation technique, multiple rounds of cross-validation may be performed under different randomized partitions of the data set, and the performance metrics are then combined for all of the rounds to estimate the model's true performance. This combined estimate for the model usually is the mean along with a variability measure (variance, standard deviation). In this analysis, we used a specific kind of cross-validation called k-fold cross-validation, which is known to work well for limited data samples. The parameter, k, has two definitions namely

- the number of rounds we repeat the iterative process mentioned above
- the number of groups we separate the data set

First, we split the entire data set into k groups. For a given iteration, k-1 groups are combined to be used as the training data, while the last one is used for testing. Then, one of the training groups is swapped for the testing group and another round commences. This procedure is repeated for a total of k times and the performance metrics are extracted and combined. A common value for k (also used in our analysis) is 5, meaning the data set is separated into 5 groups and there are 5 training-testing rounds. An important note is that tuning hyper-parameters (e.g., number of features, feature selection) be done on the training set as not doing so might result in "data leakage" and unrealistic performance of the model.

4.1.3 Performance metrics

In a binary (two-class) classification task, there are four possible meta-labels for data points, depending on the source and predicted label. These are:

- True Positive (TP): true positive samples predicted as positives
- False Positive (FP): true negative samples predicted as positives
- True Negative (TN): true negative samples predicted as negatives
- False Negative (FN): true positive samples predicted as negatives

The metrics utilized in our project are:

- **True Positive Rate (Sensitivity, Recall)**: the fraction of true positive samples that are labelled as positives in the testing phase

$$TPR = \frac{TP}{TP + FN}$$

- **False Positive Rate (1-Specificity)**: the fraction of true negative samples that are labelled as positives in the testing phase

$$FPR = \frac{FP}{FP + TN}$$

- **Accuracy**: the fraction of overall samples that are labelled correctly

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN}$$

- **F1-score**: the weighted average (harmonic mean) of the precision and recall.

$$f1 = 2 * (\frac{precision * recall}{precision + recall})$$

Here, precision (positive predictive value) is a measure of how many positively predicted labels

are positive and recall is another name for TPR mentioned above.

$$precision = \frac{TP}{TP + FP}$$

This measure is highly useful in imbalanced classification tasks, as it takes into account the prevalence of the assigned labels. The F1-score for a equally balanced classification problem reduced to the measure of precision.

- **Support**: This is not a performance metric, but rather shows how many samples are presents in each class to *support* the hypothesis.

4.1.4 Receiver Operating Characteristic (ROC) Curve

To measure the performance of a discriminative model, we first have to predict the labels using the assigned probability score. This label is assigned based on the deviation of a probability score from a discrimination threshold. The performance is then calculated as some metric that accounts for the correctly (or incorrectly) labelled points. Since the threshold that maximizes the desired performance metric is often not known, a sweep for all values of the threshold is conducted. ROC is a graphical plot that shows the performance of a probabilistic binary classifier for all values of this discrimination threshold. The two axes of a ROC plot are:

- True Positive Rate (TPR) in the y direction
- False Positive Rate (FPR) in the x direction

Another interpretation of the ROC is the graphical plot of the TPR as a function of the FPR.

Both the TPR and FPR axes start out at (0, 0). As the threshold is varied from the highest value to the lowest, more points are afforded the positive label and both the FPR and TPR increase. The sweep is performed until the minimum value for the threshold is selected, where we have 100% TPR and FPR. A straight line that connects the two extreme points ((0,0) and (1,1)) is referred to as the chance line. The chance line signifies the performance of a model that is guessing the labels off a random uniform distribution.

The area under the ROC curve (AUROC) is a performance metric that is commonly used when showing a model's skill independently of this threshold. A good model would have an area of 1 and bad model will have an area close to 0.5 (equivalent to uniformly guessing) or less. AUROC enables comparison between two models trained on a discrete number of data points. It is often used when two models are compared to each other independently of a discrimination threshold.

The ideal operating point, which can be calculated few different ways depending on the needs of the problem, is between the two extremes (0,0) and (1,1). In this project, Youden's J-statistic is used to select this optimal operating point. This point can be calculated as

$$J = \arg\max_{thr} TPR(thr) - FPR(thr)$$

This value gives the threshold that maximizes the difference between TPR and FPR. Intuitively, this is the point on the ROC curve that has the largest vertical distance from the chance line.

4.1.5 Recursive Feature Elimination (RFE)

When dealing with a large feature space, like the one in this analysis, there are some challenges to consider namely:

- a large amount of data is needed to accurately represent possible values for each feature
- there is a higher chance of overfitting, as the model will start learning the noise associated with the training data
- it takes more time to train the model with a large feature space
- the distance / similarity is more difficult to calculate than for a smaller feature space

Hence, it is wise to reduce the space to a set of strong features. Feature Selection is the process of selecting the features in the data set that matter the most to the target label (strong features). In this analysis a feature's importance is quantified as the performance (model's area under the ROC curve, more in 4.1.4) dip in a supervised classification model trained without that feature. We choose this feature importance method as it directly gauges the performance increase associated

with each feature. Hence, weak features are ones that reduce or fail to improve the performance when included. Recursive Feature Elimination (RFE) is a preprocessing feature selection method that fits a model to data and removes these weak features. It operates by recursively dropping a small number of features (≤ 10) per loop until the desired number of strong features is reached. Since the desired number of features to keep is often not known, cross-validation is also used with RFE to score different feature subsets and select the best performing subset. This cross-validation (RFECV) step hence automates the number of features selected. An important note here is that RFE and RFECV are both only used with the training data, and not the testing data. We can use these principles to also generate a chart along with the feature elimination procedure. This chart has the number of features on the X-axis and performance of the model on the Y-axis to show the performance dip as more features are added.

4.1.6 Random Forests

Random Forest [43] is a well-known ensemble classification algorithm. It is generally considered to work well with large and diverse feature sets, and has been used in this domain to achieve good results [2]. The relative ease of use combined with the ability to show feature importance makes Random Forest among the most useful machine learning algorithms. The internal architecure is comprised of individual decision trees as base estimators (see Figure 4.1). A decision tree is a non-parametric supervised learning method that predicts the value of a target label by learning simple decision rules from the features. It does this by using a tree-like structure to simulate decisions and outcomes.

This algorithm follows these steps to between against the different data classes:

1. For every base estimator (decision tree), select a random sample set of features with approximate size of the square root of the feature space

2. Select a random feature from this feature set

3. Select a value for that feature that minimizes the data set's impurity (find a split value that best separates the data set). In this analysis, the method to achieve this was Gini impurity.

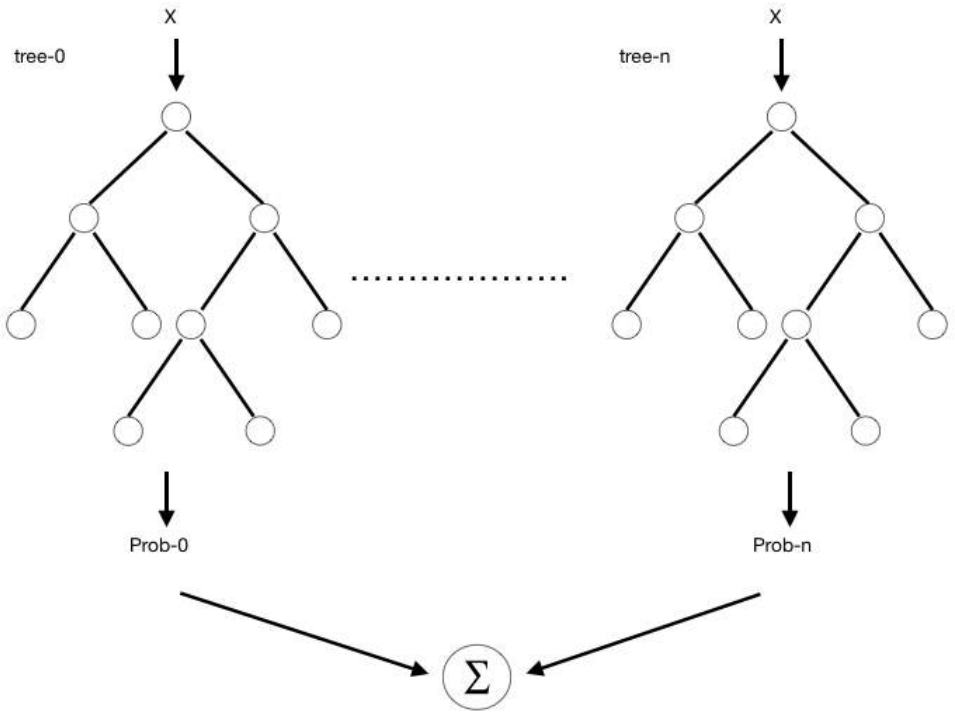


Figure 4.1: Random Forest architecure

4. Split the data on that value for that feature

5. Repeat steps 2 - 4 until either the sample data points are fully separated or the stopping criterion (maximum depth) is hit

6. The terminating (leaf) nodes at the bottom contain counts for each class, of which the label assigned is the class with highest support

The number of base estimators to use depends on the problem, but as rule of thumb, is set to 100-150. The class probability output converges around this range for most data sets. Out of an abundance of caution, this project uses a value of 200 for all the RF classification tasks. During testing, a data point follows the path down a decision tree based on its features. The label associated with the data point in an individual tree is the class with the highest support at the terminating node. The output for a data point is a probability score that is the ratio of predictions across the trees in the forest. A threshold would need to be applied to complete this as a classification model

(by varying this threshold, we get an ROC curve as seen in Section 4.1.4).

Many modifications exist for the Random Forest algorithm. Of these, the most useful addition to this project is Bagging (Bootstrap aggregating). Bagging is the process of randomly selecting data points with replacement to learn the individual decision trees. This sampling with replacement creates a data set per individual tree that has not been used in training specific trees. This data set, called Out-Of-Bag samples, can be used as an internal testing set on the individual trees that have not seen those observations, to gauge the real-world performance of the model. The resulting performance score from the Out-Of-Bag testing is referred to as the Out-Of-Bag (OOB) score.

Since the optimum maximum depth for the trees is unknown, it is tuned for every problem to avoid under-fitting. This is done by selecting a smaller sample (10 - 25%) of the data set, then sweeping the depth until the individual trees are complex enough to learn the problem. The ideal value is where the performance of the model converges, and there is no performance that can be gained through using trees of greater complexity.

4.1.7 Outlier Detection and Isolation Forest

Outliers are observations with rare occurrences in a data sample, and hence, are statistically different from most other observations. Anomalies are patterns of different data within given data, whereas outliers are extreme observations within data. The main difference between anomalies and outliers is that during training we assume there are no anomalies, but, may contain outliers. Outlier Detection (Unsupervised Anomaly Detection) is a method of identifying observations that are outliers. It is widely used in fields such as credit card fraud detection, insurance, intrusion detection, critical system fault detection, and military surveillance. There are many ways an outlier could exist in a data set (variability in measurement, error, etc.), and are often highly relevant to the analyst. The two outlier detection mechanisms are:

- Univariate Outlier Detection: find observations with extreme value for one feature (variable), e.g., Box Plot Rule
- Multivariate Outlier Detection: find observations with extreme value for multiple features,

e.g., Mahalanobis Distance, OneClassSVM, EllipticEnvelope, Local Outlier Factor

Most existing model-based approaches to outlier detection construct a profile of normal instances [5]. They then identify instances that do not conform to the normal profile as anomalies. Isolation Forest [5] is a ensemble decision tree algorithm that is used to explicitly identify anomalies instead of profiling normal points. The algorithm works as such:

- For every base estimator (decision tree), select a sample of 256 data points from the data set
- Select a random sample set of features with size approximately equal to the square root of the feature space (generally, both anomalies and outliers can be explained by a few features)
- Select a random feature from this set
- Select a random value between the minimum and maximum for that feature
- Split the data on that value for that feature
- Repeat the above steps until the sample data points are fully separated

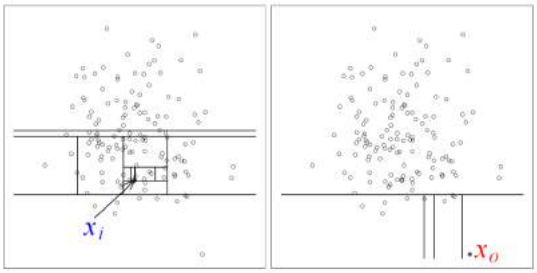
After these steps, each data point will have a set of associated depths (where in the decision tree this data point separates from the others). The intuition is that anomalies / outliers will produce noticeably shorter paths on random partitions of the feature space, as we can see in Figure 4.2.

Since outliers will be isolated more easily on random partitions of the feature space, the probabilistic expectation (mean) for this depth will be less for outliers than the other observations. The scoring function [5] used in the Isolation forest algorithm can be seen in Eqn. 4.1

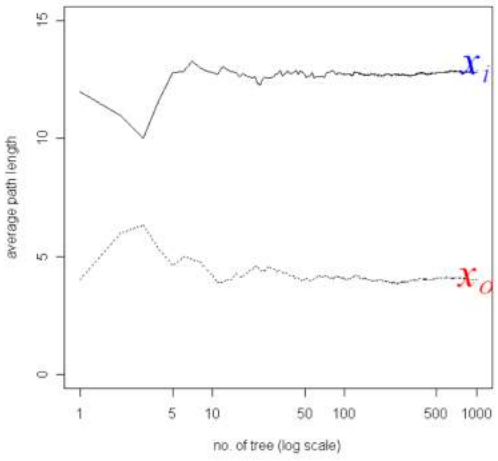
$$s(x, n) = 2^{-\frac{E(h(x))}{c(n)}} \tag{4.1}$$

where:

- $s(x, n)$ is the outlier score for a data point, x, in the sample data set, n
- $E()$ is the expectation function (mean) of a value across all trees
- $h(x)$ is a function that returns the depth for a data point, x



(a) Isolating x_i (left) and x_o (right)



(b) Average depths converge

Figure 4.2: Isolation Forest [5]

- $c(n)$ is the average path length of unsuccessful search in a binary search tree (average height of the base trees)

From this equation, if $E(h(x))$ (average height of a data point) [5]

- is $n-1$, $s(x,n)$ goes towards 0 (data point is always hard to isolate, hence is an inlier)
- is 0, $s(x,n)$ goes towards 1 (data point is always easy to isolate, hence is an outlier)
- is $c(n)$, $s(x,n)$ equals 0.5 (data point is ambiguous, this is the decision threshold)

Since the Isolation Forest algorithm samples the data set before identifying outliers, the maximum depth of the tree is the number of samples used to build the tree [5]. Isolation Forest is superior to other anomaly detection techniques because:

1. it does not need to learn the underlying distribution (not a generative model)

2. It does not require any parameters other than the data and number of decision trees, which reduces developer's bias

Although not used in this project, another note-worthy outlier detection algorithm is Local Outlier Factor (LOF). It measures the deviation / distance of a given observation from its neighbors. The LOF / anomaly score is based on how isolated an observation is from its surrounding neighborhood (k-nearest). This locality measure is used to calculate the local density. This local density value is compared with the k-neighbors, and ones with lower density are deemed outliers. This algorithm works well if you have values for k and the threshold. However, we did not have these values, hence Isolation Forest was used for outlier detection.

4.2 Experimental setup

In this section, we setup the experiment as a machine learning problem. The target label is whether an organization **has** or **has not** reported a security incident. The features are the individual hosts' configurations in an organization's network. The challenge here is how to learn a model when the labels are at the organization level and the features are at the host level (See Figure 4.3). In short, how to map the features to the label, and set this up as a machine learning problem, when the features and label are at different resolution.

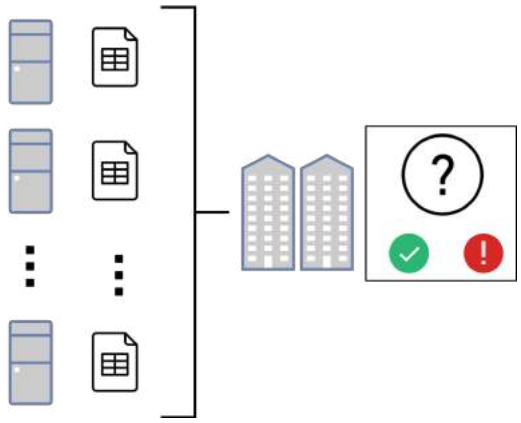


Figure 4.3: Challenge with experimental setup: features and target label at different levels

One naive way is to average all the features across an organization's hosts. However, any idiosyncratic feature associated to a host is lost in a large enough organization; which is the case for

numerous organizations in our cohort. Moreover, the features we are most interested in are misconfigurations and they are likely to be the minority in a network. Another approach is to assign the label to each of an organization's hosts. However, this again will be sub-optimal for two reasons:

1. This subtly wipes away the idiosyncrasies by overloading the model with numerous hosts that are not peculiar

2. If an organization is large enough, then looking at all of the hosts is time consuming and impractical

This project uses unsupervised anomaly detection to mitigate the above issues. Using this approach, we can identify hosts that are a representative sample of machines in a network (inliers) and hosts that are different (outliers) from the majority of the organization's hosts. An Isolation Forest model with 200 base estimators is learned with the host feature vectors and used to identify liers (outliers and inliers) for each organization. The number of base estimators is the only parameter passed to this algorithm, for the reasons mentioned in Section 4.1.6. It then assigns scores (probability of being an outlier) as well as lier labels (using the probability score and the internal threshold) for each host.

Some caveats / edge cases did exist in the lier identification step. These edge cases are:

- For organizations with one host, the host is assigned an outlier label

- For organizations with two hosts, the host with higher number of ports is the outlier and the other is an inlier

- For organizations with more than two hosts, if the algorithm could not locate an outlier, the analysis takes one host with the highest outlier score as the outlier and one with the lowest outlier score as the inlier

However, these cases are very rare, and the bulk of the organizations have well-defined outliers and inliers. After the lier identification stage, we identified 45,329 outliers and 45,225 inliers from the 714,244 total host counts as seen in Table 4.1.

Table 4.1: Outlier and inlier counts for cohort subsets

Cohort Subset	Inliers	Outliers
VICTIM(BREACH)	3650	3666
CERT	27743	27758
DNS	11978	12017
SEC500	4321	4355

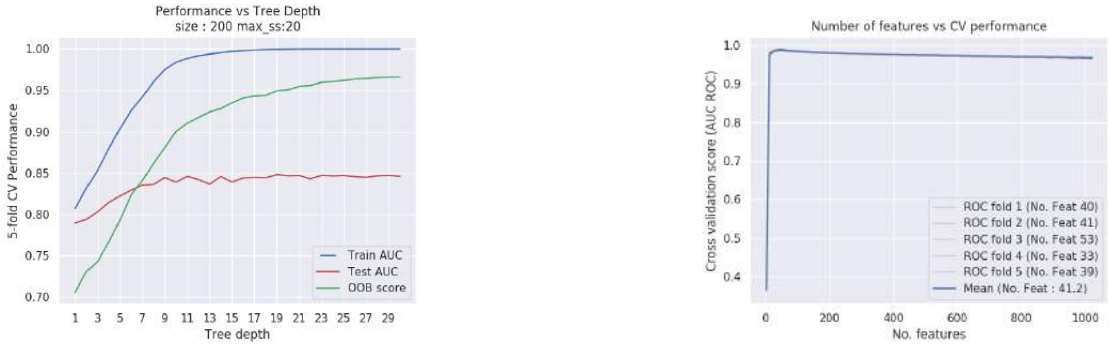
This is about 6% outliers and 6% inliers, which reduced our data space to about 12% of the original size.

4.2.1 Outlier vs. inlier analysis

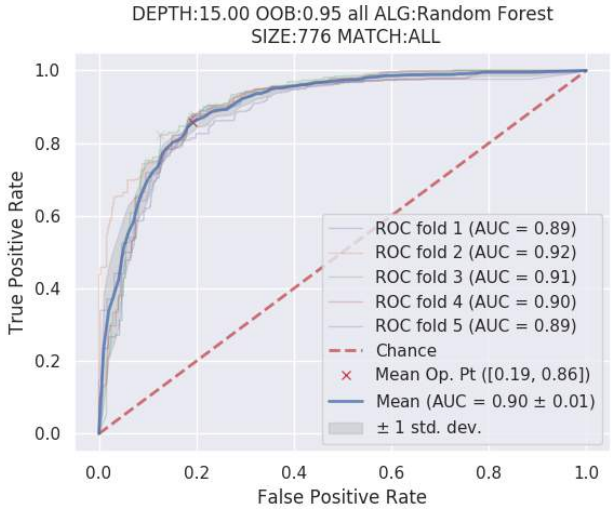
After identifying sample liers, the project attempts to answer the question "does the criteria that make a host an outlier transcend organizational boundaries?" That is, if a host is an outlier in a certain organization, will it be an outlier in another organization? To perform this analysis, we select a set of 20 sample outliers and 20 sample inliers for each organization from the cohort as our data set, resulting in 7,208 inliers and 7,312 outliers. This is set up as a binary classification problem where the class label is 1 for outliers and 0 for inliers. A random forest algorithm is used, along with RFECV and depth tuning, to discern between these two liers, as seen in Figure 4.4 (a). An optimal maximum tree depth of 15 is tuned on a sample of 200 randomly selected organizations (20 inliers and 20 outliers from each organization), as seen in Figure 4.4 (a). With this parameter, an ROC curve is generated for the tpr and fpr values as can be seen in Figure 4.4 (b).

During the analysis, we noticed that the performance for discerning outliers from inliers depend on the organization size (number of hosts in an organization). Hence, this analysis is repeated for organizations grouped by their organization size. The ROC curves for the different organization sizes can be seen in Figure 4.5. We calculate the optimal threshold using Youden's J score as described in Section 4.1.4. Using the optimal threshold ('Mean Op. Pt' in ROC graph), we are able the gauge the performance of the model. The performance metrics (f1-score, accuracy, fpr, number of important features, and support for outliers and inliers) for this analysis can be seen in Table 4.2.

Since the host attribution step during the data collection is susceptible to historical noise, the above analysis is repeated for hosts that are attributed through domain name in host certificates.



(a) Tree depth tuning and recursive feature selection



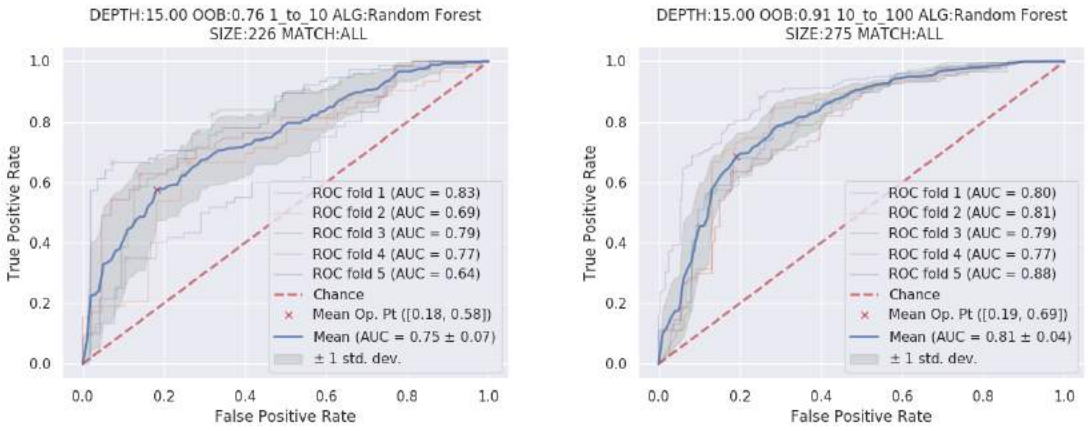
(b) ROC curve

Figure 4.4: Outlier vs. inlier classification using all attributions

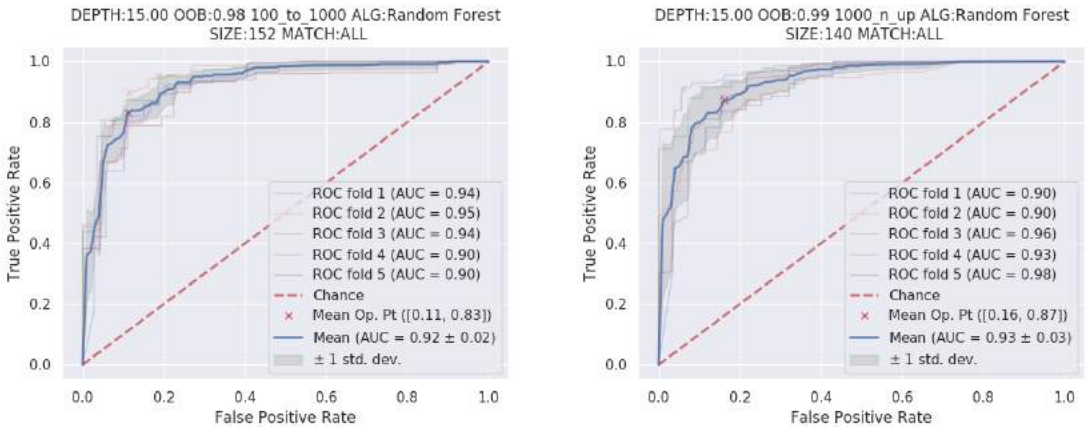
Attribution through domain name in certificates is more historically reliable but is less likely to capture hosts that do not have a certificate (attackers will more likely leverage because of less visibility) [17, 26, 27]. The ROC curve for lier discrimination using only certificate attribution can be seen in Figures A.3 and A.4. The performance of the model using the optimal discrimination threshold is tabulated in Table 4.3.

4.2.2 Victim vs. non-victim host analysis

At this stage, the organization label is assigned to these representative lier samples (inliers and outliers). This enables us to setup three separate classification scenarios: one victim cohort subset against three non-victim cohort subsets. This is a result of the three different methods of identifying



(a) Organization size: [0 to 10] (left) and [10 to 100] (right)



(b) Organization size: [100 to 1000] (left) and [1000 and above] (right)

Figure 4.5: Outlier vs. inlier classification for different organization sizes using all attributions

non-victim organizations. This is repeated for the two types of liers (outliers and inliers) separately as seen in Figure 4.6.

Hence we have learned six models, two liers for three different scenarios. Due to resource constraints on this project, we select up to a maximum of 350 samples per lier per organization for this analysis. These models are used to generate ROC curves that can be seen in Figure 4.7. The performance of the model at the optimal discrimination threshold is tabulated in Tables 4.4 and 4.5 for the two lier classification problems.

The lier classification is again repeated for hosts that are attributed through domain name in host digital certificates. The ROC curves using only certificate attribution can be seen in Figure A.5. The performance of the model at the optimal discrimination threshold is tabulated in Tables

	f1-score	accuracy	fpr	supp0	supp1	no feats
≤ 10	0.70 ± 0.07	0.71 ± 0.06	0.18 ± 0.09	259	335	231
10 - 100	0.76 ± 0.04	0.76 ± 0.04	0.27 ± 0.05	1788	1816	71
100 - 1000	0.87 ± 0.02	0.87 ± 0.02	0.14 ± 0.05	2423	2423	61
≥ 1000	0.87 ± 0.04	0.87 ± 0.04	0.14 ± 0.05	2798	2798	40
all sizes	0.84 ± 0.01	0.84 ± 0.01	0.18 ± 0.04	7208	7312	41

Table 4.2: Outlier vs. inlier using all attributions

	f1-score	accuracy	fpr	supp0	supp1	no feats
≤ 10	0.65 ± 0.05	0.66 ± 0.05	0.22 ± 0.13	249	380	48
10 - 100	0.76 ± 0.04	0.76 ± 0.04	0.22 ± 0.08	1359	1406	39
100 - 1000	0.88 ± 0.04	0.88 ± 0.04	0.12 ± 0.04	2419	2419	38
≥ 1000	0.91 ± 0.05	0.91 ± 0.05	0.13 ± 0.07	1940	1940	37
all sizes	0.83 ± 0.02	0.83 ± 0.02	0.17 ± 0.04	5925	6103	39

Table 4.3: Outlier vs. inlier using only certificate attributions

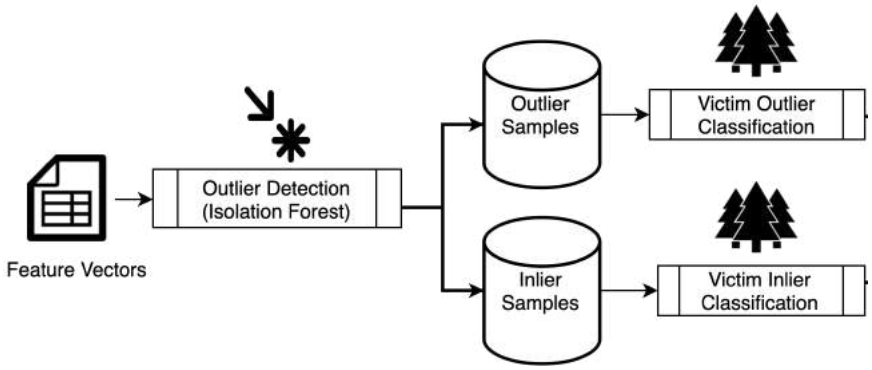


Figure 4.6: Host classification using outlier and inlier hosts

4.6 and 4.7.

4.2.3 Victim vs. non-victim organization analysis

Once the inlier and outlier hosts are attributed to victim and non-victim organizations, the challenge was to reduce these hosts to an organizational risk profile. The intuition behind this profile is that it should be a "summary" of the sample liers' probability scores. Hence, this project approximates this risk profile as the summary statistics for the distribution of scores across the lier machines. These summary statistics are:

1. 5 quartiles: [0, 25, 50, 75, 100] (0 = minimum, 100 = maximum, 50 = median)

2. average of scores

	f1-score	accuracy	fpr	supp0	supp1
SEC500	0.57 ± 0.14	0.62 ± 0.09	0.21 ± 0.23	3892	3166
CERT	0.67 ± 0.19	0.64 ± 0.20	0.39 ± 0.25	16501	3166
DNS	0.85 ± 0.05	0.85 ± 0.05	0.11 ± 0.03	9049	3166

Table 4.4: Victim vs. non-victim inlier host using all attributions

	f1-score	accuracy	fpr	supp0	supp1
SEC500	0.64 ± 0.09	0.64 ± 0.09	0.44 ± 0.19	3926	3182
CERT	0.62 ± 0.13	0.59 ± 0.15	0.41 ± 0.23	16516	3182
DNS	0.73 ± 0.09	0.72 ± 0.08	0.35 ± 0.16	9088	3182

Table 4.5: Victim vs. non-victim outlier host using all attributions

	f1-score	accuracy	fpr	supp0	supp1
SEC500	0.76 ± 0.14	0.77 ± 0.14	0.21 ± 0.22	2792	2328
CERT	0.64 ± 0.09	0.58 ± 0.09	0.50 ± 0.11	15119	2328
DNS	0.73 ± 0.08	0.73 ± 0.09	0.25 ± 0.18	4882	2328

Table 4.6: Victim vs. non-victim inlier host using only certificate attributions

	f1-score	accuracy	fpr	supp0	supp1
SEC500	0.67 ± 0.12	0.66 ± 0.13	0.32 ± 0.14	2855	2378
CERT	0.71 ± 0.13	0.68 ± 0.15	0.27 ± 0.20	15152	2378
DNS	0.74 ± 0.08	0.73 ± 0.08	0.26 ± 0.15	4919	2378

Table 4.7: Victim vs. non-victim outlier host using only certificate attributions

3. variance of scores

4. count (length) of liers

These result in 16 summary statistics (8 per lier) for each organization that can now be used to train a risk model. The summary of the training probabilities are used as the training risk profiles, and the same is applied for the testing probabilities, as seen in Figure 4.8.

These are used to train a Random Forest risk profile model that will predict the likelihood to report a security incident based on the lier samples. The ROC curve for the models ran against each cohort subset can be seen in Figure 4.9. The performance of this model at the optimal discrimination threshold is tabulated in Table 4.8.

The victim vs. non-victim organization analysis was repeated for certificate attributed hosts as well. The ROC curve for each cohort subset can be seen in Figure A.6 The performance of this model at the optimal discrimination threshold is tabulated in Table 4.9.

	f1-score	accuracy	fpr	supp0	supp1
SEC500	0.72 ± 0.04	0.72 ± 0.04	0.26 ± 0.09	200	199
CERT	0.75 ± 0.05	0.75 ± 0.05	0.27 ± 0.08	198	199
DNS	0.72 ± 0.08	0.72 ± 0.07	0.23 ± 0.14	194	199
Mean	0.73 ± 0.06	0.73 ± 0.05	0.25 ± 0.10	197	199

Table 4.8: Victim vs. non-victim organization using all attributions

	f1-score	accuracy	fpr	supp0	supp1
SEC500	0.71 ± 0.05	0.71 ± 0.05	0.27 ± 0.15	180	177
CERT	0.77 ± 0.04	0.77 ± 0.04	0.25 ± 0.05	187	177
DNS	0.71 ± 0.06	0.72 ± 0.05	0.35 ± 0.17	164	177
Mean	0.73 ± 0.05	0.73 ± 0.05	0.29 ± 0.12	177	177

Table 4.9: Victim vs. non-victim organization using only certificate attributions

4.3 Analysis of results

In Section 4.2.1 (Outlier vs. Inlier analysis), we can see the following results

1. Outliers are easier to identify for large organizations compared to smaller ones. As the organization size increase, the AUROC increases and that the number of important features that are required to predict the target label decreases. These can be seen in Figure 4.5 and Tables 4.2 and 4.3.

2. If one has no knowledge of the organization size, then for a given set of hosts this model can use 41 features to assign an outlier label with 0.84 ± 0.01 accuracy and 0.18 ± 0.04 fpr. However, in the worst case scenario (small organization size), the performance could drop to 0.70 ± 0.07 accuracy and 0.18 ± 0.09 fpr.

These statements are consistent for the hosts that are attributed through only certificates as well. Identifying outliers in smaller organizations is difficult because the rules that make a host an outlier are not well-defined. Moreover, the support (number of hosts) is much lower for smaller organizations, hence, there might not be enough observations to learn a good model.

In Section 4.2.2 (Victim vs. Non-Victim host analysis), we can see the following results

1. Inlier analysis for the DNS and CERT cohort subsets perform well (discernable against VICTIM subset). However the SEC500 inliers do not perform as well. This is due to the organi-

zation size being an important feature for the two cohorts.

2. Outlier analysis performs about the same for all of the cohort subsets, while performing the best for the DNS cohort subsets. This, again, is the effect of the ORG_SIZE feature.

3. It is important to note that there is training support imbalance in the DNS and CERT classification task. This is due to the large sizes of those non-victim organizations.

4. The certificate attributed liers perform better, as we can see in Tables 4.6 and 4.7, compared with Tables 4.4 and 4.5.

Through the individual lier classification, it is not apparent which type of lier is better for security incident prediction. In other words, looking at the sample inliers or sample outliers separately will not lead to the most optimal classification performance. However, when combined to predict the label for an organization, the model performs better. In Section 4.2.3 (Victim vs. Non-Victim Organization Analysis), we can see that summarizing the scores for individual liers into a risk vector shows an average classification performance of 0.73 ± 0.06 accuracy, 0.73 ± 0.06 f1-score, and 0.25 ± 0.10 fpr (see Figure 4.9 and Table 4.8). This is consistent for the certificate attributed hosts as well as seen in Table 4.9.

4.3.1 Feature importance

From the above ROCs and performance metrics, we can see the classification performances of the models. The Outlier vs. Inlier classification shows that, depending on the organization size, we can identify outliers with 70 - 85% accuracy. However, in order to fully answer the question what makes an outlier in this project's feature space, we need to identify the rules that are important in the classification. Sarabi et al. have stated "in the context of security, simply building black-box models is not sufficient, as one cannot readily infer why a model is making a certain prediction" [10]. To overcome this challenge, we take the features that are important in the Random Forest model and combine them with the Spearman correlation between the feature and the target label to identify these rules. This collection is then sorted based on correlation and the top 20 are presented as a feature importance chart. The feature importance chart for the Outlier vs. Inlier classification using

all attributions can be seen in Figure 4.10. On the X-axis of this chart, we can see the Organization size associated with the model, while the Y-axis represents the feature. Each entry shows how correlated that important feature was with the target outlier label. Using certificate attribution, a similar chart can be seen in Figure 4.11.

There are numerous speculations that can be made from the charts in Figures 4.10 and 4.11. For the sake of brevity, we will only analyze a small subset of the features. The following are positively correlated with the outlier label:

1. Running an SSH server (We can see that most of the top 20 features are taken up by the SSH protocol)
2. An HTTPS web server that is configured with 'P443_HTTPS_DHE' (Diffie Hellman) features
3. A web server on port 80 that returns a "200" (OK)

The following are negatively correlated with the outlier label:

1. An HTTPS service with a valid, browser-trusted certificate
2. An HTTPS web server that is not configured with Diffie Hellman features
3. An 'Akamai' web server on port 80 that returns a "400" bad request

The same feature importance chart analysis can be extended to victim vs. non-victim host analysis. The charts can be seen in Figures 4.12, 4.13, 4.14,and 4.15.

Again, many speculations can be made using the outliers in Figures 4.14, 4.15 The following are a select sample:

1. Larger organization sizes are negatively correlated with the victim label for the CERT and DNS cohorts. However, they are positively correlated with the victim label in the SEC500 analysis. This means that victim organizations are smaller in size compared with randomly sampled organizations from the Internet.

2. Running an SSH server is negatively correlated with the victim label in the SEC500 and DNS cohort subsets, and indifferent in the CERT subset.

3. Running a HTTPS server that has 'P443_HTTPS_RSA_EXPORT_SUPPORT' (FREAK vulnerability which allow an attacker to force export-grade encryption [71]) enabled is slightly positively correlated with the victim label.

4. Running an HTTPS certificate with Diffie-Hellman or Key Encipherment is slightly positively correlated with the victim label.

The following can be deduced from the inlier charts:

1. Larger organization sizes, again depending on the cohort, are positively correlated with the victim label.

2. In the feature importance chart using all attribution techniques (Figure 4.14), we can see that compared to the VICTIM cohort, the DNS cohort has far fewer HTTPS certificates. This is a direct example of how selecting a non-victim cohort can be a difficult task, and the criteria for select an exemplary cohort are very subjective. In this case, the inlier model has identified the criterion to be that victim organizations' profile contain more certificates than non-victims. This, depending on what the developer wants the model to learn, may or may not be an issue.

3. Running an "Akamai" web server with a "400" response is positively correlated with the victim label in the SEC500 and DNS subsets, but negatively correlated (or indifferent) with the CERT subset

4. A HTTPS certificate that has organization level validation ('VALIDATION_LEVEL_OV') is positively correlated with the victim label in all subsets

5. Running an HTTPS certificate issued by Comodo and GeoTrust is slightly positively correlated with the victim label.

6. The length of validity for the HTTPS certificate is negatively correlated with the victim label. This means a longer validity length is associated with non-victim organizations.

The summary statistics risk vectors analysis is also afforded the feature importance chart analysis as we can see in Figures 4.16 and 4.17.

From Figures 4.16 and 4.17 we can see that:

1. The inlier statistics are strongly positively correlated with the victim label in the CERT subset. However, in the other subsets, the correlation is equally distributed among the liers. Moreover, in DNS subset for the certificate-only attribution, the outlier statistics are more correlated with the target label.

2. Having more outliers is negatively correlated across all cohort subsets. This does not extend to inliers since the number of inliers in this analysis was chosen to match the number of outliers.

3. High variance in the probability scores of the outliers is negatively correlated with the victim label.

4.4 Discussion

The feature importance charts exemplify the fact that this analysis does not conduct vulnerability analysis, but rather profiles network postures. This is because the important features do not necessarily represent weaknesses in the network, but rather similarity to organizations that have reported security incidents. We can see that the presence of the SSH protocol is positively correlated with the outlier label, but negatively correlated with the victim label. This is consistent with this protocol's intuition of being rare but also secure. However, a misconfigured HTTPS server (Diffie-Hellman [72] and FREAK [71]) is positively correlated with the outlier label and the victim label. We found that a known vulnerability (FREAK [71]) is slightly positively correlated with the victim label in almost all cohort subsets. This is an indication that victim organizations have under-managed networks, as seen in previous works [1, 2]. Moreover, the indication we find is in line with previous works by Zhang et al. that found untrusted HTTPS is by far the most important mismanagement features [1, 2]. Finally, the performance we achieved is also consistent with existing works that use configuration information [29, 73].

Referring back to the the contributions mentioned in Figure 1.3, we can take away the following:

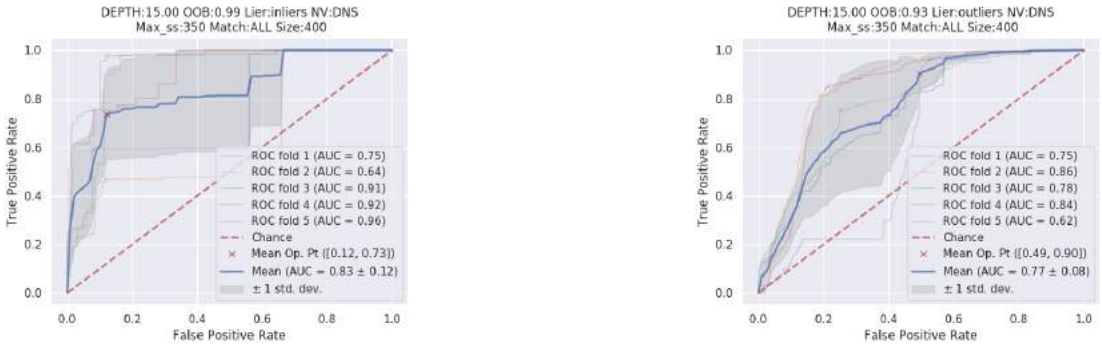
1. The heterogenous non-victim collection methods that show us the rules that discern between a victim and a non-victim depend upon how we collect our non-victim samples.

2. The footprinting techniques reveal that SSH is a secure outlier, as this was not visible without this attribution technique

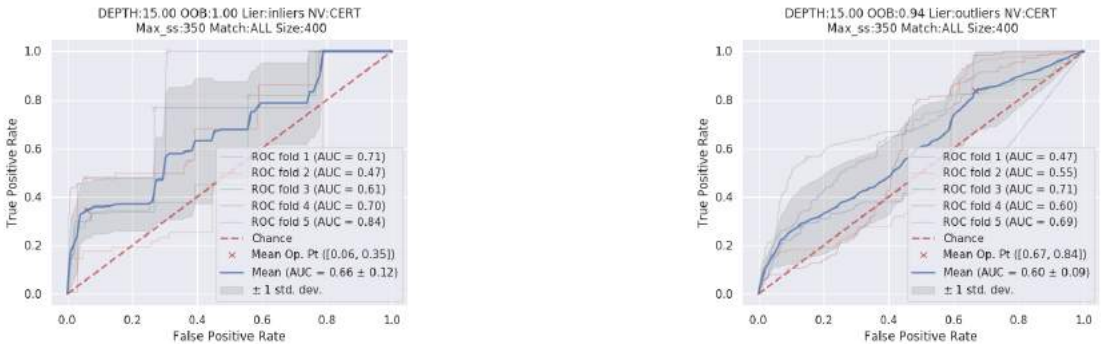
3. A more holistic representation, a result of the large feature space, of the hosts allows us to generate the feature importance charts with greater detail and dig even deeper into each protocol

4. The state of the art outlier detection techniques show us that non-victims tend to have more outliers, and higher variance in those outliers

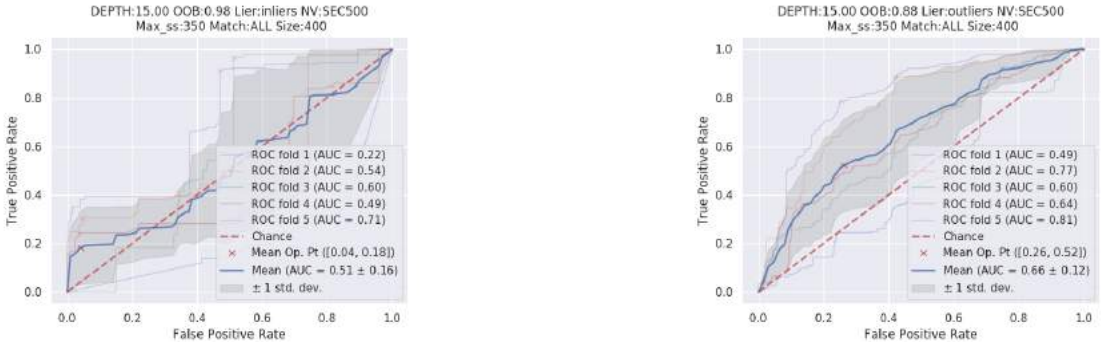
In addition to these contributions, this analysis also enables the expansion of effective risk management sectors like cyber insurance, and aids underwriters in better customization of their policies.



(a) Random sampled DNS non-victims: inliers (left) and outliers (right)



(b) Random sampled CERT non-victims: inliers (left) and outliers (right)



(c) Random sampled SEC500 non-victims: inliers (left) and outliers (right)

Figure 4.7: Non-victim vs. victim host classification using all attributions

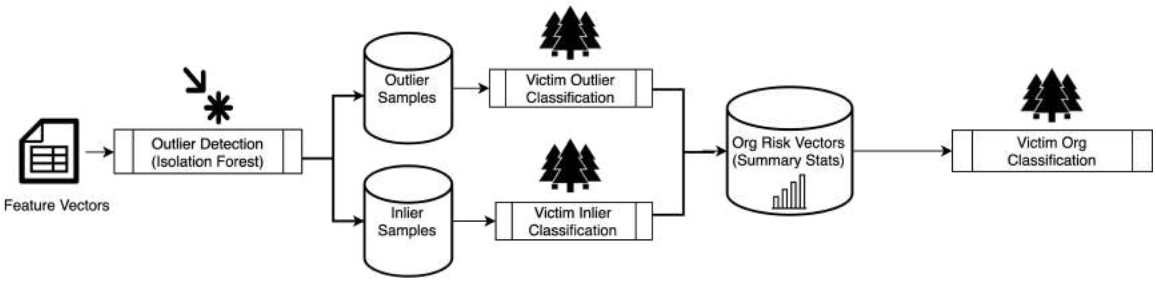
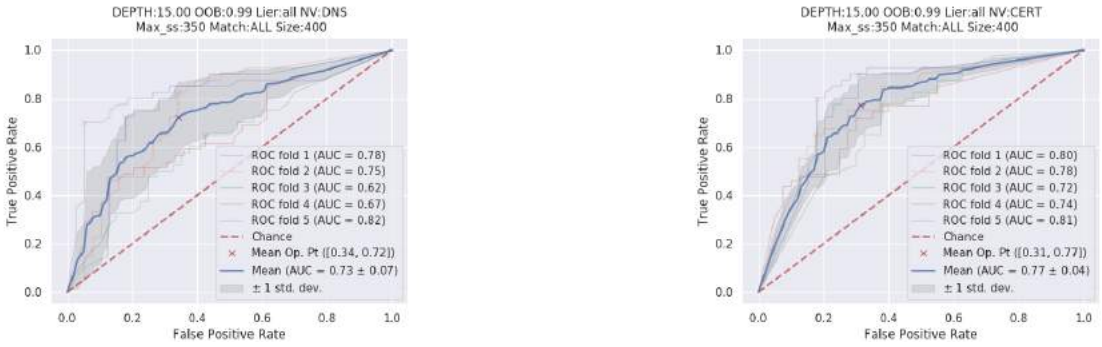
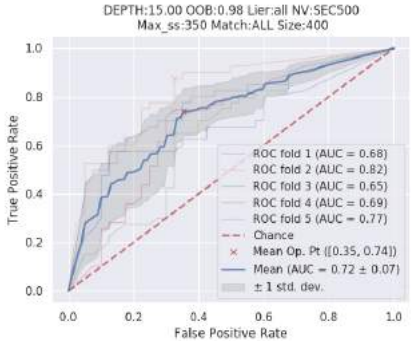


Figure 4.8: Organization classification using the probability distributions from outlier and inlier classifications



(a) Random sampled DNS (left) and random sampled CERT (right)



(b) SEC500

Figure 4.9: Victim vs. non-victim organization classification using all attributions

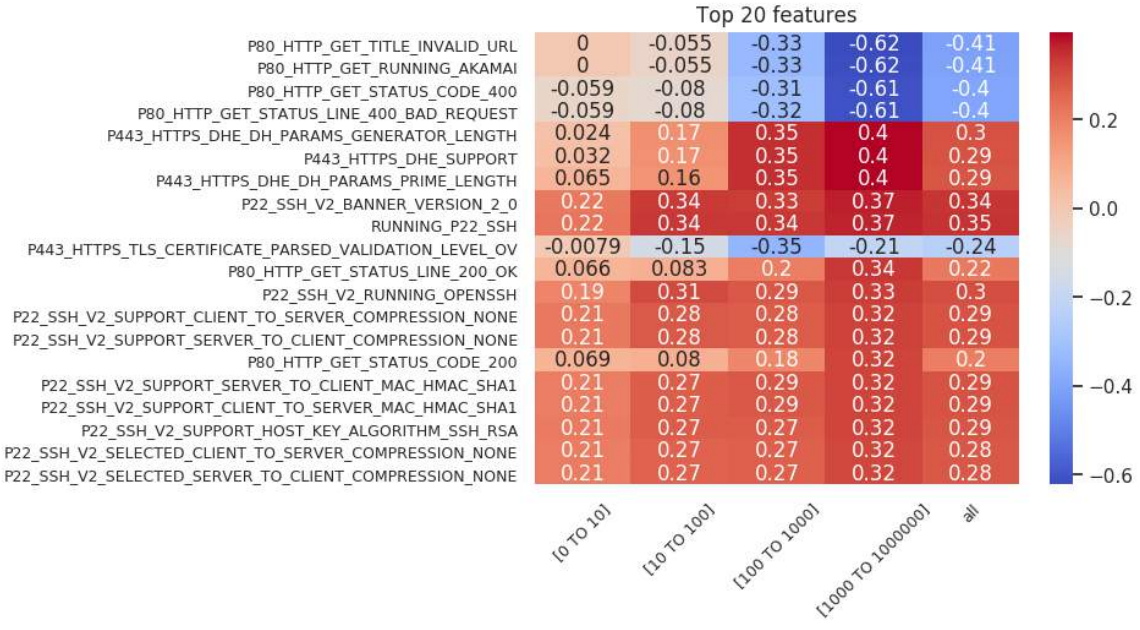


Figure 4.10: Outlier vs. inlier classification feature importance chart using all attributions

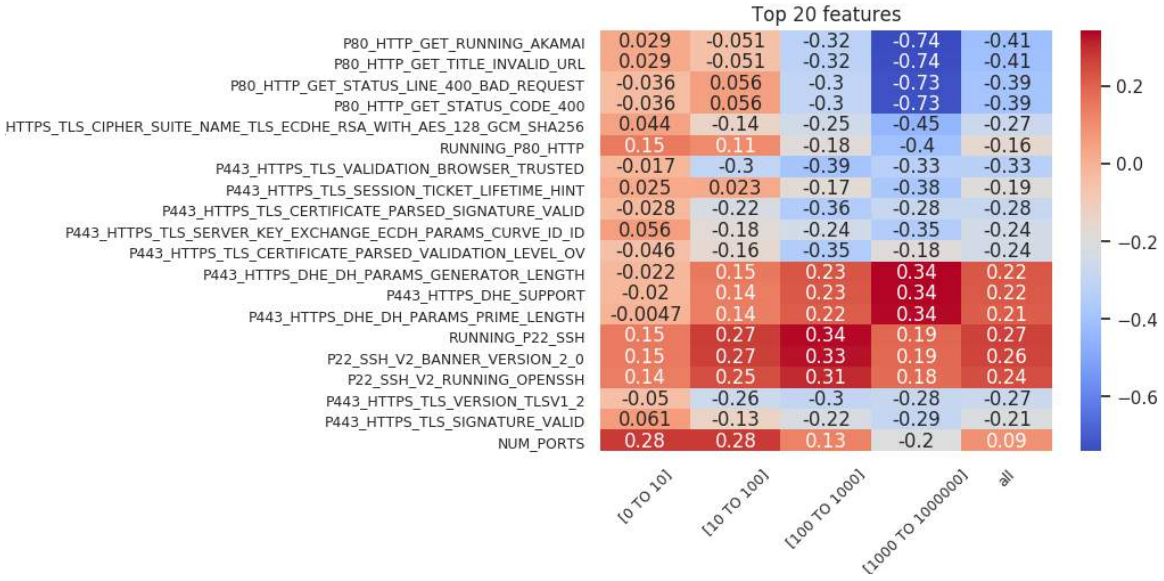


Figure 4.11: Outlier vs. inlier classification feature importance chart using only certificate attributions

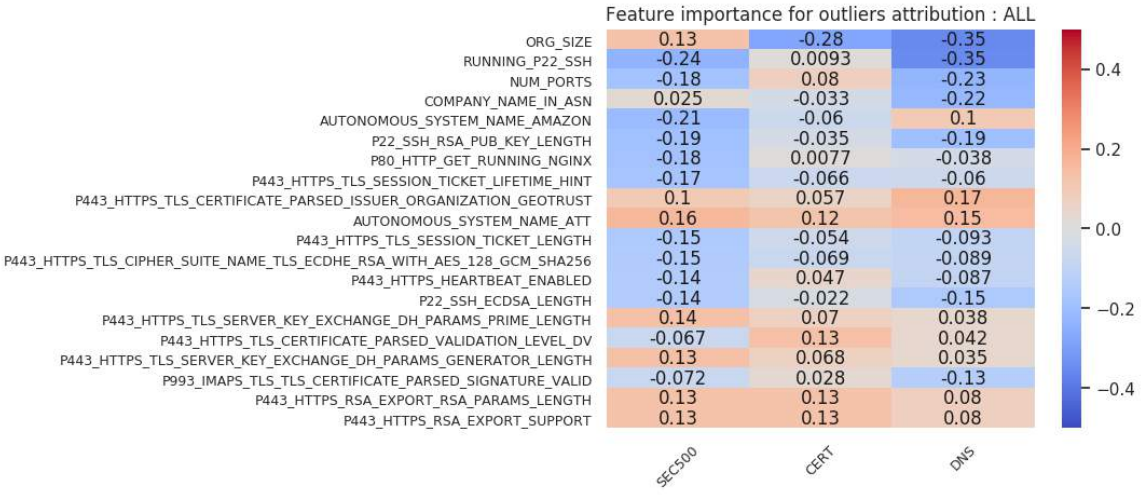


Figure 4.12: Victim vs. non-victim outlier host classification using all attributions

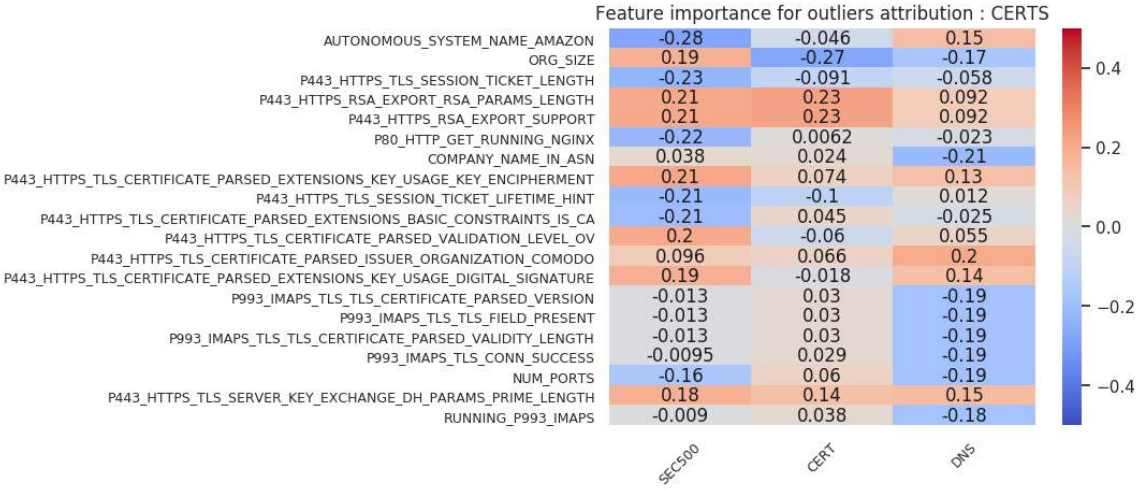


Figure 4.13: Victim vs. non-victim outlier host classification using only certificate attributions

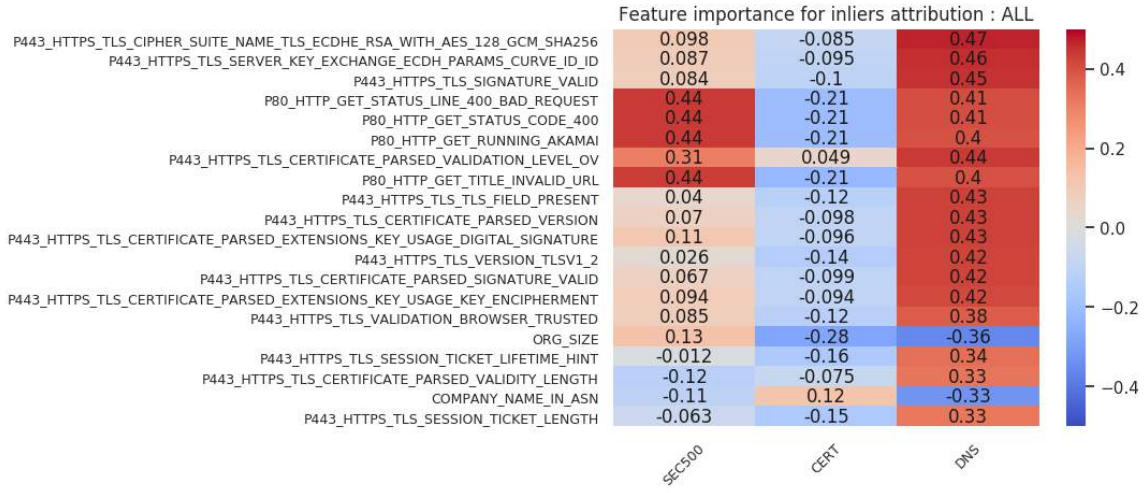


Figure 4.14: Victim vs. non-victim inlier host classification using all attributions

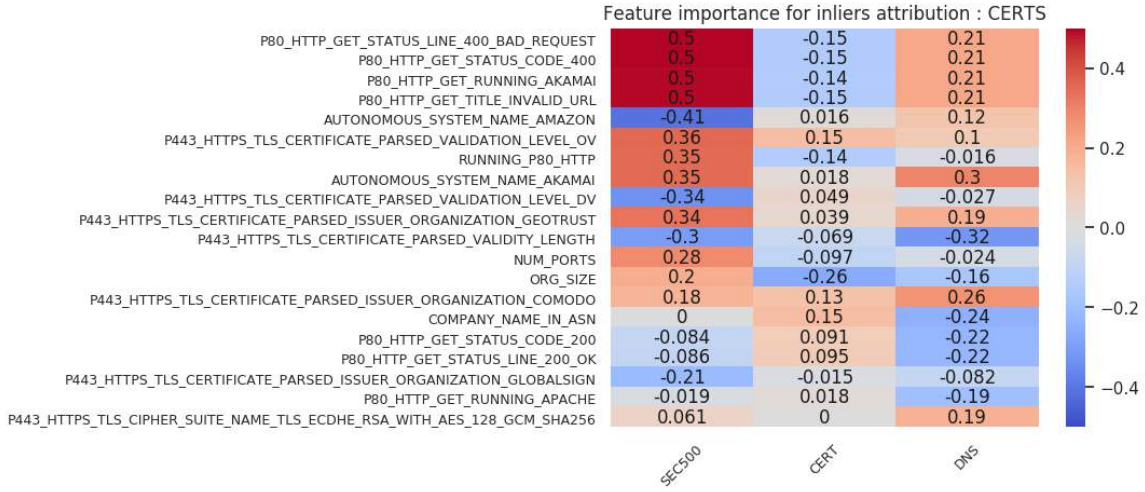


Figure 4.15: Victim vs. non-victim inlier host classification using only certificate attributions

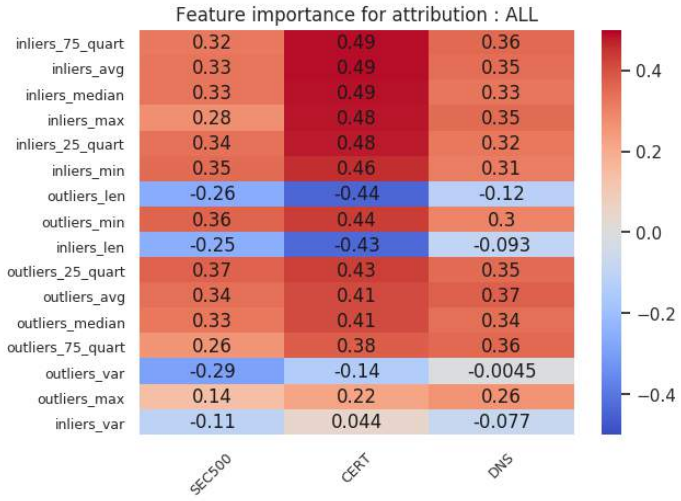


Figure 4.16: Victim vs. non-victim organization classification using all attributions

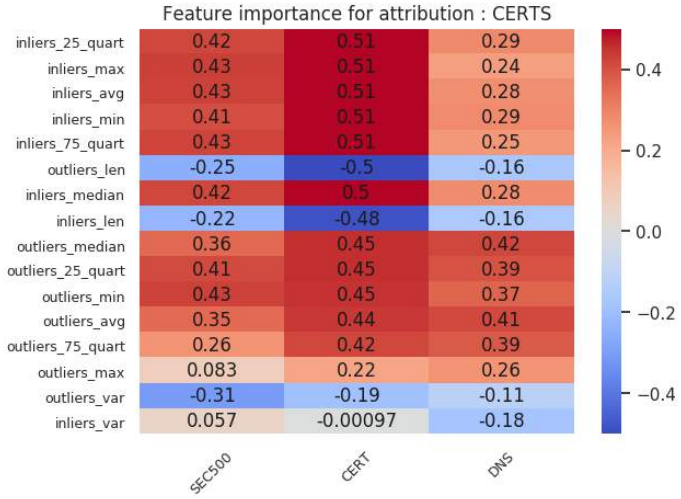


Figure 4.17: Victim vs. non-victim organization classification using only certificate attributions

Chapter 5: Conclusion

5.1 Performance comparison

The works conducted by Soska et al. [34] and Liu et al. [2] (covered in Chapter 2) are the two most similar to our analysis. The performance of our model compared to these works can be seen in Table 5.1.

Table 5.1: Performance comparison with contemporary methods

	Accuracy	TPR	FPR
"Cloudy with a Chance of Breach: Forecasting Cyber Security Incidents", 2015 [2]	0.90	0.90	0.10
"Automatically Detecting Vulnerable Websites Before They Turn Malicious", 2014 [34]	N/A	0.66	0.17
Our Method	0.73	0.77	0.25

From Table 5.1, we can see that Liu et al. have achieved a higher accuracy with a lower fpr than our analysis. We believe this to be attributed to the use of maliciousness features identified by third party sources, as it is much easier to predict the likelihood of a security incident if an organization's network is known to be malicious. We can see this in Figure 5.1. In this figure, the performance of the misconfiguration, most similar to our feature space, performs worse compared to the other features.

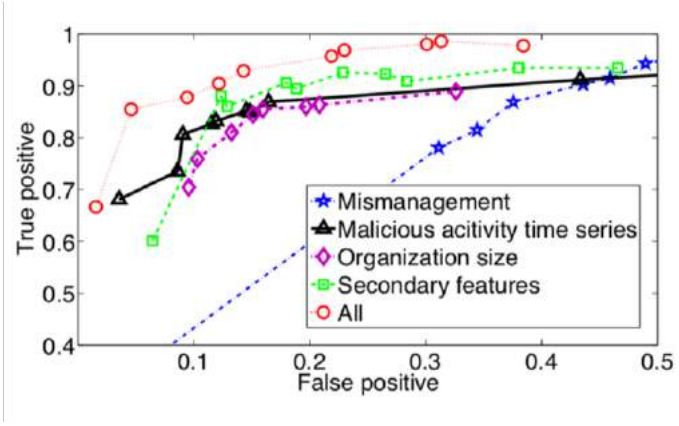


Figure 5.1: Liu et al. model performance of separate features [2]

We have covered the down sides of using RBLs in Section 2.2, however, one avenue for future work is to assign the hosts the RBL label and identify what configurations in our feature space are associated with a host being perceived as malicious.

5.2 Future work

There are many directions to extend the work described in this paper. A subset of these are:

1. In this analysis, we only analyzed digital assets in the form of IPv4 addresses and not the newer IPv6. This was because the data sources and techniques our project utilized were only equipped to handle the former version. One direction for future work is to build a more robust technique that adapts to both IP versions.

2. During the analysis, we make the assumption that the classification problem does not change over time. Meaning, a vulnerability that is present at a certain time in the analysis is present through out the time span of the analysis. However, vulnerabilities are not stationary and they evolve over time. One possible avenue for future work is to create an adaptive model to handle these novel weakness.

3. This project performed the posture analysis by taking a snapshot of an organization's network at a given moment in time. However, previous works [2, 1] analyzed a time series of network snapshots. Time series analysis is an excellent direction for future work as it accounts for the dynamic nature of network configurations over a time period.

4. One direction for future work, mentioned in a previous section, is to assign the RBL labels to the individual hosts. Since Zhang et al. [1] have shown that maliciousness is a cause for security incidents, one could perform a more thorough analysis by assigning these RBL instances to their respective hosts.

5. In our analysis, we represented an organization as a summary statistic vector of the the individual host probabilities. This was a solution to the resolution issue where the features and target label were at disparate levels. However, this is a sub-optimal approach because of the

loss in information about the inter-host configuration for the overall networks. A better approach is to layout the network's hosts in a graphical scheme. In this scenario, the hosts would be nodes and the edges would represent "connectivity" between these hosts. This scheme should, presumably, perform better because it not only captures the host features, but also the inter-host features. As an example if an organization has three hosts running NGINX servers, a graphical based approach would capture in this property where our approach would not. This ideal, more holistic, approach is left as an avenue for future work.

6. As we have previously mentioned, selecting a non-victim is a challenging task. We can see this being prevalent in the inlier analysis for the DNS cohort subset (Figure 4.14) where the presence of HTTPS certificate is correlated with the victim label. Now that we have found that the sampling technique affects the rules that discern victim from non-victim organizations, we need to collect more non-victims using other techniques for analysis. This extension to other incident data sources also applies to victim organizations. Since, in the victim collection stages, it is possible that many security incidents go unreported to more than one incident reporting source [7].

7. In this project, we attempted to introduce a few asset attribution technique to this problem domain. However, identifying all the digital resources an organization owns on the public internet is still an open problem. The reasons for the difficulty are covered in Section 3.2.2. Identifying other attribution techniques is left as a direction for future work.

8. During the asset attribution stage, we resolved a large batch of domain names using massdns [68]. This tool used a list of resolvers that included some malicious servers [74]. We used a quick and conservative calculation to identify these malicious servers. However, one direction is identifying a more reliable method of identifying malicious DNS servers.

9. One challenge with these sorts of analysis is the lack of benchmark data sets to test security incident prediction techniques. With the appearance of Censys [4], hopefully, a standard technique to collect public IPv4 data will be set in the near future. Moreover,one direction for

future work is to aggregate Censys data with other sources of intelligence, e.g., Binary Edge [55].

10. Another direction for future work is to predict what sort of security incident (phishing, malware etc.) an organization will encounter. This will essentially be set as a multi-class classification problem.

11. Reading the incident reports was a tedious process that will not scale well in production. One direction for future work is identifying a more robust way of extracting information from incident reports. Guo et al. has proposed a prediction model for inferring held-out events in stories, particularly stories involving data breaches [75]. A worth-while experiment would be to use the model that we see in [75] to automatically extract the features from incident reports. This would result in a more systematic and uniform incident reporting that solves the challenge mentioned in Liu et al.'s work [2].

Previous works [2, 1, 15] have shown that the security community should pay attention to networks' configuration in order to maintain the health of the public internet. However, identifying a way to predict security incidents within organizations' infrastructure on the internet is a rather difficult endeavour for the many reasons mentioned in this paper. In addition to these challenges, compared to internal information, external posture data does not reveal much about organization state.

In this paper, we collected external network posture information for a cohort of victim and non-victim organizations. We investigate the extent to which these publicly available configurations can be used to predict likelihood of a security incident. Finally, we compare and contrast the performance of the model we built against other contemporary models in the same problem space.

Appendix A:

A.1 Sample security incident notification letter

A.2 Description of some subdomain enumeration techniques

Service record

A Service record (SRV record) is a specification of data in the Domain Name System defining the location, i.e., the hostname and port number, of servers for specified services.

Certificate Transparency Logs

This can be used to find subdomains used by a company in order to search for security vulnerabilities. Some certificate authorities already submit all certificates automatically to public logs. The certificate authority Let's Encrypt and the CDN company Cloudflare submit all certificates to logs voluntarily. But even certificates that aren't logged by their certificate authority or hosted usually end up in the logs quickly, because Google's search engine crawler automatically submits all certificates of sites it finds. The Google Chrome developers had announced that they would require logging of all certificates in September, but the deadline has been moved to April 2018. However, in practice most certificates are already logged.

AXFR / Zone transfer

A simple and basic technique is to try an AXFR request directly on the DNS server.

Listing A.1: AXFR Request

```
~ $ dig @ns.example.com example=.com AXFR
```

A zone transfer is used to copy the content of the zone across primary and secondary DNS servers. The best practice advises administrators to allow AXFR requests only from authorized DNS servers, so the above technique will probably not work. But if it does, it is equivalent to finding a subdomain goldmine.

DNSSEC zonewalk

The Domain Name System Security Extensions (DNSSEC) is a suite of Internet Engineering Task Force (IETF) specifications for securing certain kinds of information provided by the Domain Name System (DNS) as used on Internet Protocol (IP) networks. It is a set of extensions to DNS which provide to DNS clients (resolvers) origin authentication of DNS data, authenticated denial of existence, and data integrity, but not availability or confidentiality. Using NSEC allows anyone to list the zone content by following the linked list of NSEC records. This is called 'zone walking'. Similar to zone transfer, this technique enumerates DNSSEC-signed zones. However, we did not perform any sort of hash cracking of domains that were present and encrypted in the 'Zone Walking' portion.

Google dorking

Many subdomains can be found using web crawling. Google (and also other search engines like Bing) crawl for sub domains as a byproduct of their primary intention. We can use the site operator, *site:example.com*, to find all subdomains that Google has found.

A.3 Exhaustive list of feature space

Listing A.2: Full List of Feature Space

```
[
"RUNNING_P995_POP3S",
"RUNNING_P993_IMAPS",
"RUNNING_P8888_HTTP",
"RUNNING_P80_HTTP",
"RUNNING_P8080_HTTP",
"RUNNING_P7547_CWMP",
"RUNNING_P631_IPP",
"RUNNING_P587_SMTP",
"RUNNING_P5432_POSTGRES",
"RUNNING_P53_DNS",
"RUNNING_P502_MODBUS",
"RUNNING_P47808_BACNET",
```

"RUNNING_P443_HTTPS",

"RUNNING_P3306_MYSQL",

"RUNNING_P25_SMTP",

"RUNNING_P23_TELNET",

"RUNNING_P2323_TELNET",

"RUNNING_P22_SSH",

"RUNNING_P21_FTP",

"RUNNING_P1911_FOX",

"RUNNING_P1900_UPNP",

"RUNNING_P1521_ORACLE",

"RUNNING_P143_IMAP",

"RUNNING_P1433_MSSQL",

"RUNNING_P110_POP3",

"RUNNING_P102_S7",

"P995_POP3S_TLS_TLS_VERSION_TLSV1_2",

"P995_POP3S_TLS_TLS_VERSION_TLSV1_1",

"P995_POP3S_TLS_TLS_VERSION_TLSV1_0",

"P995_POP3S_TLS_TLS_VERSION_SSLV3",

"P995_POP3S_TLS_TLS_VALIDATION_BROWSER_TRUSTED",

"P995_POP3S_TLS_TLS_SIGNATURE_VALID",

"P995_POP3S_TLS_TLS_SERVER_KEY_EXCHANGE_ECDH_PARAMS_CURVE_ID_ID",

"P995_POP3S_TLS_TLS_OCSP_STAPLING",

"P995_POP3S_TLS_TLS_FIELD_PRESENT",

"P995_POP3S_TLS_TLS_CIPHER_SUITE_NAME_TLS_RSA_WITH_RC4_128_SHA",

"P995_POP3S_TLS_TLS_CIPHER_SUITE_NAME_TLS_RSA_WITH_AES_256_CBC_SHA",

"P995_POP3S_TLS_TLS_CIPHER_SUITE_NAME_TLS_RSA_WITH_AES_128_CBC_SHA",

"P995_POP3S_TLS_TLS_CIPHER_SUITE_NAME_TLS_ECDHE_RSA_WITH_AES_256_CBC_SHA",

"P995_POP3S_TLS_TLS_CIPHER_SUITE_NAME_TLS_ECDHE_RSA_WITH_AES_128_GCM_SHA256",

"P995_POP3S_TLS_TLS_CERT_PAST_VALID_END_DATE",

"P995_POP3S_TLS_TLS_CERTIFICATE_PARSED_VERSION",

"P995_POP3S_TLS_TLS_CERTIFICATE_PARSED_VALIDITY_LENGTH",

"P995_POP3S_TLS_TLS_CERTIFICATE_PARSED_VALIDATION_LEVEL_UNKNOWN",

"P995_POP3S_TLS_TLS_CERTIFICATE_PARSED_VALIDATION_LEVEL_OV",

"P995_POP3S_TLS_TLS_CERTIFICATE_PARSED_VALIDATION_LEVEL_EV",

"P995_POP3S_TLS_TLS_CERTIFICATE_PARSED_VALIDATION_LEVEL_DV",

"P995_POP3S_TLS_TLS_CERTIFICATE_PARSED_SIGNATURE_VALID",

"P995_POP3S_TLS_TLS_CERTIFICATE_PARSED_SIGNATURE_SELF_SIGNED",

"P995_POP3S_TLS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_UNKNOWN",

"P995_POP3S_TLS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_STARFIELD",

"P995_POP3S_TLS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_SOMEORGANIZATION",

"P995_POP3S_TLS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_RAPIDSSL",

"P995_POP3S_TLS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_LOCALHOST",

"P995_POP3S_TLS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_LETS_ENCRYPT",

"P995_POP3S_TLS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_HOME_PL",

"P995_POP3S_TLS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_GOOGLE",

"P995_POP3S_TLS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_GODADDY",

"P995_POP3S_TLS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_GLOBALSIGN",

"P995_POP3S_TLS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_GEOTRUST",

"P995_POP3S_TLS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_FORTINET",

"P995_POP3S_TLS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_ENTRUST",

"P995_POP3S_TLS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_DIGICERT",

"P995_POP3S_TLS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_CPANEL",

"P995_POP3S_TLS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_COMODO",

"P995_POP3S_TLS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_CISCO",

"P995_POP3S_TLS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_ALPHASSL",

"P995_POP3S_TLS_TLS_CERTIFICATE_PARSED_EXTENSIONS_KEY_USAGE_KEY_ENCIPHERMENT",

"P995_POP3S_TLS_TLS_CERTIFICATE_PARSED_EXTENSIONS_KEY_USAGE_KEY_AGREEMENT",

"P995_POP3S_TLS_TLS_CERTIFICATE_PARSED_EXTENSIONS_KEY_USAGE_DIGITAL_SIGNATURE",

"P995_POP3S_TLS_TLS_CERTIFICATE_PARSED_EXTENSIONS_KEY_USAGE_DATA_ENCIPHERMENT",

"P995_POP3S_TLS_TLS_CERTIFICATE_PARSED_EXTENSIONS_KEY_USAGE_CRL_SIGN",

"P995_POP3S_TLS_TLS_CERTIFICATE_PARSED_EXTENSIONS_KEY_USAGE_CONTENT_COMMITMENT",

"P995_POP3S_TLS_TLS_CERTIFICATE_PARSED_EXTENSIONS_KEY_USAGE_CERTIFICATE_SIGN",

"P995_POP3S_TLS_TLS_CERTIFICATE_PARSED_EXTENSIONS_BASIC_CONSTRAINTS_IS_CA",

"P995_POP3S_TLS_RUNNING_ZIMBRA",

"P995_POP3S_TLS_RUNNING_XCHANGE",

"P995_POP3S_TLS_RUNNING_SUN",

"P995_POP3S_TLS_RUNNING_QPOPPER",

"P995_POP3S_TLS_RUNNING_ORACLE",

"P995_POP3S_TLS_RUNNING_NETMAIL",

"P995_POP3S_TLS_RUNNING_MICROSOFT_EXCHANGE_SERVER_2007",

"P995_POP3S_TLS_RUNNING_MICROSOFT_EXCHANGE_SERVER_2003",

"P995_POP3S_TLS_RUNNING_MICROSOFT",

"P995_POP3S_TLS_RUNNING_MERCURY",

"P995_POP3S_TLS_RUNNING_MDAEMON",

"P995_POP3S_TLS_RUNNING_KERIO_CONNECT",

"P995_POP3S_TLS_RUNNING_KERIO",

"P995_POP3S_TLS_RUNNING_IMAP",

"P995_POP3S_TLS_RUNNING_ICEWARP",

"P995_POP3S_TLS_RUNNING_GORDANO",

"P995_POP3S_TLS_RUNNING_EXCHANGE_SERVER",

"P995_POP3S_TLS_RUNNING_DOVECOT",

"P995_POP3S_TLS_RUNNING_CYRUS",

"P995_POP3S_TLS_CONN_SUCCESS",

"P995_POP3S_SSL_2_TLS_CERT_PAST_VALID_END_DATE",

"P995_POP3S_SSL_2_SSL_2_SUPPORT",

"P995_POP3S_SSL_2_CERTIFICATE_PARSED_VERSION",

"P995_POP3S_SSL_2_CERTIFICATE_PARSED_SIGNATURE_SELF_SIGNED",

"P995_POP3S_SSL_2_CERTIFICATE_PARSED_EXTENSIONS_KEY_USAGE_KEY_ENCIPHERMENT",

"P995_POP3S_SSL_2_CERTIFICATE_PARSED_EXTENSIONS_KEY_USAGE_DIGITAL_SIGNATURE",

"P995_POP3S_SSL_2_CERTIFICATE_PARSED_EXTENSIONS_BASIC_CONSTRAINTS_IS_CA",

"P993_IMAPS_TLS_TLS_VERSION_TLSV1_2",

"P993_IMAPS_TLS_TLS_VERSION_TLSV1_1",

"P993_IMAPS_TLS_TLS_VERSION_TLSV1_0",

"P993_IMAPS_TLS_TLS_VALIDATION_BROWSER_TRUSTED",

"P993_IMAPS_TLS_TLS_SIGNATURE_VALID",

"P993_IMAPS_TLS_TLS_SERVER_KEY_EXCHANGE_ECDH_PARAMS_CURVE_ID_ID",

"P993_IMAPS_TLS_TLS_OCSP_STAPLING",

"P993_IMAPS_TLS_TLS_FIELD_PRESENT",

"P993_IMAPS_TLS_TLS_CIPHER_SUITE_NAME_TLS_RSA_WITH_RC4_128_SHA",

"P993_IMAPS_TLS_TLS_CIPHER_SUITE_NAME_TLS_RSA_WITH_AES_256_CBC_SHA",

"P993_IMAPS_TLS_TLS_CIPHER_SUITE_NAME_TLS_RSA_WITH_AES_128_CBC_SHA",

"P993_IMAPS_TLS_TLS_CIPHER_SUITE_NAME_TLS_ECDHE_RSA_WITH_AES_256_CBC_SHA",

"P993_IMAPS_TLS_TLS_CIPHER_SUITE_NAME_TLS_ECDHE_RSA_WITH_AES_128_GCM_SHA256" ,

"P993_IMAPS_TLS_TLS_CERT_PAST_VALID_END_DATE" ,

"P993_IMAPS_TLS_TLS_CERTIFICATE_PARSED_VERSION" ,

"P993_IMAPS_TLS_TLS_CERTIFICATE_PARSED_VALIDITY_LENGTH" ,

"P993_IMAPS_TLS_TLS_CERTIFICATE_PARSED_VALIDATION_LEVEL_UNKNOWN" ,

"P993_IMAPS_TLS_TLS_CERTIFICATE_PARSED_VALIDATION_LEVEL_OV" ,

"P993_IMAPS_TLS_TLS_CERTIFICATE_PARSED_VALIDATION_LEVEL_EV" ,

"P993_IMAPS_TLS_TLS_CERTIFICATE_PARSED_VALIDATION_LEVEL_DV" ,

"P993_IMAPS_TLS_TLS_CERTIFICATE_PARSED_SIGNATURE_VALID" ,

"P993_IMAPS_TLS_TLS_CERTIFICATE_PARSED_SIGNATURE_SELF_SIGNED" ,

"P993_IMAPS_TLS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_UNKNOWN" ,

"P993_IMAPS_TLS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_STARFIELD" ,

"P993_IMAPS_TLS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_SOMEORGANIZATION" ,

"P993_IMAPS_TLS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_RAPIDSSL" ,

"P993_IMAPS_TLS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_LOCALHOST" ,

"P993_IMAPS_TLS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_LETS_ENCRYPT" ,

"P993_IMAPS_TLS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_HOME_PL" ,

"P993_IMAPS_TLS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_GOOGLE" ,

"P993_IMAPS_TLS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_GODADDY" ,

"P993_IMAPS_TLS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_GLOBALSIGN" ,

"P993_IMAPS_TLS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_GEOTRUST" ,

"P993_IMAPS_TLS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_FORTINET" ,

"P993_IMAPS_TLS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_ENTRUST" ,

"P993_IMAPS_TLS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_DIGICERT" ,

"P993_IMAPS_TLS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_CPANEL" ,

"P993_IMAPS_TLS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_COMODO" ,

"P993_IMAPS_TLS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_CISCO" ,

"P993_IMAPS_TLS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_ALPHASSL" ,

"P993_IMAPS_TLS_TLS_CERTIFICATE_PARSED_EXTENSIONS_KEY_USAGE_KEY_ENCIPHERMENT" ,

"P993_IMAPS_TLS_TLS_CERTIFICATE_PARSED_EXTENSIONS_KEY_USAGE_KEY_AGREEMENT" ,

"P993_IMAPS_TLS_TLS_CERTIFICATE_PARSED_EXTENSIONS_KEY_USAGE_DIGITAL_SIGNATURE" ,

"P993_IMAPS_TLS_TLS_CERTIFICATE_PARSED_EXTENSIONS_KEY_USAGE_DATA_ENCIPHERMENT" ,

"P993_IMAPS_TLS_TLS_CERTIFICATE_PARSED_EXTENSIONS_KEY_USAGE_CONTENT_COMMITMENT" ,

"P993_IMAPS_TLS_TLS_CERTIFICATE_PARSED_EXTENSIONS_KEY_USAGE_CERTIFICATE_SIGN" ,

"P993_IMAPS_TLS_TLS_CERTIFICATE_PARSED_EXTENSIONS_BASIC_CONSTRAINTS_IS_CA",

"P993_IMAPS_TLS_RUNNING_ZIMBRA",

"P993_IMAPS_TLS_RUNNING_XCHANGE",

"P993_IMAPS_TLS_RUNNING_SUN",

"P993_IMAPS_TLS_RUNNING_ORACLE",

"P993_IMAPS_TLS_RUNNING_MICROSOFT_EXCHANGE_SERVER_2003",

"P993_IMAPS_TLS_RUNNING_MICROSOFT",

"P993_IMAPS_TLS_RUNNING_MERCURY",

"P993_IMAPS_TLS_RUNNING_MDAEMON",

"P993_IMAPS_TLS_RUNNING_KERIO_CONNECT",

"P993_IMAPS_TLS_RUNNING_KERIO",

"P993_IMAPS_TLS_RUNNING_IMAP",

"P993_IMAPS_TLS_RUNNING_ICEWARP",

"P993_IMAPS_TLS_RUNNING_GROUPWISE",

"P993_IMAPS_TLS_RUNNING_EXCHANGE_SERVER",

"P993_IMAPS_TLS_RUNNING_DOVECOT",

"P993_IMAPS_TLS_RUNNING_CYRUS",

"P993_IMAPS_TLS_RUNNING_COURIER",

"P993_IMAPS_TLS_RUNNING_AXIGEN",

"P993_IMAPS_TLS_CONN_SUCCESS",

"P993_IMAPS_SSL_2_TLS_CERT_PAST_VALID_END_DATE",

"P993_IMAPS_SSL_2_SSL_2_SUPPORT",

"P993_IMAPS_SSL_2_CERTIFICATE_PARSED_VERSION",

"P993_IMAPS_SSL_2_CERTIFICATE_PARSED_SIGNATURE_SELF_SIGNED",

"P993_IMAPS_SSL_2_CERTIFICATE_PARSED_EXTENSIONS_KEY_USAGE_KEY_ENCIPHERMENT",

"P993_IMAPS_SSL_2_CERTIFICATE_PARSED_EXTENSIONS_KEY_USAGE_DIGITAL_SIGNATURE",

"P993_IMAPS_SSL_2_CERTIFICATE_PARSED_EXTENSIONS_BASIC_CONSTRAINTS_IS_CA",

"P8888_HTTP_GET_TITLE_UNAUTHORIZED",

"P8888_HTTP_GET_TITLE_NOT_FOUND",

"P8888_HTTP_GET_TITLE_IIS7",

"P8888_HTTP_GET_TITLE_ERROR_THE_REQUESTED_URL_COULD_NOT_BE_RETRIEVED",

"P8888_HTTP_GET_TITLE_404_NOT_FOUND",

"P8888_HTTP_GET_TITLE_403_FORBIDDEN",

"P8888_HTTP_GET_TITLE_401_UNAUTHORIZED",

"P8888_HTTP_GET_STATUS_LINE_503_SERVICE_UNAVAILABLE",

"P8888_HTTP_GET_STATUS_LINE_404_NOT_FOUND",

"P8888_HTTP_GET_STATUS_LINE_403_FORBIDDEN",

"P8888_HTTP_GET_STATUS_LINE_401_UNAUTHORIZED",

"P8888_HTTP_GET_STATUS_LINE_400_BAD_REQUEST",

"P8888_HTTP_GET_STATUS_LINE_302_FOUND",

"P8888_HTTP_GET_STATUS_LINE_200_OK",

"P8888_HTTP_GET_STATUS_CODE_503",

"P8888_HTTP_GET_STATUS_CODE_404",

"P8888_HTTP_GET_STATUS_CODE_403",

"P8888_HTTP_GET_STATUS_CODE_401",

"P8888_HTTP_GET_STATUS_CODE_400",

"P8888_HTTP_GET_STATUS_CODE_302",

"P8888_HTTP_GET_STATUS_CODE_200",

"P8888_HTTP_GET_RUNNING_ZOPE",

"P8888_HTTP_GET_RUNNING_TINYPROXY",

"P8888_HTTP_GET_RUNNING_SQUID",

"P8888_HTTP_GET_RUNNING_ORACLE",

"P8888_HTTP_GET_RUNNING_NGINX",

"P8888_HTTP_GET_RUNNING_MINIUPNP",

"P8888_HTTP_GET_RUNNING_MICROSOFT",

"P8888_HTTP_GET_RUNNING_LOCALHOST",

"P8888_HTTP_GET_RUNNING_LITESPEED",

"P8888_HTTP_GET_RUNNING_LINUX",

"P8888_HTTP_GET_RUNNING_LIGHTTPD",

"P8888_HTTP_GET_RUNNING_KANGLE",

"P8888_HTTP_GET_RUNNING_JETTY",

"P8888_HTTP_GET_RUNNING_IIS",

"P8888_HTTP_GET_RUNNING_HTTPAPI",

"P8888_HTTP_GET_RUNNING_COYOTE",

"P8888_HTTP_GET_RUNNING_APP",

"P8888_HTTP_GET_RUNNING_APACHE",

"P8888_HTTP_GET_BODY_SHA256_CE7127C38E30E92A021ED2BD09287713C6A923DB9FFDB43F12",

"P8888_HTTP_GET_BODY_SHA256_38FFD4972AE513A0C79A8BE4573403EDCD709F0F572105362B",

"P8888_HTTP_GET_BODY_SHA256_370BE45F65276B3B8DE42A29ADFB1220FC44A5E018C37E3E9B",

"P80_HTTP_GET_TITLE_UNKNOWN",

"P80_HTTP_GET_TITLE_UNAUTHORIZED",

"P80_HTTP_GET_TITLE_OBJECT_NOT_FOUND",

"P80_HTTP_GET_TITLE_NOT_FOUND",

"P80_HTTP_GET_TITLE_INVALID_URL",

"P80_HTTP_GET_TITLE_IIS7",

"P80_HTTP_GET_TITLE_ERROR_THE_REQUESTED_URL_COULD_NOT_BE_RETRIEVED",

"P80_HTTP_GET_TITLE_DIRECT_IP_ACCESS_NOT_ALLOWED_|_CLOUDFLARE",

"P80_HTTP_GET_TITLE_APACHE_HTTP_SERVER_TEST_PAGE_POWERED_BY_CENTOS",

"P80_HTTP_GET_TITLE_404_NOT_FOUND",

"P80_HTTP_GET_TITLE_403_FORBIDDEN",

"P80_HTTP_GET_TITLE_401_UNAUTHORIZED",

"P80_HTTP_GET_TITLE_401_NOT_AUTHORIZED",

"P80_HTTP_GET_TITLE_401_AUTHORIZATION_REQUIRED",

"P80_HTTP_GET_STATUS_LINE_UNKNOWN",

"P80_HTTP_GET_STATUS_LINE_503_SERVICE_UNAVAILABLE",

"P80_HTTP_GET_STATUS_LINE_502_BAD_GATEWAY",

"P80_HTTP_GET_STATUS_LINE_500_INTERNAL_SERVER_ERROR",

"P80_HTTP_GET_STATUS_LINE_404_NOT_FOUND",

"P80_HTTP_GET_STATUS_LINE_403_FORBIDDEN",

"P80_HTTP_GET_STATUS_LINE_401_UNAUTHORIZED",

"P80_HTTP_GET_STATUS_LINE_401_NOT_AUTHORIZED",

"P80_HTTP_GET_STATUS_LINE_401_AUTHORIZATION_REQUIRED",

"P80_HTTP_GET_STATUS_LINE_400_BAD_REQUEST",

"P80_HTTP_GET_STATUS_LINE_302_FOUND",

"P80_HTTP_GET_STATUS_LINE_200_OK",

"P80_HTTP_GET_STATUS_CODE_503",

"P80_HTTP_GET_STATUS_CODE_502",

"P80_HTTP_GET_STATUS_CODE_500",

"P80_HTTP_GET_STATUS_CODE_479",

"P80_HTTP_GET_STATUS_CODE_404",

"P80_HTTP_GET_STATUS_CODE_403",

"P80_HTTP_GET_STATUS_CODE_401",

"P80_HTTP_GET_STATUS_CODE_400",

"P80_HTTP_GET_STATUS_CODE_303",

"P80_HTTP_GET_STATUS_CODE_302",

"P80_HTTP_GET_STATUS_CODE_301",

"P80_HTTP_GET_STATUS_CODE_204",

"P80_HTTP_GET_STATUS_CODE_200",

"P80_HTTP_GET_RUNNING_ZOPE",

"P80_HTTP_GET_RUNNING_ZEUS",

"P80_HTTP_GET_RUNNING_XITAMI",

"P80_HTTP_GET_RUNNING_WEBRICK",

"P80_HTTP_GET_RUNNING_VIRTUOSO",

"P80_HTTP_GET_RUNNING_VARNISH",

"P80_HTTP_GET_RUNNING_UNKNOWN",

"P80_HTTP_GET_RUNNING_TOMCAT",

"P80_HTTP_GET_RUNNING_THTTPD",

"P80_HTTP_GET_RUNNING_TENGINE",

"P80_HTTP_GET_RUNNING_SQUID",

"P80_HTTP_GET_RUNNING_SONICWALL",

"P80_HTTP_GET_RUNNING_RESIN",

"P80_HTTP_GET_RUNNING_PWS",

"P80_HTTP_GET_RUNNING_ORACLE",

"P80_HTTP_GET_RUNNING_NGINX",

"P80_HTTP_GET_RUNNING_MONKEY",

"P80_HTTP_GET_RUNNING_MICROSOFT",

"P80_HTTP_GET_RUNNING_LOCALHOST",

"P80_HTTP_GET_RUNNING_LITESPEED",

"P80_HTTP_GET_RUNNING_LINUX",

"P80_HTTP_GET_RUNNING_LIGHTTPD",

"P80_HTTP_GET_RUNNING_KANGLE",

"P80_HTTP_GET_RUNNING_JETTY",

"P80_HTTP_GET_RUNNING_IPLANET",

"P80_HTTP_GET_RUNNING_INTEL",

"P80_HTTP_GET_RUNNING_IIS",

"P80_HTTP_GET_RUNNING_IBM",

"P80_HTTP_GET_RUNNING_HTTP_SERVER",

"P80_HTTP_GET_RUNNING_HTTPAPI",

"P80_HTTP_GET_RUNNING_GOAHEAD",

"P80_HTTP_GET_RUNNING_GLASSFISH",

"P80_HTTP_GET_RUNNING_DVRDVS",

"P80_HTTP_GET_RUNNING_DNVRS",

"P80_HTTP_GET_RUNNING_COYOTE",

"P80_HTTP_GET_RUNNING_CLOUDFRONT",

"P80_HTTP_GET_RUNNING_CLOUDFLARE",

"P80_HTTP_GET_RUNNING_CISCO",

"P80_HTTP_GET_RUNNING_CHEROKEE",

"P80_HTTP_GET_RUNNING_CENTOS",

"P80_HTTP_GET_RUNNING_CADDY",

"P80_HTTP_GET_RUNNING_BOA",

"P80_HTTP_GET_RUNNING_APP",

"P80_HTTP_GET_RUNNING_APACHE",

"P80_HTTP_GET_RUNNING_AOL",

"P80_HTTP_GET_RUNNING_AKAMAI",

"P80_HTTP_GET_BODY_SHA256_F33C27745F2BD87344BE790465EF984A972FD539DC83BD4F61",

"P80_HTTP_GET_BODY_SHA256_CE7127C38E30E92A021ED2BD09287713C6A923DB9FFDB43F12",

"P80_HTTP_GET_BODY_SHA256_9278D16ED2FDCD5DC651615B0B8ADC6B55FB667A9D106A9891",

"P80_HTTP_GET_BODY_SHA256_8B71379A4C9449B0D652659F4D7DA15D904B2744CEE3C0B17D",

"P80_HTTP_GET_BODY_SHA256_5A51100A730D5CA4B14540E26595B73CCE5B7CACFB3FA24359",

"P80_HTTP_GET_BODY_SHA256_38FFD4972AE513A0C79A8BE4573403EDCD709F0F572105362B",

"P80_HTTP_GET_BODY_SHA256_370BE45F65276B3B8DE42A29ADFB1220FC44A5E018C37E3E9B",

"P80_HTTP_GET_BODY_SHA256_2C3ADC6B6FB69D3A4E7B75B64E913DC96D21DBAF436BF69E77",

"P80_HTTP_GET_BODY_SHA256_29A8B2A2DBAC349F919923D25AF4F9162BC58C29B2DAAC41A5",

"P80_HTTP_GET_BODY_SHA256_1D08335E65DA7CF40D1C4A7BA0088E0F39B9C5A4B2E42DE95F",

"P8080_HTTP_GET_TITLE_UNAUTHORIZED",

"P8080_HTTP_GET_TITLE_NOT_FOUND",

"P8080_HTTP_GET_TITLE_IIS7",

"P8080_HTTP_GET_TITLE_ERROR_THE_REQUESTED_URL_COULD_NOT_BE_RETRIEVED",

"P8080_HTTP_GET_TITLE_DIRECT_IP_ACCESS_NOT_ALLOWED_|_CLOUDFLARE",

"P8080_HTTP_GET_TITLE_APACHE_HTTP_SERVER_TEST_PAGE_POWERED_BY_CENTOS",

"P8080_HTTP_GET_TITLE_404_NOT_FOUND",

"P8080_HTTP_GET_TITLE_403_FORBIDDEN",

"P8080_HTTP_GET_TITLE_401_UNAUTHORIZED",

"P8080_HTTP_GET_TITLE_401_NOT_AUTHORIZED",

"P8080_HTTP_GET_TITLE_401_AUTHORIZATION_REQUIRED",

"P8080_HTTP_GET_STATUS_LINE_UNKNOWN",

"P8080_HTTP_GET_STATUS_LINE_503_SERVICE_UNAVAILABLE",

"P8080_HTTP_GET_STATUS_LINE_502_BAD_GATEWAY",

"P8080_HTTP_GET_STATUS_LINE_500_INTERNAL_SERVER_ERROR",

"P8080_HTTP_GET_STATUS_LINE_404_NOT_FOUND",

"P8080_HTTP_GET_STATUS_LINE_404_CLIENT_ERROR",

"P8080_HTTP_GET_STATUS_LINE_403_FORBIDDEN",

"P8080_HTTP_GET_STATUS_LINE_401_UNAUTHORIZED",

"P8080_HTTP_GET_STATUS_LINE_401_AUTHORIZATION_REQUIRED",

"P8080_HTTP_GET_STATUS_LINE_400_BAD_REQUEST",

"P8080_HTTP_GET_STATUS_LINE_302_FOUND",

"P8080_HTTP_GET_STATUS_LINE_200_OK",

"P8080_HTTP_GET_STATUS_CODE_503",

"P8080_HTTP_GET_STATUS_CODE_502",

"P8080_HTTP_GET_STATUS_CODE_500",

"P8080_HTTP_GET_STATUS_CODE_404",

"P8080_HTTP_GET_STATUS_CODE_403",

"P8080_HTTP_GET_STATUS_CODE_401",

"P8080_HTTP_GET_STATUS_CODE_400",

"P8080_HTTP_GET_STATUS_CODE_302",

"P8080_HTTP_GET_STATUS_CODE_200",

"P8080_HTTP_GET_RUNNING_ZOPE",

"P8080_HTTP_GET_RUNNING_YAWS",

"P8080_HTTP_GET_RUNNING_VARNISH",

"P8080_HTTP_GET_RUNNING_UNKNOWN",

"P8080_HTTP_GET_RUNNING_TOMCAT",

"P8080_HTTP_GET_RUNNING_TINYPROXY",

"P8080_HTTP_GET_RUNNING_THTTPD",

"P8080_HTTP_GET_RUNNING_TENGINE",

"P8080_HTTP_GET_RUNNING_SQUID",

"P8080_HTTP_GET_RUNNING_SONICWALL",

"P8080_HTTP_GET_RUNNING_PWS",

"P8080_HTTP_GET_RUNNING_ORACLE",

"P8080_HTTP_GET_RUNNING_NGINX",

"P8080_HTTP_GET_RUNNING_MICROSOFT",

"P8080_HTTP_GET_RUNNING_LOCALHOST",

"P8080_HTTP_GET_RUNNING_LINUX",

"P8080_HTTP_GET_RUNNING_LIGHTTPD",

"P8080_HTTP_GET_RUNNING_KANGLE",

"P8080_HTTP_GET_RUNNING_JETTY",

"P8080_HTTP_GET_RUNNING_IIS",

"P8080_HTTP_GET_RUNNING_HTTP_SERVER",

"P8080_HTTP_GET_RUNNING_HTTPAPI",

"P8080_HTTP_GET_RUNNING_GOAHEAD",

"P8080_HTTP_GET_RUNNING_GLASSFISH",

"P8080_HTTP_GET_RUNNING_DVRDVS",

"P8080_HTTP_GET_RUNNING_DNVRS",

"P8080_HTTP_GET_RUNNING_COYOTE",

"P8080_HTTP_GET_RUNNING_CLOUDFLARE",

"P8080_HTTP_GET_RUNNING_CISCO",

"P8080_HTTP_GET_RUNNING_CENTOS",

"P8080_HTTP_GET_RUNNING_BOA",

"P8080_HTTP_GET_RUNNING_APP",

"P8080_HTTP_GET_RUNNING_APACHE",

"P8080_HTTP_GET_BODY_SHA256_F33C27745F2BD87344BE790465EF984A972FD539DC83BD4F61",

"P8080_HTTP_GET_BODY_SHA256_CE7127C38E30E92A021ED2BD09287713C6A923DB9FFDB43F12",

"P8080_HTTP_GET_BODY_SHA256_8B71379A4C9449B0D652659F4D7DA15D904B2744CEE3C0B17D",

"P8080_HTTP_GET_BODY_SHA256_5A51100A730D5CA4B14540E26595B73CCE5B7CACFB3FA24359",

"P8080_HTTP_GET_BODY_SHA256_38FFD4972AE513A0C79A8BE4573403EDCD709F0F572105362B",

"P8080_HTTP_GET_BODY_SHA256_370BE45F65276B3B8DE42A29ADFB1220FC44A5E018C37E3E9B",

"P8080_HTTP_GET_BODY_SHA256_2C3ADC6B6FB69D3A4E7B75B64E913DC96D21DBAF436BF69E77",

"P8080_HTTP_GET_BODY_SHA256_1D08335E65DA7CF40D1C4A7BA0088E0F39B9C5A4B2E42DE95F",

"P7547_CWMP_GET_TITLE_UNAUTHORIZED",

"P7547_CWMP_GET_TITLE_NOT_FOUND",

"P7547_CWMP_GET_TITLE_401_UNAUTHORIZED",

"P7547_CWMP_GET_STATUS_LINE_404_NOT_FOUND",

"P7547_CWMP_GET_STATUS_LINE_403_FORBIDDEN",

"P7547_CWMP_GET_STATUS_LINE_401_UNAUTHORIZED",

"P7547_CWMP_GET_STATUS_LINE_401_AUTHORIZATION_REQUIRED",

"P7547_CWMP_GET_STATUS_LINE_200_OK",

"P7547_CWMP_GET_STATUS_CODE_404",

"P7547_CWMP_GET_STATUS_CODE_403",

"P7547_CWMP_GET_STATUS_CODE_401",

"P7547_CWMP_GET_STATUS_CODE_301",

"P7547_CWMP_GET_STATUS_CODE_200",

"P7547_CWMP_GET_RUNNING_TR069",

"P7547_CWMP_GET_RUNNING_TORNADO",

"P7547_CWMP_GET_RUNNING_ROMPAGER",

"P7547_CWMP_GET_RUNNING_MICROSOFT",

"P7547_CWMP_GET_RUNNING_IIS",

"P7547_CWMP_GET_RUNNING_GSOAP",

"P7547_CWMP_GET_BODY_SHA256_FE164298CDC47A2C6BE40E5F9101B12EB7157387A9BCBA3CAE",

"P7547_CWMP_GET_BODY_SHA256_5A51100A730D5CA4B14540E26595B73CCE5B7CACFB3FA24359",

"P631_IPP_BANNER_VERSION_STRING_2_1",

"P631_IPP_BANNER_VERSION_STRING_2_0",

"P631_IPP_BANNER_VERSION_STRING_1_1",

"P631_IPP_BANNER_VERSION_STRING_1_0",

"P631_IPP_BANNER_SUPPORTED",

"P631_IPP_BANNER_CUPS_VERSION_2_2_1",

"P631_IPP_BANNER_CUPS_VERSION_2_2",

"P631_IPP_BANNER_CUPS_VERSION_2_1_3",

"P631_IPP_BANNER_CUPS_VERSION_2_1",

"P631_IPP_BANNER_CUPS_VERSION_2_0_3",

"P631_IPP_BANNER_CUPS_VERSION_2_0",

"P631_IPP_BANNER_CUPS_VERSION_1_7_2",

"P631_IPP_BANNER_CUPS_VERSION_1_7",

"P631_IPP_BANNER_CUPS_VERSION_1_6",

"P631_IPP_BANNER_CUPS_VERSION_1_5_3",

"P631_IPP_BANNER_CUPS_VERSION_1_5",

"P631_IPP_BANNER_CUPS_VERSION_1_4",

"P631_IPP_BANNER_CUPS_VERSION_1_3",

"P631_IPP_BANNER_CUPS_VERSION_1_2",

"P587_SMTP_TLS_TLS_FIELD_PRESENT",

"P587_SMTP_TLS_TLS_CERT_PAST_VALID_END_DATE",

"P587_SMTP_STARTTLS_TLS_VERSION_TLSV1_2",

"P587_SMTP_STARTTLS_TLS_VERSION_TLSV1_0",

"P587_SMTP_STARTTLS_TLS_VALIDATION_BROWSER_TRUSTED",

"P587_SMTP_STARTTLS_TLS_SIGNATURE_VALID",

"P587_SMTP_STARTTLS_TLS_SERVER_KEY_EXCHANGE_ECDH_PARAMS_CURVE_ID_ID",

"P587_SMTP_STARTTLS_TLS_OCSP_STAPLING",

"P587_SMTP_STARTTLS_TLS_CIPHER_SUITE_NAME_TLS_RSA_WITH_RC4_128_SHA",

"P587_SMTP_STARTTLS_TLS_CIPHER_SUITE_NAME_TLS_RSA_WITH_AES_256_CBC_SHA",

"P587_SMTP_STARTTLS_TLS_CIPHER_SUITE_NAME_TLS_RSA_WITH_AES_128_CBC_SHA",

"P587_SMTP_STARTTLS_TLS_CIPHER_SUITE_NAME_TLS_ECDHE_RSA_WITH_AES_256_CBC_SHA",

"P587_SMTP_STARTTLS_TLS_CIPHER_SUITE_NAME_TLS_ECDHE_RSA_WITH_AES_128_GCM_SHA256",

"P587_SMTP_STARTTLS_TLS_CERTIFICATE_PARSED_VERSION",

"P587_SMTP_STARTTLS_TLS_CERTIFICATE_PARSED_VALIDITY_LENGTH",

"P587_SMTP_STARTTLS_TLS_CERTIFICATE_PARSED_VALIDATION_LEVEL_UNKNOWN",

"P587_SMTP_STARTTLS_TLS_CERTIFICATE_PARSED_VALIDATION_LEVEL_OV",

"P587_SMTP_STARTTLS_TLS_CERTIFICATE_PARSED_VALIDATION_LEVEL_DV",

"P587_SMTP_STARTTLS_TLS_CERTIFICATE_PARSED_SIGNATURE_VALID",

"P587_SMTP_STARTTLS_TLS_CERTIFICATE_PARSED_SIGNATURE_SELF_SIGNED",

"P587_SMTP_STARTTLS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_STARFIELD",

"P587_SMTP_STARTTLS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_SOMEORGANIZATION",

"P587_SMTP_STARTTLS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_RAPIDSSL",

"P587_SMTP_STARTTLS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_LOCALHOST",

"P587_SMTP_STARTTLS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_LETS_ENCRYPT",

"P587_SMTP_STARTTLS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_HOME_PL",

"P587_SMTP_STARTTLS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_GODADDY",

"P587_SMTP_STARTTLS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_GLOBALSIGN",

"P587_SMTP_STARTTLS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_GEOTRUST",

"P587_SMTP_STARTTLS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_DIGICERT",

"P587_SMTP_STARTTLS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_CPANEL",

"P587_SMTP_STARTTLS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_COMODO",

"P587_SMTP_STARTTLS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_CISCO",

"P587_SMTP_STARTTLS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_ALPHASSL",

"P587_SMTP_STARTTLS_TLS_CERTIFICATE_PARSED_EXTENSIONS_KEY_USAGE_KEY_ENCIPHERMENT",

"P587_SMTP_STARTTLS_TLS_CERTIFICATE_PARSED_EXTENSIONS_KEY_USAGE_DIGITAL_SIGNATURE",

"P587_SMTP_STARTTLS_TLS_CERTIFICATE_PARSED_EXTENSIONS_KEY_USAGE_DATA_ENCIPHERMENT",

"P587_SMTP_STARTTLS_TLS_CERTIFICATE_PARSED_EXTENSIONS_KEY_USAGE_CONTENT_COMMITMENT",

"P587_SMTP_STARTTLS_TLS_CERTIFICATE_PARSED_EXTENSIONS_KEY_USAGE_CERTIFICATE_SIGN",

"P587_SMTP_STARTTLS_TLS_CERTIFICATE_PARSED_EXTENSIONS_BASIC_CONSTRAINTS_IS_CA",

"P587_SMTP_STARTTLS_RUNNING_XCHANGE",

"P587_SMTP_STARTTLS_RUNNING_SUN",

"P587_SMTP_STARTTLS_RUNNING_SENDMAIL",

"P587_SMTP_STARTTLS_RUNNING_POSTFIX",

"P587_SMTP_STARTTLS_RUNNING_MICROSOFT",

"P587_SMTP_STARTTLS_RUNNING_MERCURY",

"P587_SMTP_STARTTLS_RUNNING_MAILENABLE",

"P587_SMTP_STARTTLS_RUNNING_IRONPORT",

"P587_SMTP_STARTTLS_RUNNING_HMAIL",

"P587_SMTP_STARTTLS_RUNNING_EXIM",

"P587_SMTP_STARTTLS_RUNNING_EXCHANGE_SERVER",

"P587_SMTP_STARTTLS_CONN_SUCCESS",

"P5432_POSTGRES_TLS_TLS_FIELD_PRESENT",

"P5432_POSTGRES_BANNER_TLS_VERSION_TLSV1_2",

"P5432_POSTGRES_BANNER_TLS_VERSION_TLSV1_0",

"P5432_POSTGRES_BANNER_TLS_SIGNATURE_VALID",

"P5432_POSTGRES_BANNER_TLS_SERVER_KEY_EXCHANGE_ECDH_PARAMS_CURVE_ID_ID",

"P5432_POSTGRES_BANNER_TLS_CIPHER_SUITE_NAME_TLS_RSA_WITH_RC4_128_SHA",

"P5432_POSTGRES_BANNER_TLS_CIPHER_SUITE_NAME_TLS_RSA_WITH_AES_128_CBC_SHA",

"P5432_POSTGRES_BANNER_TLS_CIPHER_SUITE_NAME_TLS_ECDHE_RSA_WITH_AES_256_CBC_SHA",

"P5432_POSTGRES_BANNER_TLS_CERTIFICATE_PARSED_VERSION",

"P5432_POSTGRES_BANNER_TLS_CERTIFICATE_PARSED_VALIDITY_LENGTH",

"P5432_POSTGRES_BANNER_TLS_CERTIFICATE_PARSED_VALIDATION_LEVEL_UNKNOWN",

"P5432_POSTGRES_BANNER_TLS_CERTIFICATE_PARSED_VALIDATION_LEVEL_OV",

"P5432_POSTGRES_BANNER_TLS_CERTIFICATE_PARSED_SIGNATURE_VALID",

"P5432_POSTGRES_BANNER_TLS_CERTIFICATE_PARSED_SIGNATURE_SELF_SIGNED",

"P5432_POSTGRES_BANNER_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_AMAZON",

"P5432_POSTGRES_BANNER_TLS_CERTIFICATE_PARSED_EXTENSIONS_KEY_USAGE_KEY_ENCIPHERMENT",

"P5432_POSTGRES_BANNER_TLS_CERTIFICATE_PARSED_EXTENSIONS_KEY_USAGE_DIGITAL_SIGNATURE",

"P5432_POSTGRES_BANNER_TLS_CERTIFICATE_PARSED_EXTENSIONS_BASIC_CONSTRAINTS_IS_CA",

"P5432_POSTGRES_BANNER_SUPPORTED_VERSIONS_3_0",

"P5432_POSTGRES_BANNER_SUPPORTED_VERSIONS_2_0",

"P5432_POSTGRES_BANNER_SUPPORTED_VERSIONS_1_0",

"P5432_POSTGRES_BANNER_SUPPORTED",

"P5432_POSTGRES_BANNER_STARTUP_ERROR_SEVERITY_FATAL",

"P5432_POSTGRES_BANNER_STARTUP_ERROR_ROUTINE_PROCESSSTARTUPPACKET",

"P5432_POSTGRES_BANNER_STARTUP_ERROR_LINE",

"P5432_POSTGRES_BANNER_STARTUP_ERROR_FILE_POSTMASTER_C",

"P5432_POSTGRES_BANNER_PROTOCOL_ERROR_SEVERITY_FATAL",

"P5432_POSTGRES_BANNER_PROTOCOL_ERROR_ROUTINE_PROCESSSTARTUPPACKET",

"P5432_POSTGRES_BANNER_PROTOCOL_ERROR_LINE",

"P5432_POSTGRES_BANNER_PROTOCOL_ERROR_FILE_POSTMASTER_C",

"P5432_POSTGRES_BANNER_IS_SSL",

"P53_DNS_LOOKUP_SUPPORT",

"P53_DNS_LOOKUP_RESOLVES_CORRECTLY",

"P53_DNS_LOOKUP_OPEN_RESOLVER",

"P53_DNS_LOOKUP_ERRORS",

"P502_MODBUS_DEVICE_ID_SUPPORT",

"P502_MODBUS_DEVICE_ID_METADATA_DESCRIPTION_SCHNEIDER",

"P502_MODBUS_DEVICE_ID_MEI_RESPONSE_CONFORMITY_LEVEL",

"P502_MODBUS_DEVICE_ID_FUNCTION_CODE",

"P47808_BACNET_DEVICE_ID_VENDOR_ID",

"P47808_BACNET_DEVICE_ID_SUPPORT",

"P47808_BACNET_DEVICE_ID_RUNNING_WEBCTRL_SERVER",

"P47808_BACNET_DEVICE_ID_RUNNING_WEBCTRL_500",

"P47808_BACNET_DEVICE_ID_RUNNING_WEBCTRL_249999",

"P47808_BACNET_DEVICE_ID_RUNNING_WC17",

"P47808_BACNET_DEVICE_ID_RUNNING_V3_40" ,

"P47808_BACNET_DEVICE_ID_RUNNING_UNITED" ,

"P47808_BACNET_DEVICE_ID_RUNNING_TRIDIUM" ,

"P47808_BACNET_DEVICE_ID_RUNNING_TRIACTA" ,

"P47808_BACNET_DEVICE_ID_RUNNING_TRANE" ,

"P47808_BACNET_DEVICE_ID_RUNNING_TRACER" ,

"P47808_BACNET_DEVICE_ID_RUNNING_TAC" ,

"P47808_BACNET_DEVICE_ID_RUNNING_SIEMENS" ,

"P47808_BACNET_DEVICE_ID_RUNNING_SCHNEIDER" ,

"P47808_BACNET_DEVICE_ID_RUNNING_ROUTER" ,

"P47808_BACNET_DEVICE_ID_RUNNING_RELIABLE_CONTROLS" ,

"P47808_BACNET_DEVICE_ID_RUNNING_PV17" ,

"P47808_BACNET_DEVICE_ID_RUNNING_POWER_METER" ,

"P47808_BACNET_DEVICE_ID_RUNNING_OBVIUS" ,

"P47808_BACNET_DEVICE_ID_RUNNING_NIAGARAAX" ,

"P47808_BACNET_DEVICE_ID_RUNNING_NIAGARA4" ,

"P47808_BACNET_DEVICE_ID_RUNNING_NAE" ,

"P47808_BACNET_DEVICE_ID_RUNNING_MITSUBISHI" ,

"P47808_BACNET_DEVICE_ID_RUNNING_MCQUAY_INTERNATIONAL" ,

"P47808_BACNET_DEVICE_ID_RUNNING_MACH_PRO" ,

"P47808_BACNET_DEVICE_ID_RUNNING_LOYTEC_ELECTRONICS" ,

"P47808_BACNET_DEVICE_ID_RUNNING_LOCAL_BACNET_DEVICE_OBJECT" ,

"P47808_BACNET_DEVICE_ID_RUNNING_JOHNSON_CONTROLS" ,

"P47808_BACNET_DEVICE_ID_RUNNING_JCI" ,

"P47808_BACNET_DEVICE_ID_RUNNING_I_VU_STANDARD" ,

"P47808_BACNET_DEVICE_ID_RUNNING_I_VU_PLUS" ,

"P47808_BACNET_DEVICE_ID_RUNNING_I_VU" ,

"P47808_BACNET_DEVICE_ID_RUNNING_HONEYWELL" ,

"P47808_BACNET_DEVICE_ID_RUNNING_FIELDSERVER" ,

"P47808_BACNET_DEVICE_ID_RUNNING_ENS_1" ,

"P47808_BACNET_DEVICE_ID_RUNNING_EMERSON_NETWORK_POWER" ,

"P47808_BACNET_DEVICE_ID_RUNNING_EATON" ,

"P47808_BACNET_DEVICE_ID_RUNNING_DISTECH_CON" ,

"P47808_BACNET_DEVICE_ID_RUNNING_DEVICE" ,

"P47808_BACNET_DEVICE_ID_RUNNING_DEV2401" ,

"P47808_BACNET_DEVICE_ID_RUNNING_DESC" ,

"P47808_BACNET_DEVICE_ID_RUNNING_DELTA" ,

"P47808_BACNET_DEVICE_ID_RUNNING_DEFAULT_DESCRIPTION" ,

"P47808_BACNET_DEVICE_ID_RUNNING_CUSTOM_SOFTWARE_ENGINEERING" ,

"P47808_BACNET_DEVICE_ID_RUNNING_CONTROL" ,

"P47808_BACNET_DEVICE_ID_RUNNING_CARRIER" ,

"P47808_BACNET_DEVICE_ID_RUNNING_BASRT_B" ,

"P47808_BACNET_DEVICE_ID_RUNNING_BACNET_STACK_AT_SOURCEFORGE" ,

"P47808_BACNET_DEVICE_ID_RUNNING_BACNET_DEVICE" ,

"P47808_BACNET_DEVICE_ID_RUNNING_BACNETIP_TO_MSTP_ROUTER" ,

"P47808_BACNET_DEVICE_ID_RUNNING_AUTOMATED_LOGIC_CORPORATION" ,

"P47808_BACNET_DEVICE_ID_RUNNING_7_0" ,

"P47808_BACNET_DEVICE_ID_RUNNING_6_5" ,

"P47808_BACNET_DEVICE_ID_RUNNING_6_1" ,

"P47808_BACNET_DEVICE_ID_RUNNING_6_0" ,

"P47808_BACNET_DEVICE_ID_RUNNING_3_1" ,

"P47808_BACNET_DEVICE_ID_RUNNING_2_6" ,

"P47808_BACNET_DEVICE_ID_RUNNING_2_10" ,

"P47808_BACNET_DEVICE_ID_RUNNING_2_0" ,

"P47808_BACNET_DEVICE_ID_RUNNING_1_4" ,

"P47808_BACNET_DEVICE_ID_RUNNING_1_0" ,

"P47808_BACNET_DEVICE_ID_INSTANCE_NUMBER" ,

"P445_SMB_BANNER_SMBV1_SUPPORT" ,

"P443_HTTPS_TLS_VERSION_TLSV1_2" ,

"P443_HTTPS_TLS_VERSION_TLSV1_1" ,

"P443_HTTPS_TLS_VERSION_TLSV1_0" ,

"P443_HTTPS_TLS_VERSION_SSLV3" ,

"P443_HTTPS_TLS_VALIDATION_BROWSER_TRUSTED" ,

"P443_HTTPS_TLS_TLS_FIELD_PRESENT" ,

"P443_HTTPS_TLS_TLS_CERT_PAST_VALID_END_DATE" ,

"P443_HTTPS_TLS_SIGNATURE_VALID" ,

"P443_HTTPS_TLS_SESSION_TICKET_LIFETIME_HINT" ,

"P443_HTTPS_TLS_SESSION_TICKET_LENGTH" ,

"P443_HTTPS_TLS_SERVER_KEY_EXCHANGE_ECDH_PARAMS_CURVE_ID_ID",

"P443_HTTPS_TLS_SERVER_KEY_EXCHANGE_DH_PARAMS_PRIME_LENGTH",

"P443_HTTPS_TLS_SERVER_KEY_EXCHANGE_DH_PARAMS_GENERATOR_LENGTH",

"P443_HTTPS_TLS_OCSP_STAPLING",

"P443_HTTPS_TLS_CIPHER_SUITE_NAME_TLS_RSA_WITH_RC4_128_SHA",

"P443_HTTPS_TLS_CIPHER_SUITE_NAME_TLS_RSA_WITH_AES_256_CBC_SHA",

"P443_HTTPS_TLS_CIPHER_SUITE_NAME_TLS_RSA_WITH_AES_128_GCM_SHA256",

"P443_HTTPS_TLS_CIPHER_SUITE_NAME_TLS_RSA_WITH_AES_128_CBC_SHA",

"P443_HTTPS_TLS_CIPHER_SUITE_NAME_TLS_ECDHE_RSA_WITH_CHACHA20_POLY1305_SHA256",

"P443_HTTPS_TLS_CIPHER_SUITE_NAME_TLS_ECDHE_RSA_WITH_AES_256_CBC_SHA",

"P443_HTTPS_TLS_CIPHER_SUITE_NAME_TLS_ECDHE_RSA_WITH_AES_128_GCM_SHA256",

"P443_HTTPS_TLS_CIPHER_SUITE_NAME_TLS_ECDHE_ECDSA_WITH_AES_128_GCM_SHA256",

"P443_HTTPS_TLS_CIPHER_SUITE_NAME_TLS_DHE_RSA_WITH_AES_256_CBC_SHA",

"P443_HTTPS_TLS_CIPHER_SUITE_NAME_TLS_DHE_RSA_WITH_AES_128_GCM_SHA256",

"P443_HTTPS_TLS_CERTIFICATE_PARSED_VERSION",

"P443_HTTPS_TLS_CERTIFICATE_PARSED_VALIDITY_LENGTH",

"P443_HTTPS_TLS_CERTIFICATE_PARSED_VALIDATION_LEVEL_UNKNOWN",

"P443_HTTPS_TLS_CERTIFICATE_PARSED_VALIDATION_LEVEL_OV",

"P443_HTTPS_TLS_CERTIFICATE_PARSED_VALIDATION_LEVEL_EV",

"P443_HTTPS_TLS_CERTIFICATE_PARSED_VALIDATION_LEVEL_DV",

"P443_HTTPS_TLS_CERTIFICATE_PARSED_SIGNATURE_VALID",

"P443_HTTPS_TLS_CERTIFICATE_PARSED_SIGNATURE_SELF_SIGNED",

"P443_HTTPS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_WATCHGUARD",

"P443_HTTPS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_UNKNOWN",

"P443_HTTPS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_UBIQUITI",

"P443_HTTPS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_TRUSTASIA",

"P443_HTTPS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_STARFIELD",

"P443_HTTPS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_SONICWALL",

"P443_HTTPS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_SOMEORGANIZATION",

"P443_HTTPS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_RAPIDSSL",

"P443_HTTPS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_MOTOROLA",

"P443_HTTPS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_MINI_WEBSERVICE",

"P443_HTTPS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_MICROSOFT",

"P443_HTTPS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_LOCALHOST",

"P443_HTTPS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_LETS_ENCRYPT",

"P443_HTTPS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_HUAWEI",

"P443_HTTPS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_HOME_PL",

"P443_HTTPS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_GOOGLE",

"P443_HTTPS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_GODADDY",

"P443_HTTPS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_GLOBALSIGN",

"P443_HTTPS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_GEOTRUST",

"P443_HTTPS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_FORTINET",

"P443_HTTPS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_ENTRUST",

"P443_HTTPS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_DRAYTEK",

"P443_HTTPS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_DIGICERT",

"P443_HTTPS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_CPANEL",

"P443_HTTPS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_COMODO",

"P443_HTTPS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_CISCO",

"P443_HTTPS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_BMS",

"P443_HTTPS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_AMAZON",

"P443_HTTPS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_ALPHASSL",

"P443_HTTPS_TLS_CERTIFICATE_PARSED_EXTENSIONS_KEY_USAGE_KEY_ENCIPHERMENT",

"P443_HTTPS_TLS_CERTIFICATE_PARSED_EXTENSIONS_KEY_USAGE_KEY_AGREEMENT",

"P443_HTTPS_TLS_CERTIFICATE_PARSED_EXTENSIONS_KEY_USAGE_ENCIPHER_ONLY",

"P443_HTTPS_TLS_CERTIFICATE_PARSED_EXTENSIONS_KEY_USAGE_DIGITAL_SIGNATURE",

"P443_HTTPS_TLS_CERTIFICATE_PARSED_EXTENSIONS_KEY_USAGE_DECIPHER_ONLY",

"P443_HTTPS_TLS_CERTIFICATE_PARSED_EXTENSIONS_KEY_USAGE_DATA_ENCIPHERMENT",

"P443_HTTPS_TLS_CERTIFICATE_PARSED_EXTENSIONS_KEY_USAGE_CRL_SIGN",

"P443_HTTPS_TLS_CERTIFICATE_PARSED_EXTENSIONS_KEY_USAGE_CONTENT_COMMITMENT",

"P443_HTTPS_TLS_CERTIFICATE_PARSED_EXTENSIONS_KEY_USAGE_CERTIFICATE_SIGN",

"P443_HTTPS_TLS_CERTIFICATE_PARSED_EXTENSIONS_BASIC_CONSTRAINTS_MAX_PATH_LEN",

"P443_HTTPS_TLS_CERTIFICATE_PARSED_EXTENSIONS_BASIC_CONSTRAINTS_IS_CA",

"P443_HTTPS_SSL_3_SUPPORT",

"P443_HTTPS_SSL_2_TLS_CERT_PAST_VALID_END_DATE",

"P443_HTTPS_SSL_2_SUPPORT",

"P443_HTTPS_SSL_2_SSL_2_SUPPORT",

"P443_HTTPS_SSL_2_EXTRA_CLEAR",

"P443_HTTPS_SSL_2_EXPORT",

"P443_HTTPS_SSL_2_CERTIFICATE_PARSED_VERSION",

"P443_HTTPS_SSL_2_CERTIFICATE_PARSED_SIGNATURE_SELF_SIGNED",

"P443_HTTPS_SSL_2_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_STARFIELD",

"P443_HTTPS_SSL_2_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_SOMEORGANIZATION",

"P443_HTTPS_SSL_2_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_RAPIDSSL",

"P443_HTTPS_SSL_2_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_LOCALHOST",

"P443_HTTPS_SSL_2_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_GOOGLE",

"P443_HTTPS_SSL_2_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_GODADDY",

"P443_HTTPS_SSL_2_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_GLOBALSIGN",

"P443_HTTPS_SSL_2_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_GEOTRUST",

"P443_HTTPS_SSL_2_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_ENTRUST",

"P443_HTTPS_SSL_2_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_DIGICERT",

"P443_HTTPS_SSL_2_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_COMODO",

"P443_HTTPS_SSL_2_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_CISCO",

"P443_HTTPS_SSL_2_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_ALPHASSL",

"P443_HTTPS_SSL_2_CERTIFICATE_PARSED_EXTENSIONS_KEY_USAGE_KEY_ENCIPHERMENT",

"P443_HTTPS_SSL_2_CERTIFICATE_PARSED_EXTENSIONS_KEY_USAGE_KEY_AGREEMENT",

"P443_HTTPS_SSL_2_CERTIFICATE_PARSED_EXTENSIONS_KEY_USAGE_DIGITAL_SIGNATURE",

"P443_HTTPS_SSL_2_CERTIFICATE_PARSED_EXTENSIONS_KEY_USAGE_DATA_ENCIPHERMENT",

"P443_HTTPS_SSL_2_CERTIFICATE_PARSED_EXTENSIONS_KEY_USAGE_CONTENT_COMMITMENT",

"P443_HTTPS_SSL_2_CERTIFICATE_PARSED_EXTENSIONS_KEY_USAGE_CERTIFICATE_SIGN",

"P443_HTTPS_SSL_2_CERTIFICATE_PARSED_EXTENSIONS_BASIC_CONSTRAINTS_MAX_PATH_LEN",

"P443_HTTPS_SSL_2_CERTIFICATE_PARSED_EXTENSIONS_BASIC_CONSTRAINTS_IS_CA",

"P443_HTTPS_RSA_EXPORT_SUPPORT",

"P443_HTTPS_RSA_EXPORT_RSA_PARAMS_LENGTH",

"P443_HTTPS_HEARTBLEED_VULNERABLE",

"P443_HTTPS_HEARTBEAT_ENABLED",

"P443_HTTPS_DHE_SUPPORT",

"P443_HTTPS_DHE_EXPORT_SUPPORT",

"P443_HTTPS_DHE_EXPORT_DH_PARAMS_PRIME_LENGTH",

"P443_HTTPS_DHE_EXPORT_DH_PARAMS_GENERATOR_LENGTH",

"P443_HTTPS_DHE_DH_PARAMS_PRIME_LENGTH",

"P443_HTTPS_DHE_DH_PARAMS_GENERATOR_LENGTH",

"P3306_MYSQL_BANNER_TLS_VERSION_TLSV1_2",

"P3306_MYSQL_BANNER_TLS_VERSION_TLSV1_1",

"P3306_MYSQL_BANNER_TLS_VERSION_TLSV1_0",

"P3306_MYSQL_BANNER_TLS_VALIDATION_BROWSER_TRUSTED",

"P3306_MYSQL_BANNER_TLS_TLS_FIELD_PRESENT",

"P3306_MYSQL_BANNER_TLS_SIGNATURE_VALID",

"P3306_MYSQL_BANNER_TLS_SERVER_KEY_EXCHANGE_ECDH_PARAMS_CURVE_ID_ID",

"P3306_MYSQL_BANNER_TLS_CIPHER_SUITE_NAME_TLS_RSA_WITH_RC4_128_SHA",

"P3306_MYSQL_BANNER_TLS_CIPHER_SUITE_NAME_TLS_RSA_WITH_AES_256_CBC_SHA",

"P3306_MYSQL_BANNER_TLS_CIPHER_SUITE_NAME_TLS_RSA_WITH_AES_128_CBC_SHA",

"P3306_MYSQL_BANNER_TLS_CIPHER_SUITE_NAME_TLS_ECDHE_RSA_WITH_AES_128_GCM_SHA256",

"P3306_MYSQL_BANNER_TLS_CERTIFICATE_PARSED_VERSION",

"P3306_MYSQL_BANNER_TLS_CERTIFICATE_PARSED_VALIDITY_LENGTH",

"P3306_MYSQL_BANNER_TLS_CERTIFICATE_PARSED_VALIDATION_LEVEL_UNKNOWN",

"P3306_MYSQL_BANNER_TLS_CERTIFICATE_PARSED_VALIDATION_LEVEL_DV",

"P3306_MYSQL_BANNER_TLS_CERTIFICATE_PARSED_SIGNATURE_VALID",

"P3306_MYSQL_BANNER_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_LETS_ENCRYPT",

"P3306_MYSQL_BANNER_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_COMODO",

"P3306_MYSQL_BANNER_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_AMAZON",

"P3306_MYSQL_BANNER_TLS_CERTIFICATE_PARSED_EXTENSIONS_KEY_USAGE_KEY_ENCIPHERMENT",

"P3306_MYSQL_BANNER_TLS_CERTIFICATE_PARSED_EXTENSIONS_KEY_USAGE_DIGITAL_SIGNATURE",

"P3306_MYSQL_BANNER_TLS_CERTIFICATE_PARSED_EXTENSIONS_BASIC_CONSTRAINTS_IS_CA",

"P3306_MYSQL_BANNER_SUPPORT",

"P3306_MYSQL_BANNER_STATUS_FLAGS_SERVER_STATUS_AUTOCOMMIT",

"P3306_MYSQL_BANNER_SERVER_VERSION_8_0",

"P3306_MYSQL_BANNER_SERVER_VERSION_6_0",

"P3306_MYSQL_BANNER_SERVER_VERSION_5_7",

"P3306_MYSQL_BANNER_SERVER_VERSION_5_6",

"P3306_MYSQL_BANNER_SERVER_VERSION_5_5",

"P3306_MYSQL_BANNER_SERVER_VERSION_5_1",

"P3306_MYSQL_BANNER_SERVER_VERSION_5_0",

"P3306_MYSQL_BANNER_SERVER_VERSION_4_2",

"P3306_MYSQL_BANNER_SERVER_VERSION_4_1",

"P3306_MYSQL_BANNER_SERVER_VERSION_4_0",

"P3306_MYSQL_BANNER_SERVER_VERSION_3_2",

"P3306_MYSQL_BANNER_SERVER_VERSION_3_1" ,

"P3306_MYSQL_BANNER_PROTOCOL_VERSION" ,

"P3306_MYSQL_BANNER_ERROR_MESSAGE_NOT_ALLOWED_TO_CONNECT_TO_THIS_MYSQL_SERVER" ,

"P3306_MYSQL_BANNER_ERROR_MESSAGE_NOT_ALLOWED_TO_CONNECT_TO_THIS_MARIADB_SERVER" ,

"P3306_MYSQL_BANNER_ERROR_MESSAGE_BLOCKED_BECAUSE_OF_MANY_CONNECTION_ERRORS" ,

"P3306_MYSQL_BANNER_ERROR_ID_ER_HOST_NOT_PRIVILEGED" ,

"P3306_MYSQL_BANNER_ERROR_ID_ER_HOST_IS_BLOCKED" ,

"P3306_MYSQL_BANNER_ERROR_CODE" ,

"P3306_MYSQL_BANNER_CAPABILITY_FLAGS_CLIENT_TRANSACTIONS" ,

"P3306_MYSQL_BANNER_CAPABILITY_FLAGS_CLIENT_SSL" ,

"P3306_MYSQL_BANNER_CAPABILITY_FLAGS_CLIENT_SESSION_TRACK" ,

"P3306_MYSQL_BANNER_CAPABILITY_FLAGS_CLIENT_SECURE_CONNECTION" ,

"P3306_MYSQL_BANNER_CAPABILITY_FLAGS_CLIENT_RESERVED" ,

"P3306_MYSQL_BANNER_CAPABILITY_FLAGS_CLIENT_PS_MULTI_RESULTS" ,

"P3306_MYSQL_BANNER_CAPABILITY_FLAGS_CLIENT_PROTOCOL_41" ,

"P3306_MYSQL_BANNER_CAPABILITY_FLAGS_CLIENT_PLUGIN_AUTH_LEN_ENC_CLIENT_DATA" ,

"P3306_MYSQL_BANNER_CAPABILITY_FLAGS_CLIENT_PLUGIN_AUTH" ,

"P3306_MYSQL_BANNER_CAPABILITY_FLAGS_CLIENT_ODBC" ,

"P3306_MYSQL_BANNER_CAPABILITY_FLAGS_CLIENT_NO_SCHEMA" ,

"P3306_MYSQL_BANNER_CAPABILITY_FLAGS_CLIENT_MULTI_STATEMENTS" ,

"P3306_MYSQL_BANNER_CAPABILITY_FLAGS_CLIENT_MULTI_RESULTS" ,

"P3306_MYSQL_BANNER_CAPABILITY_FLAGS_CLIENT_LONG_PASSWORD" ,

"P3306_MYSQL_BANNER_CAPABILITY_FLAGS_CLIENT_LONG_FLAG" ,

"P3306_MYSQL_BANNER_CAPABILITY_FLAGS_CLIENT_LOCAL_FILES" ,

"P3306_MYSQL_BANNER_CAPABILITY_FLAGS_CLIENT_INTERACTIVE" ,

"P3306_MYSQL_BANNER_CAPABILITY_FLAGS_CLIENT_IGNORE_SPACE" ,

"P3306_MYSQL_BANNER_CAPABILITY_FLAGS_CLIENT_IGNORE_SIGPIPE" ,

"P3306_MYSQL_BANNER_CAPABILITY_FLAGS_CLIENT_FOUND_ROWS" ,

"P3306_MYSQL_BANNER_CAPABILITY_FLAGS_CLIENT_DEPRECATED_EOF" ,

"P3306_MYSQL_BANNER_CAPABILITY_FLAGS_CLIENT_CONNECT_WITH_DB" ,

"P3306_MYSQL_BANNER_CAPABILITY_FLAGS_CLIENT_CONNECT_ATTRS" ,

"P3306_MYSQL_BANNER_CAPABILITY_FLAGS_CLIENT_COMPRESS" ,

"P3306_MYSQL_BANNER_CAPABILITY_FLAGS_CLIENT_CAN_HANDLE_EXPIRED_PASSWORDS" ,

"P25_SMTP_TLS_TLS_FIELD_PRESENT" ,

"P25_SMTP_TLS_TLS_CERT_PAST_VALID_END_DATE",

"P25_SMTP_STARTTLS_TLS_VERSION_TLSV1_2",

"P25_SMTP_STARTTLS_TLS_VERSION_TLSV1_1",

"P25_SMTP_STARTTLS_TLS_VERSION_TLSV1_0",

"P25_SMTP_STARTTLS_TLS_VALIDATION_BROWSER_TRUSTED",

"P25_SMTP_STARTTLS_TLS_SIGNATURE_VALID",

"P25_SMTP_STARTTLS_TLS_SERVER_KEY_EXCHANGE_ECDH_PARAMS_CURVE_ID_ID",

"P25_SMTP_STARTTLS_TLS_OCSP_STAPLING",

"P25_SMTP_STARTTLS_TLS_CIPHER_SUITE_NAME_TLS_RSA_WITH_RC4_128_SHA",

"P25_SMTP_STARTTLS_TLS_CIPHER_SUITE_NAME_TLS_RSA_WITH_AES_256_CBC_SHA",

"P25_SMTP_STARTTLS_TLS_CIPHER_SUITE_NAME_TLS_RSA_WITH_AES_128_CBC_SHA",

"P25_SMTP_STARTTLS_TLS_CIPHER_SUITE_NAME_TLS_ECDHE_RSA_WITH_AES_256_CBC_SHA",

"P25_SMTP_STARTTLS_TLS_CIPHER_SUITE_NAME_TLS_ECDHE_RSA_WITH_AES_128_GCM_SHA256",

"P25_SMTP_STARTTLS_TLS_CIPHER_SUITE_NAME_TLS_ECDHE_ECDSA_WITH_AES_128_GCM_SHA256",

"P25_SMTP_STARTTLS_TLS_CERTIFICATE_PARSED_VERSION",

"P25_SMTP_STARTTLS_TLS_CERTIFICATE_PARSED_VALIDITY_LENGTH",

"P25_SMTP_STARTTLS_TLS_CERTIFICATE_PARSED_VALIDATION_LEVEL_UNKNOWN",

"P25_SMTP_STARTTLS_TLS_CERTIFICATE_PARSED_VALIDATION_LEVEL_OV",

"P25_SMTP_STARTTLS_TLS_CERTIFICATE_PARSED_VALIDATION_LEVEL_EV",

"P25_SMTP_STARTTLS_TLS_CERTIFICATE_PARSED_VALIDATION_LEVEL_DV",

"P25_SMTP_STARTTLS_TLS_CERTIFICATE_PARSED_SIGNATURE_VALID",

"P25_SMTP_STARTTLS_TLS_CERTIFICATE_PARSED_SIGNATURE_SELF_SIGNED",

"P25_SMTP_STARTTLS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_WATCHGUARD",

"P25_SMTP_STARTTLS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_UNKNOWN",

"P25_SMTP_STARTTLS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_STARFIELD",

"P25_SMTP_STARTTLS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_SONICWALL",

"P25_SMTP_STARTTLS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_SOMEORGANIZATION",

"P25_SMTP_STARTTLS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_RAPIDSSL",

"P25_SMTP_STARTTLS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_LOCALHOST",

"P25_SMTP_STARTTLS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_LETS_ENCRYPT",

"P25_SMTP_STARTTLS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_HOME_PL",

"P25_SMTP_STARTTLS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_GODADDY",

"P25_SMTP_STARTTLS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_GLOBALSIGN",

"P25_SMTP_STARTTLS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_GEOTRUST",

"P25_SMTP_STARTTLS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_FORTINET",

"P25_SMTP_STARTTLS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_ENTRUST",

"P25_SMTP_STARTTLS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_DIGICERT",

"P25_SMTP_STARTTLS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_CPANEL",

"P25_SMTP_STARTTLS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_COMODO",

"P25_SMTP_STARTTLS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_CISCO",

"P25_SMTP_STARTTLS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_ALPHASSL",

"P25_SMTP_STARTTLS_TLS_CERTIFICATE_PARSED_EXTENSIONS_KEY_USAGE_KEY_ENCIPHERMENT",

"P25_SMTP_STARTTLS_TLS_CERTIFICATE_PARSED_EXTENSIONS_KEY_USAGE_KEY_AGREEMENT",

"P25_SMTP_STARTTLS_TLS_CERTIFICATE_PARSED_EXTENSIONS_KEY_USAGE_DIGITAL_SIGNATURE",

"P25_SMTP_STARTTLS_TLS_CERTIFICATE_PARSED_EXTENSIONS_KEY_USAGE_DATA_ENCIPHERMENT",

"P25_SMTP_STARTTLS_TLS_CERTIFICATE_PARSED_EXTENSIONS_KEY_USAGE_CONTENT_COMMITMENT",

"P25_SMTP_STARTTLS_TLS_CERTIFICATE_PARSED_EXTENSIONS_KEY_USAGE_CERTIFICATE_SIGN",

"P25_SMTP_STARTTLS_TLS_CERTIFICATE_PARSED_EXTENSIONS_BASIC_CONSTRAINTS_IS_CA",

"P25_SMTP_STARTTLS_STARTTLS_UNKNOWN",

"P25_SMTP_STARTTLS_RUNNING_ZMAILER",

"P25_SMTP_STARTTLS_RUNNING_ZIMBRA",

"P25_SMTP_STARTTLS_RUNNING_XCHANGE",

"P25_SMTP_STARTTLS_RUNNING_UNKNOWN",

"P25_SMTP_STARTTLS_RUNNING_SUN",

"P25_SMTP_STARTTLS_RUNNING_SMAIL",

"P25_SMTP_STARTTLS_RUNNING_SENDMAIL",

"P25_SMTP_STARTTLS_RUNNING_QMAIL",

"P25_SMTP_STARTTLS_RUNNING_POSTFIX",

"P25_SMTP_STARTTLS_RUNNING_ORACLE",

"P25_SMTP_STARTTLS_RUNNING_OPENSMTPD",

"P25_SMTP_STARTTLS_RUNNING_NOVELL",

"P25_SMTP_STARTTLS_RUNNING_NETMAIL",

"P25_SMTP_STARTTLS_RUNNING_MICROSOFT",

"P25_SMTP_STARTTLS_RUNNING_MERCURY",

"P25_SMTP_STARTTLS_RUNNING_MDAEMON",

"P25_SMTP_STARTTLS_RUNNING_MAILSITE",

"P25_SMTP_STARTTLS_RUNNING_MAILENABLE",

"P25_SMTP_STARTTLS_RUNNING_KERIO_CONNECT",

"P25_SMTP_STARTTLS_RUNNING_KERIO",

"P25_SMTP_STARTTLS_RUNNING_IRONPORT",

"P25_SMTP_STARTTLS_RUNNING_IMAP",

"P25_SMTP_STARTTLS_RUNNING_ICEWARP",

"P25_SMTP_STARTTLS_RUNNING_IBM",

"P25_SMTP_STARTTLS_RUNNING_HMAIL",

"P25_SMTP_STARTTLS_RUNNING_HARAKA",

"P25_SMTP_STARTTLS_RUNNING_GROUPWISE",

"P25_SMTP_STARTTLS_RUNNING_GOOGLE",

"P25_SMTP_STARTTLS_RUNNING_FIRSTCLASS",

"P25_SMTP_STARTTLS_RUNNING_EXIM",

"P25_SMTP_STARTTLS_RUNNING_EXCHANGE_SERVER",

"P25_SMTP_STARTTLS_RUNNING_CYRUS",

"P25_SMTP_STARTTLS_RUNNING_AXIGEN",

"P25_SMTP_STARTTLS_CONN_SUCCESS",

"P25_SMTP_SSL_2_TLS_CERT_PAST_VALID_END_DATE",

"P25_SMTP_SSL_2_SSL_2_SUPPORT",

"P25_SMTP_SSL_2_CERTIFICATE_PARSED_VERSION",

"P25_SMTP_SSL_2_CERTIFICATE_PARSED_SIGNATURE_SELF_SIGNED",

"P25_SMTP_SSL_2_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_RAPIDSSL",

"P25_SMTP_SSL_2_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_GODADDY",

"P25_SMTP_SSL_2_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_GLOBALSIGN",

"P25_SMTP_SSL_2_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_GEOTRUST",

"P25_SMTP_SSL_2_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_COMODO",

"P25_SMTP_SSL_2_CERTIFICATE_PARSED_EXTENSIONS_KEY_USAGE_KEY_ENCIPHERMENT",

"P25_SMTP_SSL_2_CERTIFICATE_PARSED_EXTENSIONS_KEY_USAGE_KEY_AGREEMENT",

"P25_SMTP_SSL_2_CERTIFICATE_PARSED_EXTENSIONS_KEY_USAGE_DIGITAL_SIGNATURE",

"P25_SMTP_SSL_2_CERTIFICATE_PARSED_EXTENSIONS_KEY_USAGE_DATA_ENCIPHERMENT",

"P25_SMTP_SSL_2_CERTIFICATE_PARSED_EXTENSIONS_KEY_USAGE_CERTIFICATE_SIGN",

"P23_TELNET_BANNER_SUPPORT",

"P23_TELNET_BANNER_CONN_SUCCESS",

"P2323_TELNET_BANNER_SUPPORT",

"P22_SSH_V2_SUPPORT_SERVER_TO_CLIENT_MAC_HMAC_SHA2_256",

"P22_SSH_V2_SUPPORT_SERVER_TO_CLIENT_MAC_HMAC_SHA1_96",

"P22_SSH_V2_SUPPORT_SERVER_TO_CLIENT_MAC_HMAC_SHA1" ,

"P22_SSH_V2_SUPPORT_SERVER_TO_CLIENT_MAC_HMAC_MD5_96" ,

"P22_SSH_V2_SUPPORT_SERVER_TO_CLIENT_MAC_HMAC_MD5" ,

"P22_SSH_V2_SUPPORT_SERVER_TO_CLIENT_COMPRESSION_ZLIB" ,

"P22_SSH_V2_SUPPORT_SERVER_TO_CLIENT_COMPRESSION_NONE" ,

"P22_SSH_V2_SUPPORT_SERVER_TO_CLIENT_CIPHER_ARCFOUR256" ,

"P22_SSH_V2_SUPPORT_SERVER_TO_CLIENT_CIPHER_ARCFOUR128" ,

"P22_SSH_V2_SUPPORT_SERVER_TO_CLIENT_CIPHER_ARCFOUR" ,

"P22_SSH_V2_SUPPORT_SERVER_TO_CLIENT_CIPHER_AES256_CTR" ,

"P22_SSH_V2_SUPPORT_SERVER_TO_CLIENT_CIPHER_AES192_CTR" ,

"P22_SSH_V2_SUPPORT_SERVER_TO_CLIENT_CIPHER_AES128_GCM_OPENSSH_COM" ,

"P22_SSH_V2_SUPPORT_SERVER_TO_CLIENT_CIPHER_AES128_CTR" ,

"P22_SSH_V2_SUPPORT_SERVER_TO_CLIENT_CIPHER_AES128_CBC" ,

"P22_SSH_V2_SUPPORT_SERVER_TO_CLIENT_CIPHER_3DES_CBC" ,

"P22_SSH_V2_SUPPORT_KEX_ALGORITHM_ECDH_SHA2_NISTP521" ,

"P22_SSH_V2_SUPPORT_KEX_ALGORITHM_ECDH_SHA2_NISTP384" ,

"P22_SSH_V2_SUPPORT_KEX_ALGORITHM_ECDH_SHA2_NISTP256" ,

"P22_SSH_V2_SUPPORT_KEX_ALGORITHM_DIFFIE_HELLMAN_GROUP_EXCHANGE_SHA256" ,

"P22_SSH_V2_SUPPORT_KEX_ALGORITHM_DIFFIE_HELLMAN_GROUP_EXCHANGE_SHA1" ,

"P22_SSH_V2_SUPPORT_KEX_ALGORITHM_DIFFIE_HELLMAN_GROUP1_SHA1" ,

"P22_SSH_V2_SUPPORT_KEX_ALGORITHM_DIFFIE_HELLMAN_GROUP14_SHA1" ,

"P22_SSH_V2_SUPPORT_KEX_ALGORITHM_CURVE25519_SHA256_LIBSSH_ORG" ,

"P22_SSH_V2_SUPPORT_HOST_KEY_ALGORITHM_SSH_RSA_CERT_V01_OPENSSH_COM" ,

"P22_SSH_V2_SUPPORT_HOST_KEY_ALGORITHM_SSH_RSA" ,

"P22_SSH_V2_SUPPORT_HOST_KEY_ALGORITHM_SSH_ED25519_CERT_V01_OPENSSH_COM" ,

"P22_SSH_V2_SUPPORT_HOST_KEY_ALGORITHM_SSH_ED25519" ,

"P22_SSH_V2_SUPPORT_HOST_KEY_ALGORITHM_SSH_DSS" ,

"P22_SSH_V2_SUPPORT_HOST_KEY_ALGORITHM_ECDSA_SHA2_NISTP521" ,

"P22_SSH_V2_SUPPORT_HOST_KEY_ALGORITHM_ECDSA_SHA2_NISTP384" ,

"P22_SSH_V2_SUPPORT_HOST_KEY_ALGORITHM_ECDSA_SHA2_NISTP256" ,

"P22_SSH_V2_SUPPORT_CLIENT_TO_SERVER_MAC_HMAC_SHA2_256" ,

"P22_SSH_V2_SUPPORT_CLIENT_TO_SERVER_MAC_HMAC_SHA1_96" ,

"P22_SSH_V2_SUPPORT_CLIENT_TO_SERVER_MAC_HMAC_SHA1" ,

"P22_SSH_V2_SUPPORT_CLIENT_TO_SERVER_MAC_HMAC_MD5_96" ,

"P22_SSH_V2_SUPPORT_CLIENT_TO_SERVER_MAC_HMAC_MD5",

"P22_SSH_V2_SUPPORT_CLIENT_TO_SERVER_COMPRESSION_ZLIB",

"P22_SSH_V2_SUPPORT_CLIENT_TO_SERVER_COMPRESSION_NONE",

"P22_SSH_V2_SUPPORT_CLIENT_TO_SERVER_CIPHER_ARCFOUR256",

"P22_SSH_V2_SUPPORT_CLIENT_TO_SERVER_CIPHER_ARCFOUR128",

"P22_SSH_V2_SUPPORT_CLIENT_TO_SERVER_CIPHER_ARCFOUR",

"P22_SSH_V2_SUPPORT_CLIENT_TO_SERVER_CIPHER_AES256_CTR",

"P22_SSH_V2_SUPPORT_CLIENT_TO_SERVER_CIPHER_AES192_CTR",

"P22_SSH_V2_SUPPORT_CLIENT_TO_SERVER_CIPHER_AES128_GCM_OPENSSH_COM",

"P22_SSH_V2_SUPPORT_CLIENT_TO_SERVER_CIPHER_AES128_CTR",

"P22_SSH_V2_SUPPORT_CLIENT_TO_SERVER_CIPHER_AES128_CBC",

"P22_SSH_V2_SUPPORT_CLIENT_TO_SERVER_CIPHER_3DES_CBC",

"P22_SSH_V2_SERVER_HOST_KEY_KEY_ALGORITHM_SSH_RSA_CERT_V01_OPENSSH_COM",

"P22_SSH_V2_SERVER_HOST_KEY_KEY_ALGORITHM_SSH_RSA",

"P22_SSH_V2_SERVER_HOST_KEY_KEY_ALGORITHM_SSH_ED25519",

"P22_SSH_V2_SERVER_HOST_KEY_KEY_ALGORITHM_SSH_DSS",

"P22_SSH_V2_SERVER_HOST_KEY_KEY_ALGORITHM_ECDSA_SHA2_NISTP521",

"P22_SSH_V2_SERVER_HOST_KEY_KEY_ALGORITHM_ECDSA_SHA2_NISTP384",

"P22_SSH_V2_SERVER_HOST_KEY_KEY_ALGORITHM_ECDSA_SHA2_NISTP256",

"P22_SSH_V2_SERVER_HOST_KEY_CERTKEY_PUBLIC_KEY_TYPE_NAME_HOST",

"P22_SSH_V2_SELECTED_SERVER_TO_CLIENT_MAC_HMAC_SHA2_256",

"P22_SSH_V2_SELECTED_SERVER_TO_CLIENT_MAC_HMAC_SHA1",

"P22_SSH_V2_SELECTED_SERVER_TO_CLIENT_COMPRESSION_NONE",

"P22_SSH_V2_SELECTED_SERVER_TO_CLIENT_CIPHER_ARCFOUR256",

"P22_SSH_V2_SELECTED_SERVER_TO_CLIENT_CIPHER_ARCFOUR128",

"P22_SSH_V2_SELECTED_SERVER_TO_CLIENT_CIPHER_ARCFOUR",

"P22_SSH_V2_SELECTED_SERVER_TO_CLIENT_CIPHER_AES256_CTR",

"P22_SSH_V2_SELECTED_SERVER_TO_CLIENT_CIPHER_AES192_CTR",

"P22_SSH_V2_SELECTED_SERVER_TO_CLIENT_CIPHER_AES128_GCM_OPENSSH_COM",

"P22_SSH_V2_SELECTED_SERVER_TO_CLIENT_CIPHER_AES128_CTR",

"P22_SSH_V2_SELECTED_KEX_ALGORITHM_ECDH_SHA2_NISTP256",

"P22_SSH_V2_SELECTED_KEX_ALGORITHM_DIFFIE_HELLMAN_GROUP1_SHA1",

"P22_SSH_V2_SELECTED_KEX_ALGORITHM_DIFFIE_HELLMAN_GROUP14_SHA1",

"P22_SSH_V2_SELECTED_KEX_ALGORITHM_CURVE25519_SHA256_LIBSSH_ORG",

"P22_SSH_V2_SELECTED_HOST_KEY_ALGORITHM_SSH_RSA_CERT_V01_OPENSSH_COM",

"P22_SSH_V2_SELECTED_HOST_KEY_ALGORITHM_SSH_RSA",

"P22_SSH_V2_SELECTED_HOST_KEY_ALGORITHM_SSH_ED25519",

"P22_SSH_V2_SELECTED_HOST_KEY_ALGORITHM_SSH_DSS",

"P22_SSH_V2_SELECTED_HOST_KEY_ALGORITHM_ECDSA_SHA2_NISTP521",

"P22_SSH_V2_SELECTED_HOST_KEY_ALGORITHM_ECDSA_SHA2_NISTP384",

"P22_SSH_V2_SELECTED_HOST_KEY_ALGORITHM_ECDSA_SHA2_NISTP256",

"P22_SSH_V2_SELECTED_CLIENT_TO_SERVER_MAC_HMAC_SHA2_256",

"P22_SSH_V2_SELECTED_CLIENT_TO_SERVER_MAC_HMAC_SHA1",

"P22_SSH_V2_SELECTED_CLIENT_TO_SERVER_COMPRESSION_NONE",

"P22_SSH_V2_SELECTED_CLIENT_TO_SERVER_CIPHER_ARCFOUR256",

"P22_SSH_V2_SELECTED_CLIENT_TO_SERVER_CIPHER_ARCFOUR128",

"P22_SSH_V2_SELECTED_CLIENT_TO_SERVER_CIPHER_ARCFOUR",

"P22_SSH_V2_SELECTED_CLIENT_TO_SERVER_CIPHER_AES256_CTR",

"P22_SSH_V2_SELECTED_CLIENT_TO_SERVER_CIPHER_AES192_CTR",

"P22_SSH_V2_SELECTED_CLIENT_TO_SERVER_CIPHER_AES128_GCM_OPENSSH_COM",

"P22_SSH_V2_SELECTED_CLIENT_TO_SERVER_CIPHER_AES128_CTR",

"P22_SSH_V2_RUNNING_ZYXEL",

"P22_SSH_V2_RUNNING_XLIGHTFTPD",

"P22_SSH_V2_RUNNING_WS_FTP_SSH",

"P22_SSH_V2_RUNNING_WRI",

"P22_SSH_V2_RUNNING_UNKNOWN",

"P22_SSH_V2_RUNNING_SYSAXSSH",

"P22_SSH_V2_RUNNING_SYNCPLIFY",

"P22_SSH_V2_RUNNING_SUN",

"P22_SSH_V2_RUNNING_SRTSSHSERVER",

"P22_SSH_V2_RUNNING_SERV_U",

"P22_SSH_V2_RUNNING_ROUTEROS",

"P22_SSH_V2_RUNNING_ROSSSH",

"P22_SSH_V2_RUNNING_ROMSSHELL",

"P22_SSH_V2_RUNNING_REBEXSSH",

"P22_SSH_V2_RUNNING_OPENSSH",

"P22_SSH_V2_RUNNING_MPSSH",

"P22_SSH_V2_RUNNING_MOD_SFTP",

"P22_SSH_V2_RUNNING_MAVERICK",

"P22_SSH_V2_RUNNING_LIBSSH",

"P22_SSH_V2_RUNNING_LANCOM",

"P22_SSH_V2_RUNNING_IPSSH",

"P22_SSH_V2_RUNNING_HP",

"P22_SSH_V2_RUNNING_DROPBEAR",

"P22_SSH_V2_RUNNING_CRUSH",

"P22_SSH_V2_RUNNING_COMPLETEFTP",

"P22_SSH_V2_RUNNING_CISCO",

"P22_SSH_V2_RUNNING_CERBERUS",

"P22_SSH_V2_RUNNING_ADTRAN",

"P22_SSH_V2_KEY_EXCHANGE_ECDH_PARAMS_SERVER_PUBLIC_Y_LENGTH",

"P22_SSH_V2_KEY_EXCHANGE_ECDH_PARAMS_SERVER_PUBLIC_X_LENGTH",

"P22_SSH_V2_KEY_EXCHANGE_DH_PARAMS_PRIME_LENGTH",

"P22_SSH_V2_KEY_EXCHANGE_DH_PARAMS_GENERATOR_LENGTH",

"P22_SSH_V2_BANNER_VERSION_2_0",

"P22_SSH_V2_BANNER_VERSION_1_9",

"P22_SSH_V2_BANNER_VERSION_1_5",

"P22_SSH_RSA_PUB_KEY_LENGTH",

"P22_SSH_ECDSA_LENGTH",

"P22_SSH_CERT_KEY_VALID_LENGTH",

"P22_SSH_BANNER_SERVER_KEY_EXCHANGE_SERVER_TO_CLIENT_MAC_HMAC_SHA2_256",

"P22_SSH_BANNER_SERVER_KEY_EXCHANGE_SERVER_TO_CLIENT_MAC_HMAC_SHA1_96",

"P22_SSH_BANNER_SERVER_KEY_EXCHANGE_SERVER_TO_CLIENT_MAC_HMAC_SHA1",

"P22_SSH_BANNER_SERVER_KEY_EXCHANGE_SERVER_TO_CLIENT_MAC_HMAC_MD5_96",

"P22_SSH_BANNER_SERVER_KEY_EXCHANGE_SERVER_TO_CLIENT_MAC_HMAC_MD5",

"P22_SSH_BANNER_SERVER_KEY_EXCHANGE_SERVER_TO_CLIENT_COMPRESSION_ZLIB",

"P22_SSH_BANNER_SERVER_KEY_EXCHANGE_SERVER_TO_CLIENT_COMPRESSION_NONE",

"P22_SSH_BANNER_SERVER_KEY_EXCHANGE_SERVER_TO_CLIENT_CIPHER_ARCFOUR256",

"P22_SSH_BANNER_SERVER_KEY_EXCHANGE_SERVER_TO_CLIENT_CIPHER_ARCFOUR128",

"P22_SSH_BANNER_SERVER_KEY_EXCHANGE_SERVER_TO_CLIENT_CIPHER_ARCFOUR",

"P22_SSH_BANNER_SERVER_KEY_EXCHANGE_SERVER_TO_CLIENT_CIPHER_AES256_CTR",

"P22_SSH_BANNER_SERVER_KEY_EXCHANGE_SERVER_TO_CLIENT_CIPHER_AES192_CTR",

"P22_SSH_BANNER_SERVER_KEY_EXCHANGE_SERVER_TO_CLIENT_CIPHER_AES128_GCM_OPENSSH_COM",

"P22_SSH_BANNER_SERVER_KEY_EXCHANGE_SERVER_TO_CLIENT_CIPHER_AES128_CTR",

"P22_SSH_BANNER_SERVER_KEY_EXCHANGE_SERVER_TO_CLIENT_CIPHER_AES128_CBC",

"P22_SSH_BANNER_SERVER_KEY_EXCHANGE_SERVER_TO_CLIENT_CIPHER_3DES_CBC",

"P22_SSH_BANNER_SERVER_KEY_EXCHANGE_KEX_ALGORITHM_ECDH_SHA2_NISTP521",

"P22_SSH_BANNER_SERVER_KEY_EXCHANGE_KEX_ALGORITHM_ECDH_SHA2_NISTP384",

"P22_SSH_BANNER_SERVER_KEY_EXCHANGE_KEX_ALGORITHM_ECDH_SHA2_NISTP256",

"P22_SSH_BANNER_SERVER_KEY_EXCHANGE_KEX_ALGORITHM_DIFFIE_HELLMAN_GROUP_EXCHANGE_SHA256",

"P22_SSH_BANNER_SERVER_KEY_EXCHANGE_KEX_ALGORITHM_DIFFIE_HELLMAN_GROUP_EXCHANGE_SHA1",

"P22_SSH_BANNER_SERVER_KEY_EXCHANGE_KEX_ALGORITHM_DIFFIE_HELLMAN_GROUP1_SHA1",

"P22_SSH_BANNER_SERVER_KEY_EXCHANGE_KEX_ALGORITHM_DIFFIE_HELLMAN_GROUP14_SHA1",

"P22_SSH_BANNER_SERVER_KEY_EXCHANGE_KEX_ALGORITHM_CURVE25519_SHA256_LIBSSH_ORG",

"P22_SSH_BANNER_SERVER_KEY_EXCHANGE_HOST_KEY_ALGORITHM_SSH_RSA",

"P22_SSH_BANNER_SERVER_KEY_EXCHANGE_HOST_KEY_ALGORITHM_SSH_ED25519",

"P22_SSH_BANNER_SERVER_KEY_EXCHANGE_HOST_KEY_ALGORITHM_SSH_DSS",

"P22_SSH_BANNER_SERVER_KEY_EXCHANGE_HOST_KEY_ALGORITHM_ECDSA_SHA2_NISTP521",

"P22_SSH_BANNER_SERVER_KEY_EXCHANGE_HOST_KEY_ALGORITHM_ECDSA_SHA2_NISTP384",

"P22_SSH_BANNER_SERVER_KEY_EXCHANGE_HOST_KEY_ALGORITHM_ECDSA_SHA2_NISTP256_CERT_V01_OPENSSH_COM",

"P22_SSH_BANNER_SERVER_KEY_EXCHANGE_HOST_KEY_ALGORITHM_ECDSA_SHA2_NISTP256",

"P22_SSH_BANNER_SERVER_KEY_EXCHANGE_CLIENT_TO_SERVER_MAC_HMAC_SHA2_256",

"P22_SSH_BANNER_SERVER_KEY_EXCHANGE_CLIENT_TO_SERVER_MAC_HMAC_SHA1_96",

"P22_SSH_BANNER_SERVER_KEY_EXCHANGE_CLIENT_TO_SERVER_MAC_HMAC_SHA1",

"P22_SSH_BANNER_SERVER_KEY_EXCHANGE_CLIENT_TO_SERVER_MAC_HMAC_MD5_96",

"P22_SSH_BANNER_SERVER_KEY_EXCHANGE_CLIENT_TO_SERVER_MAC_HMAC_MD5",

"P22_SSH_BANNER_SERVER_KEY_EXCHANGE_CLIENT_TO_SERVER_COMPRESSION_ZLIB",

"P22_SSH_BANNER_SERVER_KEY_EXCHANGE_CLIENT_TO_SERVER_COMPRESSION_NONE",

"P22_SSH_BANNER_SERVER_KEY_EXCHANGE_CLIENT_TO_SERVER_CIPHER_ARCFOUR256",

"P22_SSH_BANNER_SERVER_KEY_EXCHANGE_CLIENT_TO_SERVER_CIPHER_ARCFOUR128",

"P22_SSH_BANNER_SERVER_KEY_EXCHANGE_CLIENT_TO_SERVER_CIPHER_ARCFOUR",

"P22_SSH_BANNER_SERVER_KEY_EXCHANGE_CLIENT_TO_SERVER_CIPHER_AES256_CTR",

"P22_SSH_BANNER_SERVER_KEY_EXCHANGE_CLIENT_TO_SERVER_CIPHER_AES192_CTR",

"P22_SSH_BANNER_SERVER_KEY_EXCHANGE_CLIENT_TO_SERVER_CIPHER_AES128_GCM_OPENSSH_COM",

"P22_SSH_BANNER_SERVER_KEY_EXCHANGE_CLIENT_TO_SERVER_CIPHER_AES128_CTR",

"P22_SSH_BANNER_SERVER_KEY_EXCHANGE_CLIENT_TO_SERVER_CIPHER_AES128_CBC",

"P22_SSH_BANNER_SERVER_KEY_EXCHANGE_CLIENT_TO_SERVER_CIPHER_3DES_CBC",

"P22_SSH_BANNER_ALGORITHM_SELECTION_SERVER_TO_CLIENT_MAC_HMAC_SHA2_256",

"P22_SSH_BANNER_ALGORITHM_SELECTION_SERVER_TO_CLIENT_MAC_HMAC_SHA1",

"P22_SSH_BANNER_ALGORITHM_SELECTION_SERVER_TO_CLIENT_COMPRESSION_NONE",

"P22_SSH_BANNER_ALGORITHM_SELECTION_SERVER_TO_CLIENT_CIPHER_ARCFOUR256",

"P22_SSH_BANNER_ALGORITHM_SELECTION_SERVER_TO_CLIENT_CIPHER_ARCFOUR128",

"P22_SSH_BANNER_ALGORITHM_SELECTION_SERVER_TO_CLIENT_CIPHER_ARCFOUR",

"P22_SSH_BANNER_ALGORITHM_SELECTION_SERVER_TO_CLIENT_CIPHER_AES256_CTR",

"P22_SSH_BANNER_ALGORITHM_SELECTION_SERVER_TO_CLIENT_CIPHER_AES192_CTR",

"P22_SSH_BANNER_ALGORITHM_SELECTION_SERVER_TO_CLIENT_CIPHER_AES128_GCM_OPENSSH_COM",

"P22_SSH_BANNER_ALGORITHM_SELECTION_SERVER_TO_CLIENT_CIPHER_AES128_CTR",

"P22_SSH_BANNER_ALGORITHM_SELECTION_KEX_ALGORITHM_ECDH_SHA2_NISTP256",

"P22_SSH_BANNER_ALGORITHM_SELECTION_KEX_ALGORITHM_DIFFIE_HELLMAN_GROUP1_SHA1",

"P22_SSH_BANNER_ALGORITHM_SELECTION_KEX_ALGORITHM_DIFFIE_HELLMAN_GROUP14_SHA1",

"P22_SSH_BANNER_ALGORITHM_SELECTION_KEX_ALGORITHM_CURVE25519_SHA256_LIBSSH_ORG",

"P22_SSH_BANNER_ALGORITHM_SELECTION_HOST_KEY_ALGORITHM_SSH_RSA",

"P22_SSH_BANNER_ALGORITHM_SELECTION_HOST_KEY_ALGORITHM_SSH_ED25519",

"P22_SSH_BANNER_ALGORITHM_SELECTION_HOST_KEY_ALGORITHM_SSH_DSS",

"P22_SSH_BANNER_ALGORITHM_SELECTION_HOST_KEY_ALGORITHM_ECDSA_SHA2_NISTP521",

"P22_SSH_BANNER_ALGORITHM_SELECTION_HOST_KEY_ALGORITHM_ECDSA_SHA2_NISTP384",

"P22_SSH_BANNER_ALGORITHM_SELECTION_HOST_KEY_ALGORITHM_ECDSA_SHA2_NISTP256_CERT_V01_OPENSSH_COM",

"P22_SSH_BANNER_ALGORITHM_SELECTION_HOST_KEY_ALGORITHM_ECDSA_SHA2_NISTP256",

"P22_SSH_BANNER_ALGORITHM_SELECTION_CLIENT_TO_SERVER_MAC_HMAC_SHA2_256",

"P22_SSH_BANNER_ALGORITHM_SELECTION_CLIENT_TO_SERVER_MAC_HMAC_SHA1",

"P22_SSH_BANNER_ALGORITHM_SELECTION_CLIENT_TO_SERVER_COMPRESSION_NONE",

"P22_SSH_BANNER_ALGORITHM_SELECTION_CLIENT_TO_SERVER_CIPHER_ARCFOUR256",

"P22_SSH_BANNER_ALGORITHM_SELECTION_CLIENT_TO_SERVER_CIPHER_ARCFOUR128",

"P22_SSH_BANNER_ALGORITHM_SELECTION_CLIENT_TO_SERVER_CIPHER_ARCFOUR",

"P22_SSH_BANNER_ALGORITHM_SELECTION_CLIENT_TO_SERVER_CIPHER_AES256_CTR",

"P22_SSH_BANNER_ALGORITHM_SELECTION_CLIENT_TO_SERVER_CIPHER_AES192_CTR",

"P22_SSH_BANNER_ALGORITHM_SELECTION_CLIENT_TO_SERVER_CIPHER_AES128_GCM_OPENSSH_COM",

"P22_SSH_BANNER_ALGORITHM_SELECTION_CLIENT_TO_SERVER_CIPHER_AES128_CTR",

"P21_FTP_BANNER_RUNNING_WS_FTP",

"P21_FTP_BANNER_RUNNING_WAR",

"P21_FTP_BANNER_RUNNING_VSFTPD",

"P21_FTP_BANNER_RUNNING_TITAN",

"P21_FTP_BANNER_RUNNING_SYSAX",

"P21_FTP_BANNER_RUNNING_SERV_U",

"P21_FTP_BANNER_RUNNING_PURE_FTPD",

"P21_FTP_BANNER_RUNNING_PROFTPD",

"P21_FTP_BANNER_RUNNING_NASFTP",

"P21_FTP_BANNER_RUNNING_MICROSOFT",

"P21_FTP_BANNER_RUNNING_FILEZILLA",

"P21_FTP_BANNER_RUNNING_DREAMHOST",

"P21_FTP_BANNER_RUNNING_CRUSHFTP",

"P21_FTP_BANNER_RUNNING_CERBERUS",

"P21_FTP_BANNER_RUNNING_BLAH",

"P21_FTP_BANNER_BANNER_UNKNOWN",

"P21_FTP_BANNER_BANNER_220_VSFTPD",

"P21_FTP_BANNER_BANNER_220_PROFTPD",

"P21_FTP_BANNER_BANNER_220_MICROSOFT_FTP_SERVICE",

"P21_FTP_BANNER_BANNER_220_FTP_SERVICE_READY",

"P21_FTP_BANNER_BANNER_220_FTP_SERVER_READY",

"P21_FTP_BANNER_BANNER_220_FTP_FIRMWARE_UPDATE_UTILITY",

"P21_FTP_BANNER_BANNER_220_FILEZILLA_SERVER_VERSION_0_9_41_BETA",

"P21_FTP_BANNER_BANNER_220_FILEZILLA_SERVER_0_9_60_BETA",

"P21_FTP_BANNER_BANNER_220",

"P1911_FOX_DEVICE_ID_VM_VERSION_2_3",

"P1911_FOX_DEVICE_ID_VM_VERSION_25_74_B02",

"P1911_FOX_DEVICE_ID_VM_VERSION_24_45_B08",

"P1911_FOX_DEVICE_ID_VM_VERSION_23_7_B01",

"P1911_FOX_DEVICE_ID_VM_VERSION_1_5_0_81_B06",

"P1911_FOX_DEVICE_ID_VM_VERSION_1_5_0_81_B02",

"P1911_FOX_DEVICE_ID_VM_VERSION_1_5_0_34_B29",

"P1911_FOX_DEVICE_ID_VM_VERSION_1_5_0_34_B28",

"P1911_FOX_DEVICE_ID_VM_NAME_JAVA_HOTSPOTTM_SERVER_VM",

"P1911_FOX_DEVICE_ID_VM_NAME_JAVA_HOTSPOTTM_EMBEDDED_CLIENT_VM",

"P1911_FOX_DEVICE_ID_VM_NAME_JAVA_HOTSPOTTM_CLIENT_VM",

"P1911_FOX_DEVICE_ID_VM_NAME_JAVA_HOTSPOTTM_64_BIT_SERVER_VM",

"P1911_FOX_DEVICE_ID_VM_NAME_J9",

"P1911_FOX_DEVICE_ID_VERSION_NIAGARA_4",

"P1911_FOX_DEVICE_ID_VERSION_1_0_1",

"P1911_FOX_DEVICE_ID_VERSION_1_0",

"P1911_FOX_DEVICE_ID_SYS_INFO_BOG_77",

"P1911_FOX_DEVICE_ID_SYS_INFO_BOG_74",

"P1911_FOX_DEVICE_ID_SYS_INFO_BOG_61",

"P1911_FOX_DEVICE_ID_SYS_INFO_BOG_6",

"P1911_FOX_DEVICE_ID_SUPPORT",

"P1911_FOX_DEVICE_ID_OS_VERSION_6_5_0",

"P1911_FOX_DEVICE_ID_OS_VERSION_6_4_1",

"P1911_FOX_DEVICE_ID_OS_VERSION_6_3_2",

"P1911_FOX_DEVICE_ID_OS_VERSION_6_3_0",

"P1911_FOX_DEVICE_ID_OS_VERSION_6_3",

"P1911_FOX_DEVICE_ID_OS_VERSION_6_2",

"P1911_FOX_DEVICE_ID_OS_VERSION_6_1",

"P1911_FOX_DEVICE_ID_OS_NAME_WINDOWS_SERVER_2012",

"P1911_FOX_DEVICE_ID_OS_NAME_WINDOWS_SERVER_2008_R2",

"P1911_FOX_DEVICE_ID_OS_NAME_WINDOWS_SERVER_2008",

"P1911_FOX_DEVICE_ID_OS_NAME_WINDOWS_7",

"P1911_FOX_DEVICE_ID_OS_NAME_WINDOWS",

"P1911_FOX_DEVICE_ID_OS_NAME_QNX",

"P1911_FOX_DEVICE_ID_ID",

"P1911_FOX_DEVICE_ID_BRAND_ID_WEBSOPEN",

"P1911_FOX_DEVICE_ID_BRAND_ID_WEBS",

"P1911_FOX_DEVICE_ID_BRAND_ID_VYKON",

"P1911_FOX_DEVICE_ID_BRAND_ID_TAC",

"P1911_FOX_DEVICE_ID_BRAND_ID_STAEFA",

"P1911_FOX_DEVICE_ID_BRAND_ID_JENESYS",

"P1911_FOX_DEVICE_ID_BRAND_ID_FACEXP",

"P1911_FOX_DEVICE_ID_BRAND_ID_DISTECH",

"P1911_FOX_DEVICE_ID_AUTH_AGENT_TYPE_LDAP",

"P1911_FOX_DEVICE_ID_AUTH_AGENT_TYPE_FOX",

"P1911_FOX_DEVICE_ID_APP_VERSION_3_8_38",

"P1911_FOX_DEVICE_ID_APP_VERSION_3_8_213",

"P1911_FOX_DEVICE_ID_APP_VERSION_3_8_111",

"P1911_FOX_DEVICE_ID_APP_VERSION_3_7_106_8",

"P1911_FOX_DEVICE_ID_APP_VERSION_3_7_106_5",

"P1911_FOX_DEVICE_ID_APP_VERSION_3_7_106_4",

"P1911_FOX_DEVICE_ID_APP_VERSION_3_7_106_1",

"P1911_FOX_DEVICE_ID_APP_NAME_STATION",

"P1900_UPNP_DISCOVERY_FIELD_PRESENT",

"P1521_ORACLE_BANNER_SUPPORT",

"P1521_ORACLE_BANNER_REFUSE_VERSION_12_1",

"P1521_ORACLE_BANNER_REFUSE_VERSION_11_2",

"P1521_ORACLE_BANNER_REFUSE_VERSION_10_1",

"P143_IMAP_TLS_TLS_FIELD_PRESENT",

"P143_IMAP_TLS_TLS_CERT_PAST_VALID_END_DATE",

"P143_IMAP_STARTTLS_TLS_VERSION_TLSV1_2",

"P143_IMAP_STARTTLS_TLS_VERSION_TLSV1_0",

"P143_IMAP_STARTTLS_TLS_VALIDATION_BROWSER_TRUSTED",

"P143_IMAP_STARTTLS_TLS_SIGNATURE_VALID",

"P143_IMAP_STARTTLS_TLS_SERVER_KEY_EXCHANGE_ECDH_PARAMS_CURVE_ID_ID",

"P143_IMAP_STARTTLS_TLS_OCSP_STAPLING",

"P143_IMAP_STARTTLS_TLS_CIPHER_SUITE_NAME_TLS_RSA_WITH_RC4_128_SHA",

"P143_IMAP_STARTTLS_TLS_CIPHER_SUITE_NAME_TLS_RSA_WITH_AES_256_CBC_SHA",

"P143_IMAP_STARTTLS_TLS_CIPHER_SUITE_NAME_TLS_RSA_WITH_AES_128_CBC_SHA",

"P143_IMAP_STARTTLS_TLS_CIPHER_SUITE_NAME_TLS_ECDHE_RSA_WITH_AES_256_CBC_SHA",

"P143_IMAP_STARTTLS_TLS_CIPHER_SUITE_NAME_TLS_ECDHE_RSA_WITH_AES_128_GCM_SHA256",

"P143_IMAP_STARTTLS_TLS_CERTIFICATE_PARSED_VERSION",

"P143_IMAP_STARTTLS_TLS_CERTIFICATE_PARSED_VALIDITY_LENGTH",

"P143_IMAP_STARTTLS_TLS_CERTIFICATE_PARSED_VALIDATION_LEVEL_UNKNOWN",

"P143_IMAP_STARTTLS_TLS_CERTIFICATE_PARSED_VALIDATION_LEVEL_OV",

"P143_IMAP_STARTTLS_TLS_CERTIFICATE_PARSED_VALIDATION_LEVEL_EV",

"P143_IMAP_STARTTLS_TLS_CERTIFICATE_PARSED_VALIDATION_LEVEL_DV",

"P143_IMAP_STARTTLS_TLS_CERTIFICATE_PARSED_SIGNATURE_VALID",

"P143_IMAP_STARTTLS_TLS_CERTIFICATE_PARSED_SIGNATURE_SELF_SIGNED",

"P143_IMAP_STARTTLS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_UNKNOWN",

"P143_IMAP_STARTTLS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_STARFIELD",

"P143_IMAP_STARTTLS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_SOMEORGANIZATION",

"P143_IMAP_STARTTLS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_RAPIDSSL",

"P143_IMAP_STARTTLS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_LOCALHOST",

"P143_IMAP_STARTTLS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_LETS_ENCRYPT",

"P143_IMAP_STARTTLS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_HOME_PL",

"P143_IMAP_STARTTLS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_GODADDY",

"P143_IMAP_STARTTLS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_GLOBALSIGN",

"P143_IMAP_STARTTLS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_GEOTRUST",

"P143_IMAP_STARTTLS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_FORTINET",

"P143_IMAP_STARTTLS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_ENTRUST",

"P143_IMAP_STARTTLS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_DIGICERT",

"P143_IMAP_STARTTLS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_CPANEL",

"P143_IMAP_STARTTLS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_COMODO",

"P143_IMAP_STARTTLS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_CISCO",

"P143_IMAP_STARTTLS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_ALPHASSL",

"P143_IMAP_STARTTLS_TLS_CERTIFICATE_PARSED_EXTENSIONS_KEY_USAGE_KEY_ENCIPHERMENT",

"P143_IMAP_STARTTLS_TLS_CERTIFICATE_PARSED_EXTENSIONS_KEY_USAGE_KEY_AGREEMENT",

"P143_IMAP_STARTTLS_TLS_CERTIFICATE_PARSED_EXTENSIONS_KEY_USAGE_DIGITAL_SIGNATURE",

"P143_IMAP_STARTTLS_TLS_CERTIFICATE_PARSED_EXTENSIONS_KEY_USAGE_DATA_ENCIPHERMENT",

"P143_IMAP_STARTTLS_TLS_CERTIFICATE_PARSED_EXTENSIONS_KEY_USAGE_CONTENT_COMMITMENT",

"P143_IMAP_STARTTLS_TLS_CERTIFICATE_PARSED_EXTENSIONS_KEY_USAGE_CERTIFICATE_SIGN",

"P143_IMAP_STARTTLS_TLS_CERTIFICATE_PARSED_EXTENSIONS_BASIC_CONSTRAINTS_IS_CA",

"P143_IMAP_STARTTLS_STARTTLS_UNKNOWN",

"P143_IMAP_STARTTLS_STARTTLS_A001_OK_COMPLETED",

"P143_IMAP_STARTTLS_STARTTLS_A001_OK_BEGIN",

"P143_IMAP_STARTTLS_STARTTLS_A001_NO_ERROR_IN_IMAP_COMMAND_RECEIVED_BY_SERVER_",

"P143_IMAP_STARTTLS_STARTTLS_A001_NO",

"P143_IMAP_STARTTLS_STARTTLS_A001_BAD_UNKNOWN",

"P143_IMAP_STARTTLS_STARTTLS_A001_BAD_TLS_SUPPORT_ISNT_ENABLED_",

"P143_IMAP_STARTTLS_STARTTLS_A001_BAD_INVALID_COMMAND",

"P143_IMAP_STARTTLS_RUNNING_ZIMBRA",

"P143_IMAP_STARTTLS_RUNNING_XCHANGE",

"P143_IMAP_STARTTLS_RUNNING_UNKNOWN",

"P143_IMAP_STARTTLS_RUNNING_SUN",

"P143_IMAP_STARTTLS_RUNNING_SMAIL",

"P143_IMAP_STARTTLS_RUNNING_ORACLE",

"P143_IMAP_STARTTLS_RUNNING_MICROSOFT_EXCHANGE_SERVER_2007",

"P143_IMAP_STARTTLS_RUNNING_MICROSOFT_EXCHANGE_SERVER_2003",

"P143_IMAP_STARTTLS_RUNNING_MICROSOFT",

"P143_IMAP_STARTTLS_RUNNING_MERCURY",

"P143_IMAP_STARTTLS_RUNNING_MDAEMON",

"P143_IMAP_STARTTLS_RUNNING_MAILSITE",

"P143_IMAP_STARTTLS_RUNNING_KERIO_CONNECT",

"P143_IMAP_STARTTLS_RUNNING_KERIO",

"P143_IMAP_STARTTLS_RUNNING_IMAP",

"P143_IMAP_STARTTLS_RUNNING_ICEWARP",

"P143_IMAP_STARTTLS_RUNNING_IBM",

"P143_IMAP_STARTTLS_RUNNING_GROUPWISE",

"P143_IMAP_STARTTLS_RUNNING_FIRSTCLASS",

"P143_IMAP_STARTTLS_RUNNING_EXCHANGE_SERVER",

"P143_IMAP_STARTTLS_RUNNING_DOVECOT",

"P143_IMAP_STARTTLS_RUNNING_CYRUS",

"P143_IMAP_STARTTLS_RUNNING_COURIER",

"P143_IMAP_STARTTLS_RUNNING_AXIGEN",

"P143_IMAP_STARTTLS_CONN_SUCCESS",

"P143_IMAP_SSL_2_TLS_CERT_PAST_VALID_END_DATE",

"P143_IMAP_SSL_2_SSL_2_SUPPORT",

"P143_IMAP_SSL_2_CERTIFICATE_PARSED_VERSION",

"P143_IMAP_SSL_2_CERTIFICATE_PARSED_SIGNATURE_SELF_SIGNED",

"P143_IMAP_SSL_2_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_GODADDY",

"P143_IMAP_SSL_2_CERTIFICATE_PARSED_EXTENSIONS_KEY_USAGE_KEY_ENCIPHERMENT",

"P143_IMAP_SSL_2_CERTIFICATE_PARSED_EXTENSIONS_KEY_USAGE_DIGITAL_SIGNATURE",

"P143_IMAP_SSL_2_CERTIFICATE_PARSED_EXTENSIONS_BASIC_CONSTRAINTS_IS_CA",

"P1433_MSSQL_TLS_TLS_FIELD_PRESENT",

"P1433_MSSQL_BANNER_VERSION_9_0",

"P1433_MSSQL_BANNER_VERSION_1_0",

"P1433_MSSQL_BANNER_VERSION_14_0",

"P1433_MSSQL_BANNER_VERSION_13_0",

"P1433_MSSQL_BANNER_VERSION_12_0",

"P1433_MSSQL_BANNER_VERSION_11_0",

"P1433_MSSQL_BANNER_VERSION_10_50",

"P1433_MSSQL_BANNER_VERSION_10_5",

"P1433_MSSQL_BANNER_VERSION_10_0",

"P1433_MSSQL_BANNER_TLS_VERSION_TLSV1_2",

"P1433_MSSQL_BANNER_TLS_VERSION_TLSV1_0",

"P1433_MSSQL_BANNER_TLS_VALIDATION_BROWSER_TRUSTED",

"P1433_MSSQL_BANNER_TLS_SIGNATURE_VALID",

"P1433_MSSQL_BANNER_TLS_SERVER_KEY_EXCHANGE_ECDH_PARAMS_CURVE_ID_ID",

"P1433_MSSQL_BANNER_TLS_OCSP_STAPLING",

"P1433_MSSQL_BANNER_TLS_CIPHER_SUITE_NAME_TLS_RSA_WITH_RC4_128_SHA",

"P1433_MSSQL_BANNER_TLS_CIPHER_SUITE_NAME_TLS_RSA_WITH_AES_128_CBC_SHA",

"P1433_MSSQL_BANNER_TLS_CIPHER_SUITE_NAME_TLS_ECDHE_RSA_WITH_AES_256_CBC_SHA",

"P1433_MSSQL_BANNER_TLS_CIPHER_SUITE_NAME_TLS_ECDHE_RSA_WITH_AES_128_GCM_SHA256",

"P1433_MSSQL_BANNER_TLS_CERTIFICATE_PARSED_VERSION",

"P1433_MSSQL_BANNER_TLS_CERTIFICATE_PARSED_VALIDITY_LENGTH",

"P1433_MSSQL_BANNER_TLS_CERTIFICATE_PARSED_VALIDATION_LEVEL_UNKNOWN",

"P1433_MSSQL_BANNER_TLS_CERTIFICATE_PARSED_VALIDATION_LEVEL_OV",

"P1433_MSSQL_BANNER_TLS_CERTIFICATE_PARSED_SIGNATURE_VALID",

"P1433_MSSQL_BANNER_TLS_CERTIFICATE_PARSED_SIGNATURE_SELF_SIGNED",

"P1433_MSSQL_BANNER_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_AMAZON",

"P1433_MSSQL_BANNER_TLS_CERTIFICATE_PARSED_EXTENSIONS_KEY_USAGE_KEY_ENCIPHERMENT",

"P1433_MSSQL_BANNER_TLS_CERTIFICATE_PARSED_EXTENSIONS_KEY_USAGE_DIGITAL_SIGNATURE",

"P1433_MSSQL_BANNER_TLS_CERTIFICATE_PARSED_EXTENSIONS_KEY_USAGE_DATA_ENCIPHERMENT",

"P1433_MSSQL_BANNER_SUPPORT",

"P1433_MSSQL_BANNER_ENCRYPT_MODE_ENCRYPT_ON",

"P110_POP3_TLS_TLS_FIELD_PRESENT",

"P110_POP3_TLS_TLS_CERT_PAST_VALID_END_DATE",

"P110_POP3_STARTTLS_TLS_VERSION_TLSV1_2",

"P110_POP3_STARTTLS_TLS_VERSION_TLSV1_0",

"P110_POP3_STARTTLS_TLS_VALIDATION_BROWSER_TRUSTED",

"P110_POP3_STARTTLS_TLS_SIGNATURE_VALID",

"P110_POP3_STARTTLS_TLS_SERVER_KEY_EXCHANGE_ECDH_PARAMS_CURVE_ID_ID" ,

"P110_POP3_STARTTLS_TLS_OCSP_STAPLING" ,

"P110_POP3_STARTTLS_TLS_CIPHER_SUITE_NAME_TLS_RSA_WITH_RC4_128_SHA" ,

"P110_POP3_STARTTLS_TLS_CIPHER_SUITE_NAME_TLS_RSA_WITH_AES_256_CBC_SHA" ,

"P110_POP3_STARTTLS_TLS_CIPHER_SUITE_NAME_TLS_RSA_WITH_AES_128_CBC_SHA" ,

"P110_POP3_STARTTLS_TLS_CIPHER_SUITE_NAME_TLS_ECDHE_RSA_WITH_AES_256_CBC_SHA" ,

"P110_POP3_STARTTLS_TLS_CIPHER_SUITE_NAME_TLS_ECDHE_RSA_WITH_AES_128_GCM_SHA256" ,

"P110_POP3_STARTTLS_TLS_CERTIFICATE_PARSED_VERSION" ,

"P110_POP3_STARTTLS_TLS_CERTIFICATE_PARSED_VALIDITY_LENGTH" ,

"P110_POP3_STARTTLS_TLS_CERTIFICATE_PARSED_VALIDATION_LEVEL_UNKNOWN" ,

"P110_POP3_STARTTLS_TLS_CERTIFICATE_PARSED_VALIDATION_LEVEL_OV" ,

"P110_POP3_STARTTLS_TLS_CERTIFICATE_PARSED_VALIDATION_LEVEL_EV" ,

"P110_POP3_STARTTLS_TLS_CERTIFICATE_PARSED_VALIDATION_LEVEL_DV" ,

"P110_POP3_STARTTLS_TLS_CERTIFICATE_PARSED_SIGNATURE_VALID" ,

"P110_POP3_STARTTLS_TLS_CERTIFICATE_PARSED_SIGNATURE_SELF_SIGNED" ,

"P110_POP3_STARTTLS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_UNKNOWN" ,

"P110_POP3_STARTTLS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_STARFIELD" ,

"P110_POP3_STARTTLS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_SOMEORGANIZATION" ,

"P110_POP3_STARTTLS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_RAPIDSSL" ,

"P110_POP3_STARTTLS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_LOCALHOST" ,

"P110_POP3_STARTTLS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_LETS_ENCRYPT" ,

"P110_POP3_STARTTLS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_HOME_PL" ,

"P110_POP3_STARTTLS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_GODADDY" ,

"P110_POP3_STARTTLS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_GLOBALSIGN" ,

"P110_POP3_STARTTLS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_GEOTRUST" ,

"P110_POP3_STARTTLS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_FORTINET" ,

"P110_POP3_STARTTLS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_ENTRUST" ,

"P110_POP3_STARTTLS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_DIGICERT" ,

"P110_POP3_STARTTLS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_CPANEL" ,

"P110_POP3_STARTTLS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_COMODO" ,

"P110_POP3_STARTTLS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_CISCO" ,

"P110_POP3_STARTTLS_TLS_CERTIFICATE_PARSED_ISSUER_ORGANIZATION_ALPHASSL" ,

"P110_POP3_STARTTLS_TLS_CERTIFICATE_PARSED_EXTENSIONS_KEY_USAGE_KEY_ENCIPHERMENT" ,

"P110_POP3_STARTTLS_TLS_CERTIFICATE_PARSED_EXTENSIONS_KEY_USAGE_KEY_AGREEMENT" ,

"P110_POP3_STARTTLS_TLS_CERTIFICATE_PARSED_EXTENSIONS_KEY_USAGE_DIGITAL_SIGNATURE",

"P110_POP3_STARTTLS_TLS_CERTIFICATE_PARSED_EXTENSIONS_KEY_USAGE_DATA_ENCIPHERMENT",

"P110_POP3_STARTTLS_TLS_CERTIFICATE_PARSED_EXTENSIONS_KEY_USAGE_CRL_SIGN",

"P110_POP3_STARTTLS_TLS_CERTIFICATE_PARSED_EXTENSIONS_KEY_USAGE_CONTENT_COMMITMENT",

"P110_POP3_STARTTLS_TLS_CERTIFICATE_PARSED_EXTENSIONS_KEY_USAGE_CERTIFICATE_SIGN",

"P110_POP3_STARTTLS_TLS_CERTIFICATE_PARSED_EXTENSIONS_BASIC_CONSTRAINTS_IS_CA",

"P110_POP3_STARTTLS_STARTTLS_UNKNOWN",

"P110_POP3_STARTTLS_RUNNING_ZIMBRA",

"P110_POP3_STARTTLS_RUNNING_XCHANGE",

"P110_POP3_STARTTLS_RUNNING_SUN",

"P110_POP3_STARTTLS_RUNNING_SMAIL",

"P110_POP3_STARTTLS_RUNNING_QPOPPER",

"P110_POP3_STARTTLS_RUNNING_ORACLE",

"P110_POP3_STARTTLS_RUNNING_NETMAIL",

"P110_POP3_STARTTLS_RUNNING_MICROSOFT_EXCHANGE_SERVER_2007",

"P110_POP3_STARTTLS_RUNNING_MICROSOFT_EXCHANGE_SERVER_2003",

"P110_POP3_STARTTLS_RUNNING_MICROSOFT",

"P110_POP3_STARTTLS_RUNNING_MERCURY",

"P110_POP3_STARTTLS_RUNNING_MDAEMON",

"P110_POP3_STARTTLS_RUNNING_MAILENABLE",

"P110_POP3_STARTTLS_RUNNING_KERIO_CONNECT",

"P110_POP3_STARTTLS_RUNNING_KERIO",

"P110_POP3_STARTTLS_RUNNING_IMAP",

"P110_POP3_STARTTLS_RUNNING_ICEWARP",

"P110_POP3_STARTTLS_RUNNING_IBM",

"P110_POP3_STARTTLS_RUNNING_GROUPWISE",

"P110_POP3_STARTTLS_RUNNING_GORDANO",

"P110_POP3_STARTTLS_RUNNING_EXCHANGE_SERVER",

"P110_POP3_STARTTLS_RUNNING_DOVECOT",

"P110_POP3_STARTTLS_RUNNING_CYRUS",

"P110_POP3_STARTTLS_CONN_SUCCESS",

"P110_POP3_SSL_2_TLS_CERT_PAST_VALID_END_DATE",

"P110_POP3_SSL_2_SSL_2_SUPPORT",

"P110_POP3_SSL_2_CERTIFICATE_PARSED_VERSION",

"P110_POP3_SSL_2_CERTIFICATE_PARSED_SIGNATURE_SELF_SIGNED",

"P110_POP3_SSL_2_CERTIFICATE_PARSED_EXTENSIONS_BASIC_CONSTRAINTS_IS_CA",

"P102_S7_SZL_SUPPORT",

"ORG_SIZE",

"NUM_PORTS",

"NUM_MEDIUM_RISK_SERVICES_RUNNING",

"NUM_LOW_RISK_SERVICES_RUNNING",

"NUM_HIGH_RISK_SERVICES_RUNNING",

"METADATA_DESCRIPTION_WINDOWS",

"METADATA_DESCRIPTION_UNIX",

"METADATA_DESCRIPTION_UBUNTU",

"METADATA_DESCRIPTION_REDHAT",

"METADATA_DESCRIPTION_RASPBIAN",

"METADATA_DESCRIPTION_QNX",

"METADATA_DESCRIPTION_ORACLE",

"METADATA_DESCRIPTION_MAGEIA",

"METADATA_DESCRIPTION_LINUX",

"METADATA_DESCRIPTION_HPE",

"METADATA_DESCRIPTION_HP",

"METADATA_DESCRIPTION_FREEBSD",

"METADATA_DESCRIPTION_FEDORA",

"METADATA_DESCRIPTION_DEBIAN",

"METADATA_DESCRIPTION_CIS",

"METADATA_DESCRIPTION_CENTOS",

"METADATA_DESCRIPTION_APPLE",

"LABEL",

"IP_ADDRESS",

"DOMAIN",

"COMPANY_NAME_IN_ASN",

"CENSYS_DATE_TABLE",

"AUTONOMOUS_SYSTEM_NAME_UNIFIED",

"AUTONOMOUS_SYSTEM_NAME_PSYCHZ",

"AUTONOMOUS_SYSTEM_NAME_MICROSOFT",

"AUTONOMOUS_SYSTEM_NAME_LEASEWEB",

"AUTONOMOUS_SYSTEM_NAME_GOOGLE",

"AUTONOMOUS_SYSTEM_NAME_GODADDY",

"AUTONOMOUS_SYSTEM_NAME_EGIHOSTING",

"AUTONOMOUS_SYSTEM_NAME_DIGITALOCEAN",

"AUTONOMOUS_SYSTEM_NAME_COMCAST",

"AUTONOMOUS_SYSTEM_NAME_CLOUDFLARE",

"AUTONOMOUS_SYSTEM_NAME_CHARTER",

"AUTONOMOUS_SYSTEM_NAME_CENTURYLINK",

"AUTONOMOUS_SYSTEM_NAME_ATT",

"AUTONOMOUS_SYSTEM_NAME_AMAZON",

"AUTONOMOUS_SYSTEM_NAME_AKAMAI"

]

A.4 Additional analysis charts

Since the models learned features that were useful per cohort, another chart was generated only containing similar correlations across the cohort subsets.

Sample Security Breach Notification Letter

Date

Dear Recipient Name:

We are contacting you because we have learned of a serious data security incident that occurred on (*specific or approximate date*) *OR* between (*date, year and date, year*) that involved some of your personal information.

The breach involved (*provide a brief general description of the breach and include how many records or people it may have affected*). The information breached contained (*customer names, mailing addresses, credit card numbers, and/or Social Security numbers, etc.*). Other information (*bank account PIN, security codes, etc.*) was not released.

We are notifying you so you can take action along with our efforts to minimize or eliminate potential harm. Because this is a serious incident, we strongly encourage you to take preventive measures now to help prevent and detect any misuse of your information. We have advised the three major U.S. credit reporting agencies about this incident and have given those agencies a general report, alerting them to the fact that the incident occurred, however, we have not notified them about the presence of your specific information in the data breach.*

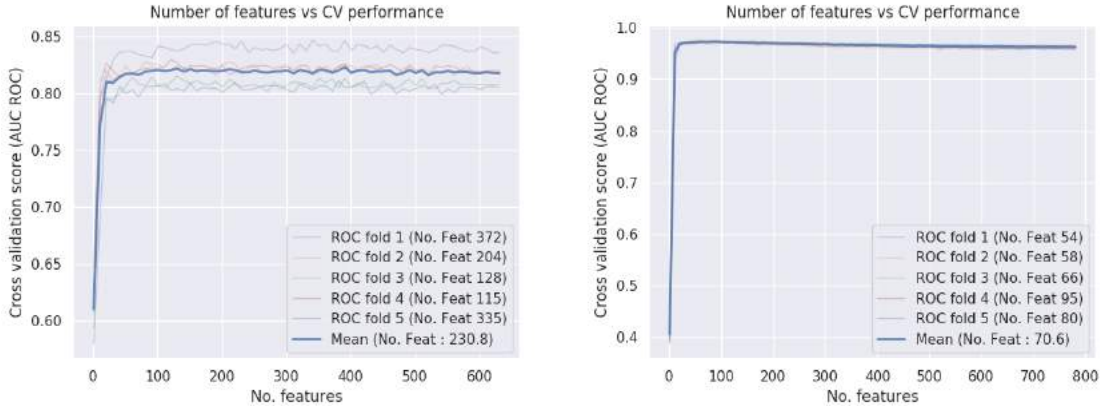
(*Optional paragraph if offering credit protection service.***)
To protect you we have retained (*name of identity theft company*), a specialist in identity theft protection, to provide you with ___ year(s) of (description of services) services, free of charge. You can enroll in the program by following the directions below. **Please keep this letter; you will need the personal access code it contains in order to register for services.**

As a first preventive step, we recommend you closely monitor your financial accounts and, if you see any unauthorized activity, promptly contact your financial institution. We also suggest you submit a complaint with the Federal Trade Commission (FTC) by calling 1-877-ID-THEFT (1-877-438-4338) or online at https://www.ftccomplaintassistant.gov/

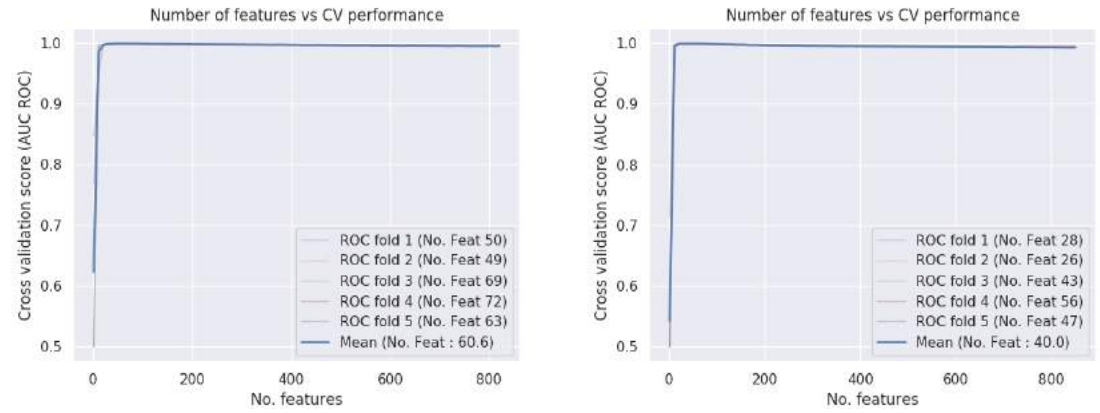
As a second step, you also may want to contact the three U.S. credit reporting agencies (Equifax, Experian and TransUnion) to obtain a free credit report from each by calling 1-877-322-8228 or by logging onto www.annualcreditreport.com.

Even if you do not find any suspicious activity on your initial credit reports, the FTC recommends that you check your credit reports periodically. A victim's personal information is sometimes held for use or shared among a group of thieves at different times. Checking your credit reports periodically can help you spot problems and address them quickly.

Figure A.1: Sample incident notification letter [6]

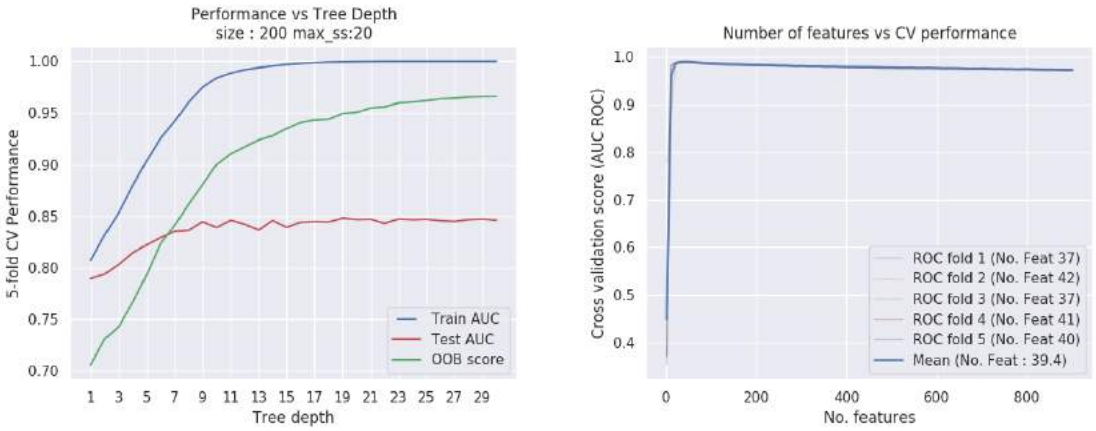


(a) Organization size: [0 to 10] (left) and [10 to 100] (right)

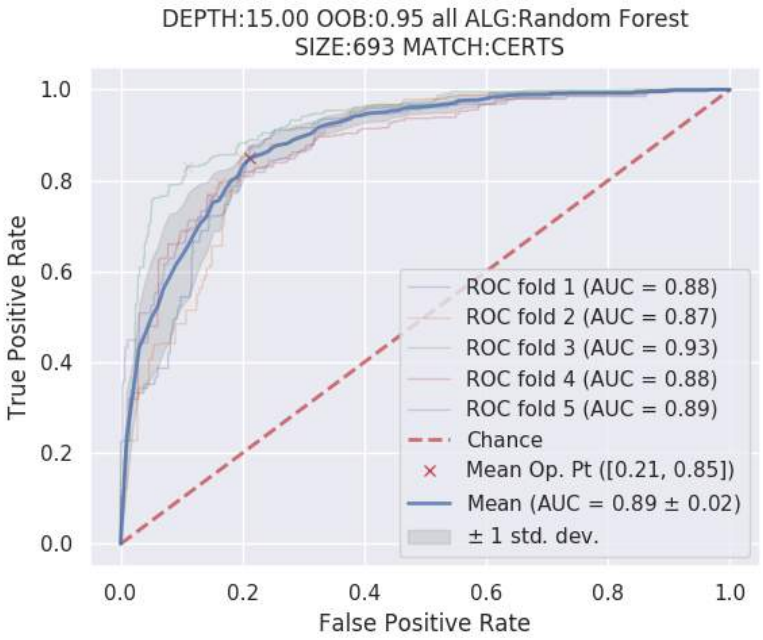


(b) Organization size: [100 to 1000] (left) and [1000 and above] (right)

Figure A.2: Outlier vs. inlier classification for different organization sizes using all attributions (RFE)

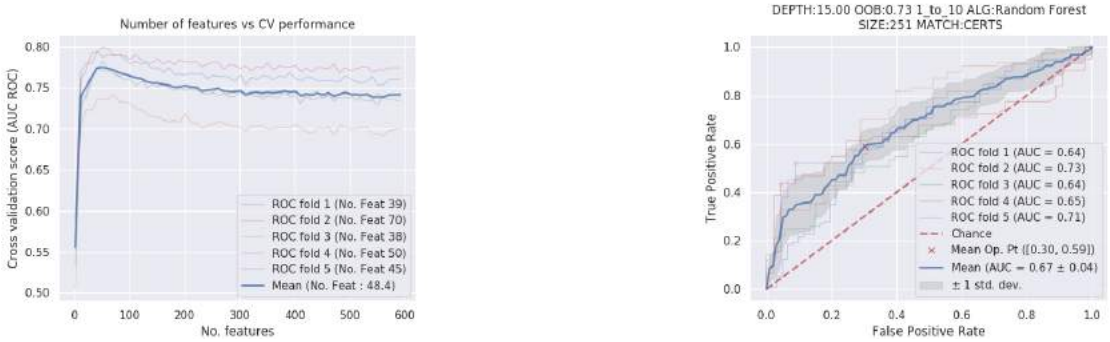


(a) Tree depth tuning and recursive feature selection

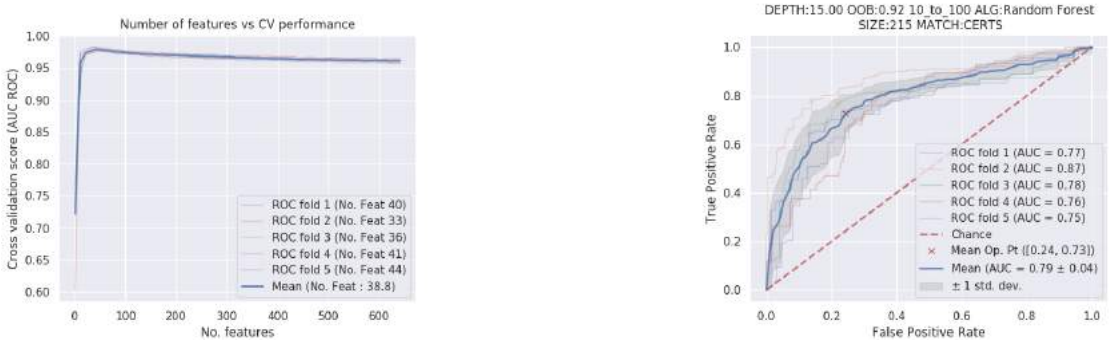


(b) ROC curve

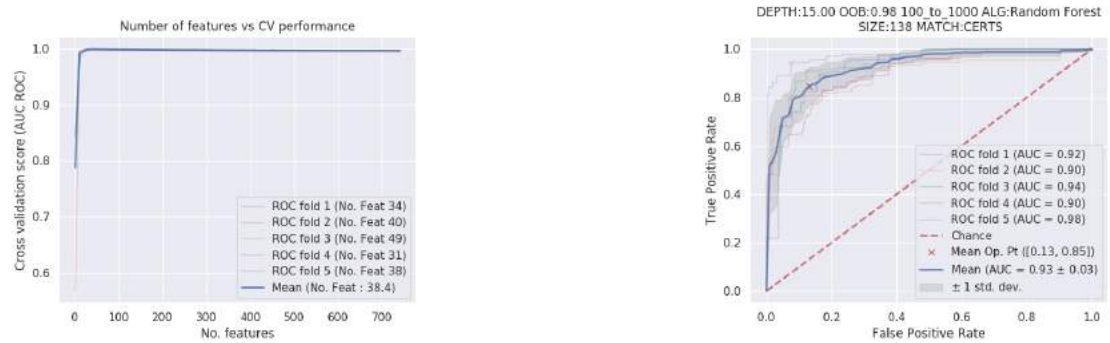
Figure A.3: Outlier vs. inlier classification using only certificate attributions



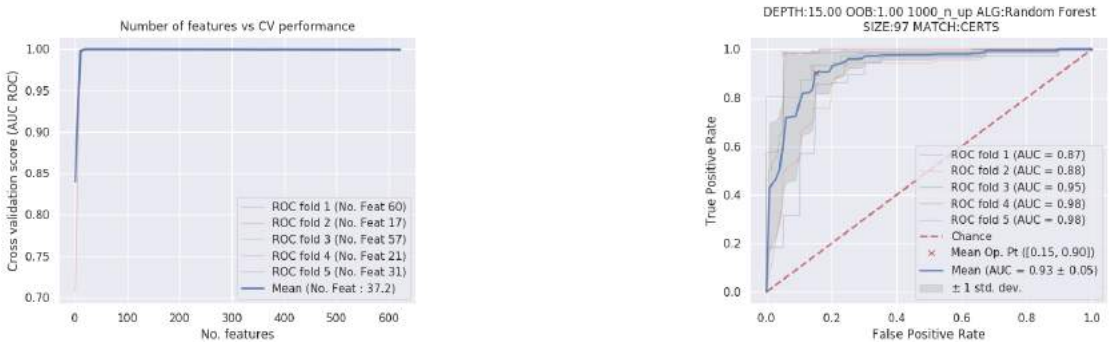
(a) Organization size: [0 to 10]



(b) Organization size: [10 to 100]

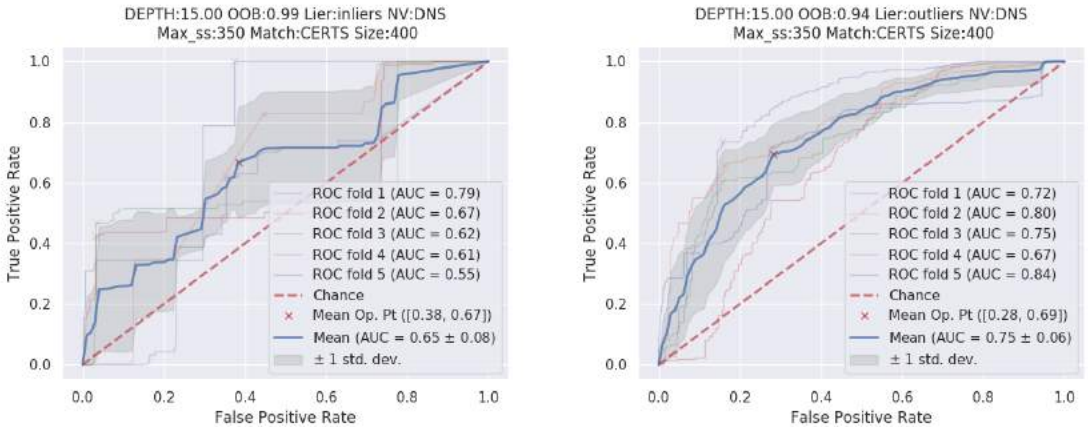


(c) Organization size: [100 to 1000]

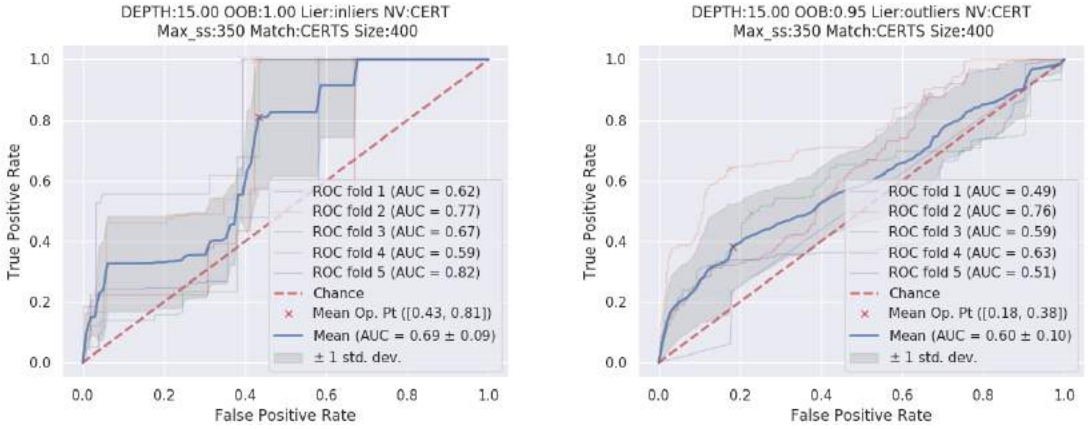


(d) Organization size: [1000 and above]

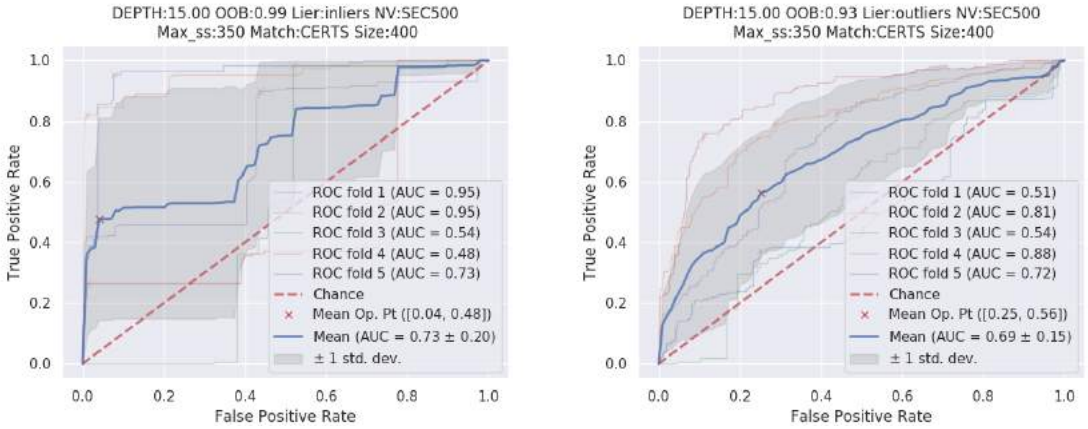
Figure A.4: Outlier vs. inlier classification for different organization sizes using only certificate attributions



(a) Random sampled DNS non-victims: inliers (left) and outliers (right)

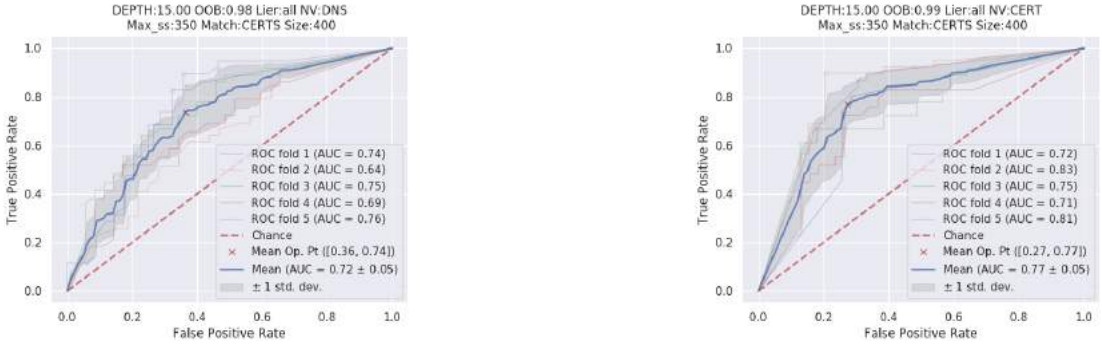


(b) Random sampled CERT non-victims: inliers (left) and outliers (right)

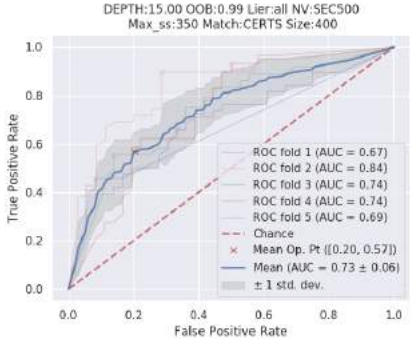


(c) Random sampled SEC500 non-victims: inliers (left) and outliers (right)

Figure A.5: Non-victim vs. victim host classification using only certificate attributions



(a) Random sampled DNS (left) and random sampled CERT (right)



(b) SEC500

Figure A.6: Victim vs. non-victim organization classification using only certificate attributions

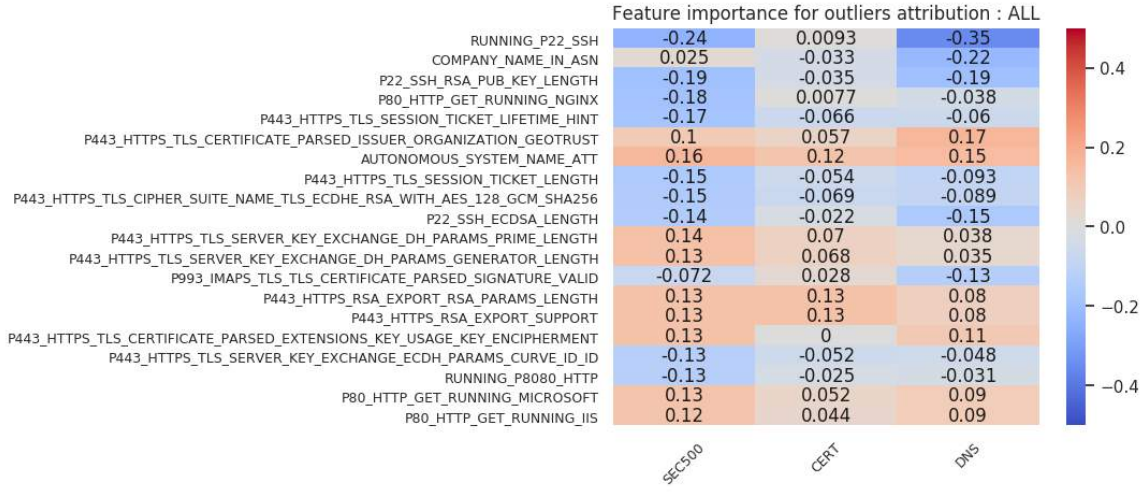


Figure A.7: Victim vs. non-victim outlier host classification using all attributions (similar correlation)

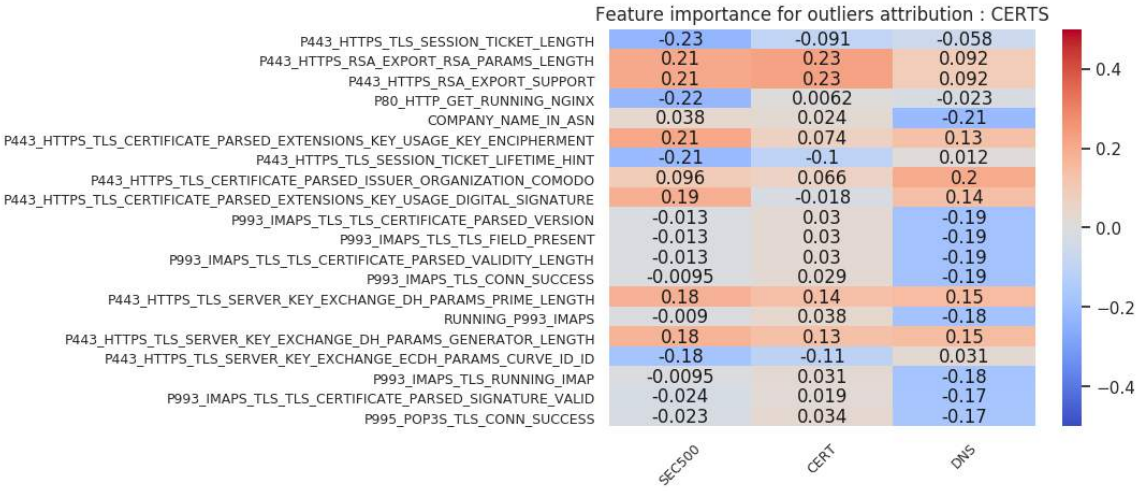


Figure A.8: Victim vs. non-victim outlier host classification using only certificate attributions (similar correlation)

Figure A.9: Victim vs. non-victim inlier host classification using all attributions (similar correlation)

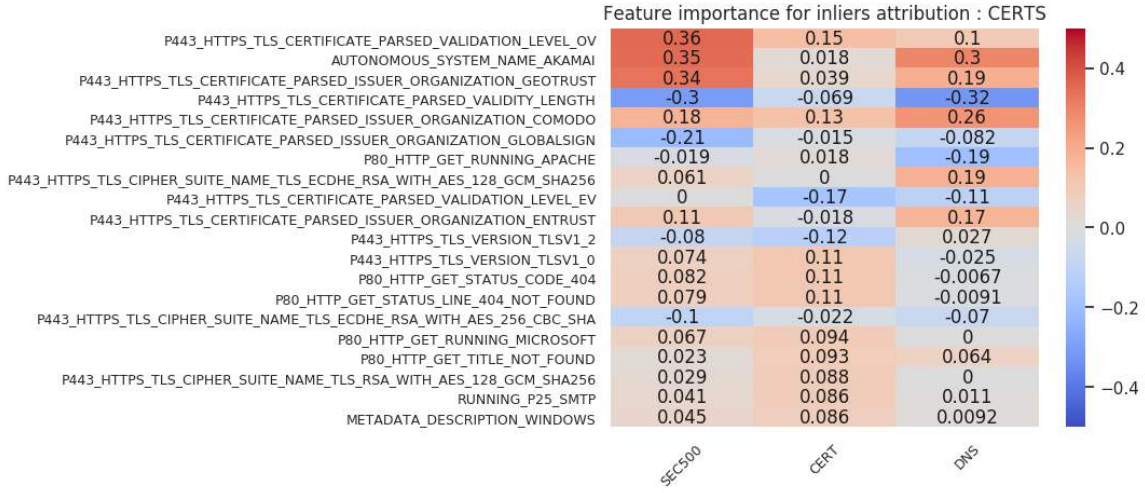


Figure A.10: Victim vs. non-victim inlier host classification using only certificate attributions (similar correlation)

Bibliography

[1] Jing Zhang, Zakir Durumeric, Michael Bailey, Mingyan Liu, and Manish Karir. On the mismanagement and maliciousness of networks. In *NDSS*, 2014.

[2] Yang Liu, Armin Sarabi, Jing Zhang, Parinaz Naghizadeh, Manish Karir, Michael Bailey, and Mingyan Liu. Cloudy with a chance of breach: Forecasting cyber security incidents. In *USENIX Security Symposium*, pages 1009–1024, 2015.

[3] American registry for Internet Numbers. American registry for internet numbers. URL `https://whois.arin.net/ui/`.

[4] Zakir Durumeric, David Adrian, Ariana Mirian, Michael Bailey, and J Alex Halderman. A search engine backed by internet-wide scanning. In *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security*, pages 542–553. ACM, 2015.

[5] Fei Tony Liu, Kai Ming Ting, and Zhi-Hua Zhou. Isolation forest. In *2008 Eighth IEEE International Conference on Data Mining*, pages 413–422. IEEE, 2008.

[6] Division of Financial Regulation. Sample breach notification letter. URL `https://dfr.oregon.gov/business/Documents/Sample-Letter-2016.pdf`.

[7] Privacy Rights Clearing House. Privacy rights clearing house, . URL `https://www.privacyrights.org/`.

[8] Privacy Rights Clearing House. Breaches for 2017-18, . URL `https://www.privacyrights.org/data-breaches?title=&taxonomy_vocabulary_11_tid%5B%5D=2436&taxonomy_vocabulary_11_tid%5B%5D=2434`.

[9] Benjamin Edwards, Steven Hofmeyr, and Stephanie Forrest. Hype and heavy tails: A closer look at data breaches. *Journal of Cybersecurity*, 2(1):3–14, 2016.

[10] Armin Sarabi. Quantifying security: Methods, challenges and applications. 2018.

[11] K Aditya, Slawomir Grzonkowski, and Nhien-An Le-Khac. Riskwriter: Predicting cyber risk of an enterprise. In *International Conference on Information Systems Security*, pages 88–106. Springer, 2018.

[12] FICO. Cyber risk score. URL `https://www.fico.com/en/products/cyber-risk-score`.

[13] BitSight. Bitsight, . URL `https://www.bitsight.com/`.

[14] SecurityScorecard. Securityscorecard. URL `https://securityscorecard.com/`.

[15] Nan Sun, Jun Zhang, Paul Rimba, Shang Gao, Yang Xiang, and Leo Yu Zhang. Data-driven cybersecurity incident prediction: A survey. *IEEE Communications Surveys & Tutorials*, 2018.

[16] CyberSecurity Ventures. Cybersecurity 500. URL `https://cybersecurityventures.com/cybersecurity-500/#home/?view_1_sort=field_4|asc&view_1_page=1&view_1_per_page=1000,https://www.prnewswire.com/news-releases/cybersecurity-500-2018-the-official-list-300648938.html`.

[17] George Kurtz Stuart Mcclure, Joel Scambray. *Hacking Exposed: Network Security secrets and solutions*. Osborne/McGraw-Hill, 2001.

[18] Zakir Durumeric, Eric Wustrow, and J Alex Halderman. Zmap: Fast internet-wide scanning and its security applications. In *Presented as part of the 22nd {USENIX} Security Symposium ({USENIX} Security 13)*, pages 605–620, 2013.

[19] Google BigQuery. Google bigquery. URL `https://bigquery.cloud.google.com`.

[20] Wayne W Daniel et al. *Applied nonparametric statistics*. Houghton Mifflin, 1978.

[21] Charles Spearman. The proof and measurement of association between two things. *American journal of Psychology*, 15(1):72–101, 1904.

[22] Judea Pearl et al. Causal inference in statistics: An overview. *Statistics surveys*, 3:96–146, 2009.

[23] Verizon RISK Team. The veris community database(vcdb). URL `http://veriscommunity.net/vcdb.html`.

[24] hackmageddon. hackmageddon. URL `https://www.hackmageddon.com/`.

[25] Web-Hacking-Incident-Database. Web-hacking-incident-database. URL `http://projects.webappsec.org/w/page/13246995/Web-Hacking-Incident-Database`.

[26] Ahana Roy, Louis Mejia, Paul Helling, and Aspen Olmsted. Automation of cyber-reconnaissance: A java-based open source tool for information gathering. In *2017 12th International Conference for Internet Technology and Secured Transactions (ICITST)*, pages 424–426. IEEE, 2017.

[27] Nihad A Hassan and Rami Hijazi. Technical footprinting. In *Open Source Intelligence Methods and Tools*, pages 313–339. Springer, 2018.

[28] Armin Sarabi, Parinaz Naghizadeh, Yang Liu, and Mingyan Liu. Prioritizing security spending: A quantitative analysis of risk distributions for different business profiles. In *WEIS*, 2015.

[29] Yang Liu, Jing Zhang, Armin Sarabi, Mingyan Liu, Manish Karir, and Michael Bailey. Predicting cyber security incidents using feature-based characterization of network-level malicious activities. In *Proceedings of the 2015 ACM International Workshop on International Workshop on Security and Privacy Analytics*, pages 3–9. ACM, 2015.

[30] Marie Vasek and Tyler Moore. Identifying risk factors for webserver compromise. In *International Conference on Financial Cryptography and Data Security*, pages 326–345. Springer, 2014.

[31] Olivier Thonnard, Leyla Bilge, Anand Kashyap, and Martin Lee. Are you at risk? profiling organizations and individuals subject to targeted attacks. In *International Conference on Financial Cryptography and Data Security*, pages 13–31. Springer, 2015.

[32] Davide Canali, Leyla Bilge, and Davide Balzarotti. On the effectiveness of risk prediction based on users browsing behavior. In *Proceedings of the 9th ACM symposium on Information, computer and communications security*, pages 171–182. ACM, 2014.

[33] Armin Sarabi and Mingyan Liu. Characterizing the internet host population using deep learning: A universal and lightweight numerical embedding. In *Proceedings of the Internet Measurement Conference 2018*, pages 133–146. ACM, 2018.

[34] Kyle Soska and Nicolas Christin. Automatically detecting vulnerable websites before they turn malicious. In *23rd {USENIX} Security Symposium ({USENIX} Security 14)*, pages 625–640, 2014.

[35] Zhiyun Qian, Zhuoqing Morley Mao, Yinglian Xie, and Fang Yu. On network-level clusters for spam detection. In *NDSS*, 2010.

[36] Anirudh Ramachandran and Nick Feamster. Understanding the network-level behavior of spammers. In *ACM SIGCOMM Computer Communication Review*, volume 36, pages 291–302. ACM, 2006.

[37] Institute of Risk Management. Cyber risk and risk management. URL `https://www.theirm.org/knowledge-and-resources/thought-leadership/cyber-risk/`.

[38] UpGuard. Upguard. URL `https://www.upguard.com/`.

[39] Python 2.7. Python 2.7. URL `https://www.python.org/download/releases/2.7/`.

[40] Jet Brains. Pycharm. URL `https://www.jetbrains.com/pycharm/`.

[41] Clinton Gormley and Zachary Tong. *Elasticsearch: The definitive guide: A distributed real-time search and analytics engine*. " O'Reilly Media, Inc.", 2015.

[42] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[43] Andy Liaw, Matthew Wiener, et al. Classification and regression by randomforest. *R news*, 2 (3):18–22, 2002.

[44] seaborn data visualization. seaborn. URL `https://seaborn.pydata.org`.

[45] Wes McKinney et al. Data structures for statistical computing in python. In *Proceedings of the 9th Python in Science Conference*, volume 445, pages 51–56. Austin, TX, 2010.

[46] Travis Oliphant. NumPy: A guide to NumPy. USA: Trelgol Publishing, 2006–. URL `http://www.numpy.org/`. [Online; accessed ¡today¿].

[47] J. D. Hunter. Matplotlib: A 2d graphics environment. *Computing In Science & Engineering*, 9 (3):90–95, 2007. doi: 10.1109/MCSE.2007.55.

[48] HHS.gov. Health information privacy, hitech act. URL `https://www.hhs.gov/hipaa/for-professionals/special-topics/hitech-act-enforcement-interim-final-rule/index.html`.

[49] Breach Report Portal. U.s department of health and human services. URL `https://ocrportal.hhs.gov/ocr/breach/breach_report.jsf`.

[50] Microsoft Bing. Bing. URL `https://www.bing.com/`.

[51] Google. Google. URL `https://www.google.com/`.

[52] Subdomain Enumeration Cheat Sheet. Subdomain enumeration cheat sheet. URL `https://pentester.land/cheatsheets/2018/11/14/subdomains-enumeration-cheatsheet.html`.

[53] The art of subdomain enumeration. The art of subdomain enumeration. URL `https://blog.sweepatic.com/art-of-subdomain-enumeration/`.

[54] RISKIQ. Riskiq. URL `https://www.riskiq.com/`.

[55] Binary Edge. Binary edge. URL `https://app.binaryedge.io/`.

[56] Security Trails. Security trails. URL `https://securitytrails.com/`.

[57] Virus Total. Virus total. URL `https://www.virustotal.com`.

[58] DNSDB(FarSight). Dnsdb(farsight). URL `https://www.farsightsecurity.com/get-started/`.

[59] Shodan. Shodan. URL https://www.shodan.io/.

[60] F-Secure. Riddler. URL https://riddler.io/.

[61] Certificate Search. crt.sh. URL https://crt.sh/.

[62] Amass. Amass. URL https://github.com/OWASP/Amass.

[63] dnsrecon. dnsrecon. URL https://github.com/darkoperator/dnsrecon.

[64] Sublist3r. Sublist3r. URL https://github.com/aboul3la/Sublist3r.

[65] subfinder. subfinder. URL https://github.com/subfinder/subfinder.

[66] assets-from spf. assets-from-spf. URL https://github.com/0xbharath/assets-from-spf.

[67] domains-from csp. domains-from-csp. URL https://github.com/0xbharath/domains-from-csp.

[68] massdns. massdns. URL https://github.com/blechschmidt/massdns.

[69] Rapid 7. Open data. URL https://opendata.rapid7.com/.

[70] BitSight. Cybersecurity is essential for merger & acquisition due diligence, . URL https://info.bitsight.com/cybersecurity-ma-due-diligence.

[71] Censys. The freak attack. URL https://censys.io/blog/freak.

[72] David Adrian, Karthikeyan Bhargavan, Zakir Durumeric, Pierrick Gaudry, Matthew Green, J Alex Halderman, Nadia Heninger, Drew Springall, Emmanuel Thomé, Luke Valenta, et al. Imperfect forward secrecy: How diffie-hellman fails in practice. In *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security*, pages 5–17. ACM, 2015.

[73] Gang Wang, Tianyi Wang, Haitao Zheng, and Ben Y Zhao. Man vs. machine: Practical adversarial detection of malicious crowdsourcing workers. In *23rd {USENIX} Security Symposium ({USENIX} Security 14)*, pages 239–254, 2014.

[74] Public DNS Server List. Public dns server list. URL https://public-dns.info/.

[75] Hui Guo, Özgür Kafali, and Munindar Singh. Extraction of natural language requirements from breach reports using event inference. In *2018 5th International Workshop on Artificial Intelligence for Requirements Engineering (AIRE)*, pages 22–28. IEEE, 2018.